Received: 31 December 2020 | Revised: 24 June 2021 | Accepted: 26 June 2021

DOI: 10.1002/mbo3.1220

ORIGINAL ARTICLE

MicrobiologyOpen Open Access WILEY

Measuring the absolute abundance of the microbiome by adding yeast containing 16S rRNA gene from a hyperthermophile

Ju Yeong Kim[1,2] | Myung-hee Yi[1] | Myungjun Kim[1] | Seogwon Lee[1] | Hye Su Moon[3] | Dongeun Yong[3] | Tai-Soon Yong[1]

[1]Department of Environmental Medical Biology, Arthropods of Medical Importance Resource Bank, Institute of Tropical Medicine, Yonsei University College of Medicine, Seoul, Korea

[2]Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea

[3]Department of Laboratory Medicine and Research Institute of Bacterial Resistance, Yonsei University College of Medicine, Seoul, Korea

Correspondence
Tai-Soon Yong, Department of Environmental Medical Biology, Institute of Tropical Medicine, Arthropods of Medical Importance Resource Bank, Yonsei University College of Medicine, Seoul 03722, Korea.
Email: tsyong212@gmail.com

Funding information
National Research Foundation of Korea, Grant/Award Number: 2019R1A2B5B01069843 and 2020R1I1A2074562

Abstract

High-throughput sequencing (HTS) of 16S rRNA gene amplicons provides compositional information regarding the microbial community, but not the absolute abundance of the bacteria. We aimed to develop a standardized method for quantifying the absolute abundance of bacteria in microbiome studies. To demonstrate the utility of our approach, we quantified the number of bacteria from the compositional data of the fecal and cecal microbiomes. The 16S rRNA gene of a hyperthermophile, *Thermus aquaticus*, was cloned into *Pichia pastoris* (yeast) genome, and an equivalent amount of the yeast was added to the stool and cecal samples of mice before DNA extraction. 16S rRNA gene library construction and HTS were performed after DNA extraction. The absolute abundances of bacteria were calculated using *T. aquaticus* reads. The average relative abundances of *T. aquaticus* in the five stool and five cecal samples were 0.95% and 0.33%, respectively, indicating that the number of bacteria in a cecum sample is 2.9 times higher than that in a stool sample. The method proposed for quantifying the absolute abundance of the bacterial population in this study is expected to overcome the limitation of showing only compositional data in most microbiome studies.

KEYWORDS
16S rRNA gene, absolute abundance, hyperthermophile, microbiome, *Thermus aquaticus*

1 | INTRODUCTION

High-throughput sequencing (HTS) of 16S rRNA gene amplicon is a reliable and widely used method used in microbiome studies (Poretsky et al., 2014). However, this technique only describes the bacterial composition (relative abundance) and does not quantify the absolute abundances in the samples because the DNA concentration of the 16S rRNA gene is compensated during the construction of the 16S rRNA gene library (Gloor et al., 2017).

Researchers investigating the microbiome have the desire to quantify not only the percentage of each bacterium in the sample but also the absolute quantity of each bacterial species. Bacterial quantification is essential in numerous studies, that is, comparison of the stool microbiome between patients with diarrhea and healthy controls, and comparison of the microbiome in different environments, such as soil and water. In a study of patients with Crohn's disease, absolute abundance of the fecal microbiome was associated with disease phenotype (Vandeputte et al., 2017).

Recently, there have been studies to estimate the absolute abundance of the microbiome by adding internal standards (Harrison et al., 2021). Cellular internal standards use a specific bacterial species mixed into the target sample during the HTS procedure, whereas DNA internal standards use genomic DNA or synthetic DNA. Cellular internal standards may have similar traits to those of microbial organisms in the sample, but they might be more difficult for quantitation and less reproducible than DNA internal standards (Venkataraman et al., 2018). An ideal internal standard should have similar traits to focal organisms, be easily cultured, not occur in the biological samples, not have copy number variation, be easily quantified, and be accurate and reproducible (Harrison et al., 2021).

Thermus aquaticus is a popular hyperthermophile bacterium that was first discovered in the hot spring of Yellowstone National Park (Brock & Freeze, 1969). This bacterium survives at a temperature of 50–80℃ and therefore is not found in the normal environment or animal gut. *T. aquaticus* is the source of Taq DNA polymerase, an important component of the polymerase chain reaction (PCR) (Chien et al., 1976; Saiki et al., 1988).

This study is aimed at devising a new standardized method for quantifying bacteria in samples when analyzing 16S rRNA gene amplicons using the Illumina MiSeq platform. To fulfill the study objectives, the *T. aquaticus* 16S rRNA gene was cloned into the *Pichia pastoris* (yeast) genome for stable amplification. An equivalent amount of this yeast was added to the initial stool and cecum contents of the mice. DNA extraction, 16S rRNA gene library construction, and HTS were performed thereafter for the microbiome analysis. The number of bacteria present in the initial stool and cecum samples was quantified based on the relative abundance of *T. aquaticus* in the samples.

2 | MATERIALS AND METHODS

2.1 | Insertion of the *T. aquaticus* YT 1 16S rRNA gene into the *P. pastoris* GS115 genome

The complete 16S rRNA gene of *T. aquaticus* YT 1 (1470 bp, NCBI Reference Sequence: NR_025900.1) was synthesized and cloned into pUC57 (Bio Basic Inc.). To clone this gene into pPIC9, primers were designed to make the PCR amplicon have Xho I and Not I restriction sites on its 5′ and 3′ end, respectively. The primers are as follows: forward, 5′-AAAAGTCGAGGAATTCGAGCTCGGTACC-3′3 and reverse, 5′-AAAAGCGGCCGCATGATTACGCCAAGCTTGC-3′. Then, the 16S rRNA gene of *T. aquaticus* was cloned into pPIC9 using Xho I and Not I restriction enzymes. The plasmid containing the 16S rRNA gene of *T. aquaticus* was linearized using the Sac I restriction enzyme and integrated into the *P. pastoris* GS115 genome using the Pichia Expression Kit (Invitrogen).

Sanger sequencing was used for sequencing the insert in *P. pastoris* using the 5′hAOX1 primer 5′-GACTGGTTCCAATTGACAAGC-3′ and 3′aAOX1 primer 5′-GCAAATGGCATTCTGACATCC-3′. To detect the integration location of the 16S rRNA gene of *T. aquaticus* in the yeast chromosome, Sanger sequencing was performed using the forward primer 5′-CTGCTGCCTCCCGTAGG-3′ and reverse primer 5′-CTACTGGCTTGGCCATAATT- 3′. The forward primer is the reverse complement sequence of the forward primer that we used in the HTS of the 16S rRNA gene V3-V4 region. The reverse primer is a part of the polyamine oxidase gene in the yeast chromosome 4 that is located close to the alcohol oxidase 1 gene, the integration target region.

2.2 | *P. pastoris* culture and aliquot

Pichia pastoris was cultured in yeast extract–peptone–dextrose (YPD) medium. One unit of *P. pastoris* harboring *T. aquaticus* 16S rRNA gene was defined as 2 million cells of *P. pastoris*. One unit of *P. pastoris* was added to fecal and cecal samples of the mice just before DNA extraction. One stool was used as the fecal sample, whereas the entire content of one cecum was used as the cecal sample.

2.3 | Stool and cecum contents of mice

Five female C57BL/6 mice (7 weeks of age) were obtained from Orient Bio Inc. Mice were euthanized, and the cecum and stool were sampled. All animal studies were approved by the Department of Laboratory Animal Resources Committee of Yonsei University College of Medicine (No. 2020–0056). Animal experiments were performed in pathogen-free conditions and following standard management practices.

2.4 | DNA extraction

DNA was extracted from stool and cecum samples of the mice using FastDNA™ SPIN Kit for Soil (MP Biomedicals), as per the manufacturer's instructions.

2.5 | Serial dilution of stool suspension

Twenty fecal pellets from 10 mice were placed in a 15 ml conical tube and suspended using 5 ml phosphate-buffered saline, and two-fold serial dilutions were performed. DNA extraction was performed using 500 μL of the diluted samples (x1, x1/2, x1/4, x1/8) containing one unit each of *P. pastoris*.

2.6 | HTS of 16S rRNA gene

The V3–V4 region of the 16S rRNA gene was amplified by PCR using the bacterial universal primer pair (forward primer: 5'- TCGTC GGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWG CAG-3'; reverse primer: 5'- GTCTCGTGGGCTCGGAGATGTGTA

TAAGAGACAGGACTACHVGGGTATCTAATCC-3'). A limited cycle amplification step was included to add multiplexing indices and Illumina sequencing adapters. The libraries were normalized, pooled and sequenced using the Illumina MiSeq platform (Illumina MiSeq V3 cartridge [600 cycles]; Illumina), according to the manufacturers' instructions.

2.7 | Bioinformatics and statistics

All analyses were performed with EzBioCloud, a commercially available ChunLab bioinformatic cloud platform for microbiome research (https://www.ezbiocloud.net/) (Kim et al., 2019). Raw reads were processed through a quality check, and low quality (<Q25) reads were filtered using Trimmomatic version 0.32 (Bolger et al., 2014). Paired-end sequence data were then merged using PandaSeq (Masella et al., 2012). Primers were trimmed using the ChunLab in-house program (ChunLab, Inc.), by including a similarity cutoff of 0.8. Background noise was removed from the sequences using the Mothur pre-clustering program, which merges sequences and extracts unique sequences, allowing up to two differences between the sequences (Schloss et al., 2009). The EzBioCloud database (Yoon et al., 2017) was used for a taxonomic assignment using the basic local alignment search tool (BLAST) version 2.2.22 (Altschul et al., 1990), and pairwise alignments were generated to calculate similarity (Myers & Miller, 1988). The UCHIME algorithm and nonchimeric 16S rRNA gene database from EzBioCloud were used to detect chimeric sequences for reads with a best-hit similarity rate of <97% (Edgar et al., 2011). Sequence data were then clustered using a cluster database at high identity with tolerance (CD Hit) and UCLUST algorithms (Edgar, 2010; Fu et al., 2012). To perform the analyses, the reads were normalized to 50,000. The Shannon index (Shannon, 1948), PCoA (Gower, 1966) and PERMANOVA (Anderson, 2001) computed based on the generalized UniFrac distance (Lozupone & Knight, 2005). Wilcoxon rank-sum test was used to evaluate the differences in the number of OTUs (richness), the Shannon index, and relative abundance of *T. aquaticus* between the two groups of samples (stool and caecum). LDA effect size (LEfSe) tool was used to identify the significantly different taxa between the two groups (Segata et al., 2011).

2.8 | Whole-genome sequencing (WGS) of the yeast

WGS was performed to confirm the copy number of the integrated 16S rRNA gene in the yeast. WGS and bioinformatics analysis were performed at Macrogen Inc. The sequencing libraries were prepared according to the manufacturer's instructions of the TruSeq DNA Nano Sample Preparation Kit (Illumina). Briefly, fragmentation of 100 ng of the yeast genomic DNA was performed using adaptive focused acoustics (Covaris Inc.), and the fragmented DNA was end-repaired to create 5′-phosphorylated, blunt-ended dsDNA molecules. Following end-repair, DNA was size-selected using the bead-based method. These DNA fragments undergo the addition of a single "A" base and ligation of TruSeq indexing adapters. Next, we performed sequencing using the HiSeq™ platform (Illumina).

After sequencing, Trimmomatic v0.36 was used to remove adapter sequences and low-quality reads for reducing biases in the analysis (Bolger et al., 2014). In the sample, filtered data were mapped using BWA v0.7.17 with the mem algorithm to the reference genome (Li & Durbin, 2010). The reference genome was chromosome 4 of *P. pastoris* GS115 (*Komagataella phaffii*) sequence (NCBI GenBank: CP014718.1) modified to insert the pPIC9 plasmid containing the 16S rRNA gene into the AOX1 promoter region in the yeast genome (CP014718.1:1592211–1593034). After read mapping, duplicated reads were removed with Sambamba v0.6.7 (Tarasov et al., 2015). The genome coverage and mapping ratio of mapped reads on the reference genome were calculated. The average read depth of a certain region was defined as the sum of the depths of all positions in that region divided by the number of bases in that region.

2.9 | Quantitative PCR

Quantitative PCR was performed to check the change in the amount of bacterial 16S rRNA gene in the yeast over time. The yeast was fully cultivated in 10 ml of YPD medium for 24 h, and 5 μl was inoculated into 10 ml of fresh YPD medium (n = 5). We collected the yeasts, calculated colony-forming units (CFU) and extracted DNA for three consecutive days. Then, quantitative PCR was performed using primers for the arginosuccinase gene of the yeast (forward primer 5′-ACCCGTGAACATGCTTTGCT-3′ and reverse primer 5′-CACCATTCTCTCAAGCTCGT-3′) (Krainer et al., 2012) and primers for the bacterial 16S rRNA gene (BACT1369 5′-CGGTGAATACGTTCYCGG-3′ and PROK1492R 5′-GGWTACCTTGTTACGACTT-3′) (Buchan et al., 2009) with AMPIGENE® qPCR Green Mixes (ENZO Life Sciences).

3 | RESULTS

The 16S rRNA gene of *T. aquaticus* inserted into the *P. pastoris* genome was 100% identical to the NCBI reference sequence, NR_025900.1 (Figure 1a). Then, PCR was used to confirm where the integrated 16S rRNA gene was inserted into the yeast chromosome. PCR was performed using the primer set capable of amplifying the inserted 16S rRNA gene and the polyamine oxidase gene of yeast chromosome 4 because this gene is located next to the AOX1 promoter, which was the suspected integration site (Figure 1b). The PCR product showed the expected band size (1887 bps) in gel electrophoresis, and DNA sequencing (Sanger method) confirmed that it contained the 16S rRNA gene fragment of *T. aquaticus*, an AOX1 promoter, and the polyamine oxidase gene fragment from the yeast chromosome (Figure 1b). WGS was performed to confirm the copy number of the integrated 16S rRNA gene in the yeast. The average

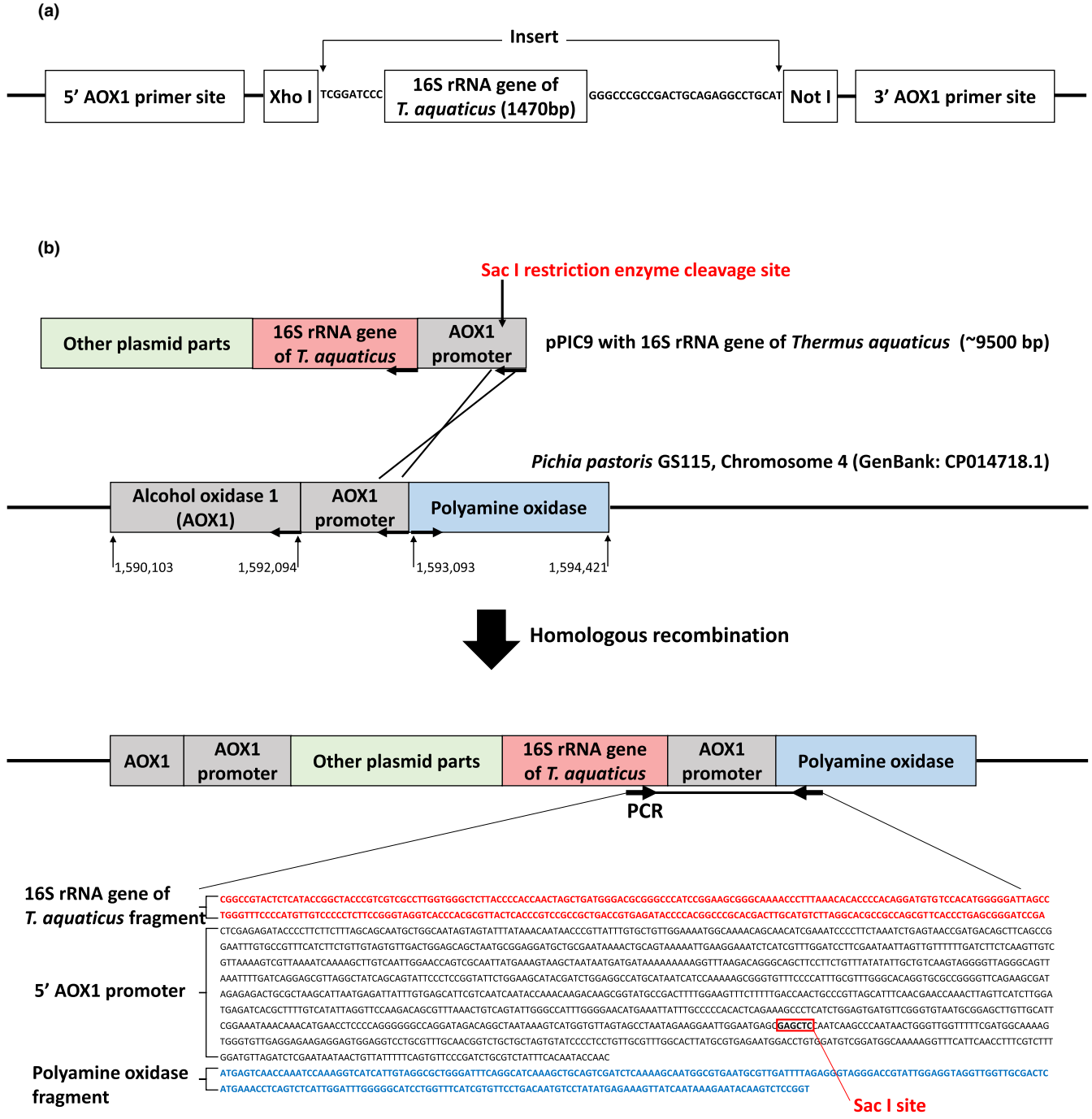


FIGURE 1 DNA sequence and the location of the *Thermus aquaticus* 16S rRNA gene in *Pichia pastoris*. (a) The inserted *T. aquaticus* 16S rRNA gene was 100% identical to the NCBI reference sequence NR_025900.1. (b) Integration of the 16S rRNA gene into the AOX1 promoter in the yeast genome was confirmed by PCR and Sanger sequencing using a primer set encompassing the 16S rRNA gene of the pPIC9 and the polyamine oxidase gene from yeast chromosome 4

read depth of the total genome of the yeast was 237.88 and that of the integrated 16S rRNA gene of *T. aquaticus* was 256.21, which means that a single copy of the 16S rRNA gene was inserted into the yeast genome because *P. pastoris* GS 115 is in a haploid state. This result is in line with the explanation of the kit's protocol that more than 90% of integration cases have one copy of an insert.

For microbiome analysis, total reads corresponding to 10 samples (five each for stool and cecum) ranged from 57,252 to 85,958 (Table 1). Total read numbers corresponding to identified bacteria were not different between the stool and cecal samples ($p = 0.345$), but a greater number of *T. aquaticus* reads were observed in the stool samples than in cecal samples ($p = 0.008$). The average relative

TABLE 1 Reads of *Thermus aquaticus* in the samples

	Total reads	Reads of *Thermus aquaticus*	Relative abundance of *Thermus aquaticus*
Stool 1	80,157	768	0.96%
Stool 2	76,305	728	0.95%
Stool 3	67,976	554	0.81%
Stool 4	85,958	1,068	1.28%
Stool 5	65,856	499	0.76%
Cecum 1	57,252	105	0.18%
Cecum 2	58,837	59	0.10%
Cecum 3	77,891	468	0.60%
Cecum 4	78,261	324	0.41%
Cecum 5	67,136	246	0.37%

TABLE 2 Comparison of relative abundance of *Thermus aquaticus* in stool and cecum

	Stools	Cecum
Average	0.95%	0.33%
Standard deviation	0.18%	0.18%
95% confidence interval	[0.79%, 1.11%]	[0.18%, 0.49%]
Median	0.95%	0.37%
p value	0.009**	

Wilcoxon rank-sum test was performed.

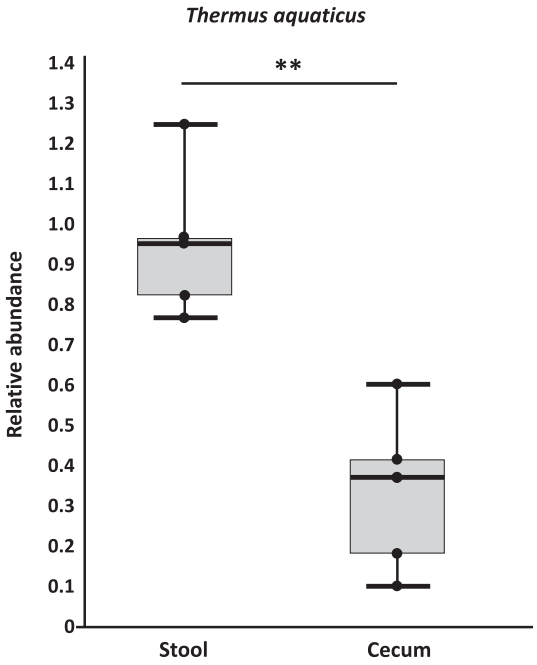


FIGURE 2 Relative abundance of *Thermus aquaticus* composition in fecal (n = 5) and cecal (n = 5) samples. The boxplots indicate the minimum, first quartile, second quartile (median), third quartile, and maximum values. Wilcoxon's rank-sum test was used (** indicates p-value <0.01)

abundances of *T. aquaticus* in the stool and cecal samples were 0.95% and 0.33%, respectively indicating significant differences between the two sample groups (Table 2 and Figure 2).

As the same amount of yeast harboring the *T. aquaticus* 16S rRNA gene was added to each of the samples before DNA extraction, the total amount of bacteria in the original samples can be estimated. When X is the relative abundance of *T. aquaticus* in the stool and Y is the relative abundance of *T. aquaticus* in the cecum, the formula to estimate the ratio of the number of bacteria in the cecum to that in the stool is as follows:

$$\frac{(100-Y)/Y}{(100-X)/X}=\frac{(100-0.33)/0.33}{(100-0.95)/0.95}=2.90$$

In other words, on average a cecum sample has 2.9 times more bacteria than a stool sample.

For individual samples, the relative abundances of bacteria belonging to the Bacteroidaceae family were very similar between stool 1 and cecum 2 samples, 37.50%, and 36.86%, respectively (Figure 3). However, the relative abundances of bacteria belonging to the Thermaceae family (the family name of *T. aquaticus*) were 0.96% and 0.1% in the stool 1 and cecum 2 samples, respectively. In the following formula, the cecum 2 sample was found to have 9.68 times more bacteria than stool 1. Therefore, it is estimated that the number of bacteria belonging to the Bacteroidaceae family in cecum 2 is 9.52 times greater than that in stool 1 (9.68 × 36.86%/37.50%).

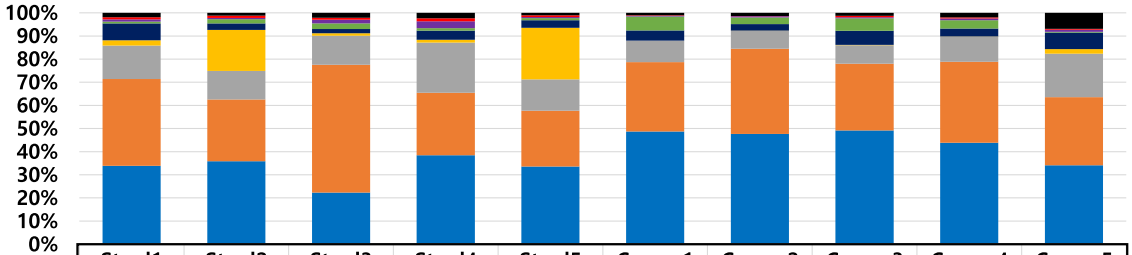
$$\frac{(100-Y)/Y}{(100-X)/X}=\frac{(100-0.10)/0.10}{(100-0.96)/0.96}=9.68$$

(X = the relative abundance of Thermaceae in stool 1, Y = the relative abundance of *T. aquaticus* in cecum 2).

Likewise, the ratios of absolute bacterial abundance from stool 1 to caecum 5 were calculated as 1, 1.00, 1.18, 0.77, 1.27, 5.26, 9.68, 1.60, 2.33, and 2.63 using the relative abundance of Thermaceae in stool 1 as the reference.

The number of operational taxonomic units (OTUs) and the Shannon index were significantly higher in cecum samples (Figure 4a, b), suggesting that the cecal microbiome is more rich and diverse than the stool microbiome. Principal coordinates analysis (PCoA) showed that the samples from the same group were clustered closely in the plot. Permutational multivariate analysis of variance (PERMANOVA) confirmed the significant differences between the microbiome composition of the stool and cecum samples (p = 0.01) (Figure 5).

LEfSe analysis was performed to identify the bacterial species that are differentially enriched in either of the two sample groups. *Mucispirillum schaedleri* had the highest linear discriminant analysis (LDA) score, 4.63, and its average relative abundance was 8.97 in the stool and 0.43 in the cecum (Table A1). The LDA score of *T. aquaticus* (3.49) was the third-highest among all bacterial species (Table A1).



	Stool1	Stool2	Stool3	Stool4	Stool5	Cecum1	Cecum2	Cecum3	Cecum4	Cecum5
Others	1.88	1.21	2.13	2.38	1.08	1.09	1.33	1.30	1.99	6.93
Thermaceae	0.96	0.95	0.82	1.24	0.76	0.18	0.10	0.60	0.41	0.37
Porphyromonadaceae	0.97	0.53	1.63	2.97	0.43	0.33	0.52	0.43	0.79	0.92
Lactobacillaceae	0.74	1.89	2.44	1.08	0.89	5.97	2.92	5.43	3.55	0.29
Christensenellaceae	7.33	2.77	1.82	3.94	3.27	4.44	2.80	6.09	3.46	7.14
Deferribacteraceae	2.27	17.74	1.05	1.23	22.35	0.00	0.00	0.17	0.00	2.00
Ruminococcaceae	14.45	12.36	12.57	21.71	13.51	9.28	7.84	7.99	10.92	18.77
Bacteroidaceae	37.50	26.68	55.20	27.03	24.17	30.00	36.86	28.80	35.03	29.44
Lachnospiraceae	33.90	35.86	22.34	38.41	33.55	48.71	47.64	49.19	43.84	34.14

FIGURE 3 The composition of stool and cecal microbiome at the family level. The arrows are the values used as examples in the result

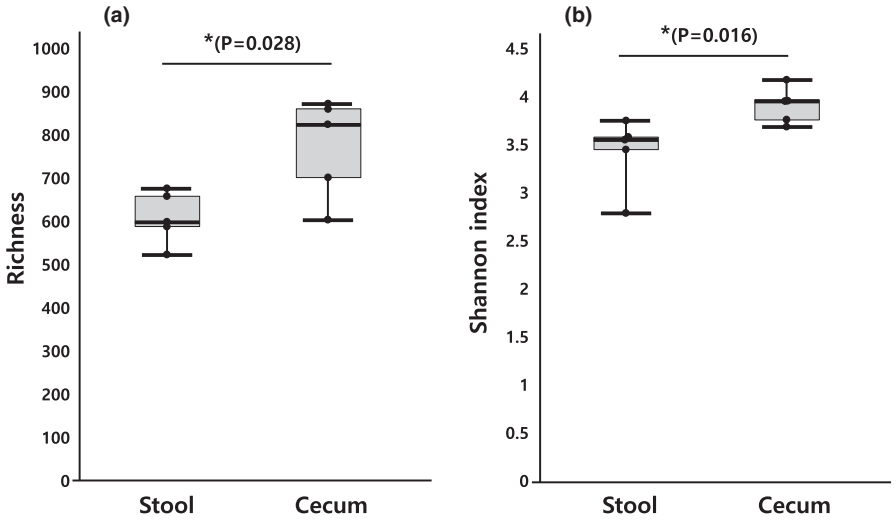


FIGURE 4 Alpha diversity of the microbiome in the stool and cecum of mice. (a) The number of OTUs (richness) and (b) Shannon index. The boxplots indicate the minimum, first quartile, second quartile (median), third quartile, and maximum values. Wilcoxon's rank-sum test was used (* indicates *p*-value <0.05)

To validate this technique, we analyzed twofold serial dilutions of the stool suspension. An equal amount of yeast was added to the diluted samples (x1, x1/2, x1/4 and x1/8); the DNA was extracted and HTS was performed. The results showed that the relative abundances of the added *T. aquaticus* sequence in the twofold serially diluted samples were 0.91%, 1.77%, 3.72%, and 6.97%, respectively (Figure 6). The ratio of the estimated absolute abundances of total bacteria in the diluted samples (x1, x1/2, x1/4 and x1/8) was calculated as 8.15: 4.20: 1.96: 1.

$$\frac{(100-0.91)/0.91}{(100-6.97)/6.97}=8.15$$

$$\frac{(100-0.91)/0.91}{(100-3.72)/3.72}=4.20$$

$$\frac{(100-0.91)/0.91}{(100-1.77)/1.77}=1.96$$

4 | DISCUSSION

HTS of 16S rRNA gene amplicon provides information on the relative abundances of bacteria (compositional data), but not the absolute abundances. In compositional data, as one taxon increases within a sample, other taxa must decrease. The compositional data alone may not reflect the true microbiome of the sample thus resulting in biased statistical results when analyzed in an inappropriate manner (Gloor et al., 2017). To address this problem, we developed a method to measure the absolute abundances of bacterial populations in stool and cecal samples by adding *Pichia pastoris* that had been genetically modified to contain the 16S rRNA gene from a hyperthermophile.

We selected the hyperthermophile *T. aquaticus* as an internal standard because it is rarely present in clinical and environmental samples. In addition, given that *T. aquaticus* is commonly used in research, it was anticipated that every bacterial database will have information on this bacterium. Indeed, the 16S rRNA gene databases,

such as EzBioCloud, SILVA, Greengenes and NCBI had information regarding the nucleotide sequence of the *T. aquaticus* 16S rRNA gene. In the beginning, we had tried to use *E. coli* transformed with pUC57 harboring the *T. aquaticus* 16S rRNA gene directly, but the ribosomal DNA of *E. coli*, the host of plasmid, was constantly detected in the tests. Therefore, we used *P. pastoris* (yeast) as the host for stable amplification of the *T. aquaticus* 16S rRNA gene, without any *E. coli* genome contamination.

There have been numerous studies to estimate absolute abundances by quantitative real-time PCR or cell counting using microscopy or flow cytometry. Compared to using an internal standard, these methods are considered to be costly, labor-intensive, time-consuming and require specialized equipment and skills (Harrison et al., 2021). In addition, the quantitative PCR method is less accurate in calculating the converted absolute abundance than the internal standard method (Stämmler et al., 2016).

Recently, studies have been conducted to estimate absolute abundances in microbiota by incorporating internal standards into samples. There are two types of internal standards for metabarcoding: cellular and DNA. Various cellular internal standards have been studied such as *Escherichia coli*, *Salinibacter ruber*, *Rhizobium radiobacter*, *Alicyclobacillus acidiphilus*, *Sporosarcina pasteurii*, and *Shewanella oneidensis* (Ji et al., 2019; Jones et al., 2015; Piwosz et al., 2018; Stämmler et al., 2016). The ideal cellular internal standard should have similar traits to focal organisms, be easily cultured, not occur in the biological samples and not have copy number variation (Harrison et al., 2021). Genomic DNA can be used as internal standards (Deagle et al., 2018; Ji et al., 2020; Lin et al., 2019; Smets et al., 2016; Venkataraman et al., 2018). Synthetic DNA has also been utilized as an internal standard (Hardwick et al., 2018; Jiang et al., 2011; Tkacz et al., 2018; Tourlousse et al., 2017; Zemb et al., 2020). Synthetic sequences do not occur naturally in samples and the problem with copy number variation can be eliminated.

The method we developed is a powerful approach that maximizes the strengths of both cellular and DNA internal standards, as it uses cells (yeasts) that contain the DNA sequence of the 16S rRNA gene of *T. aquaticus*. The 16S rRNA gene of *T. aquaticus*, a

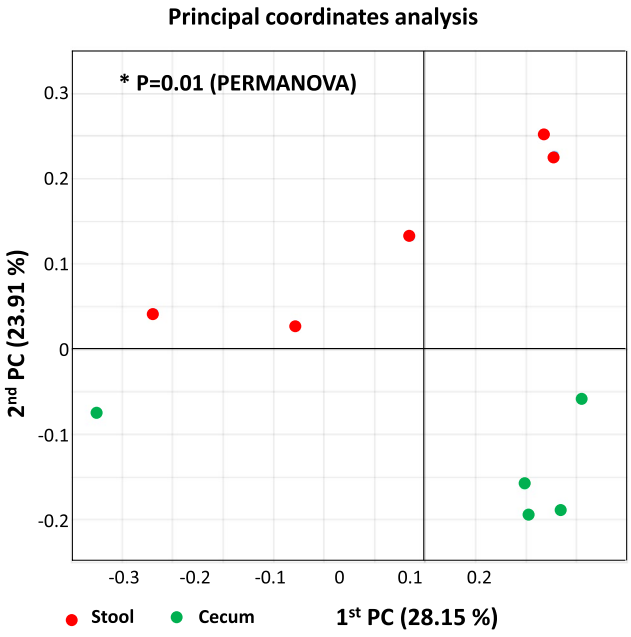
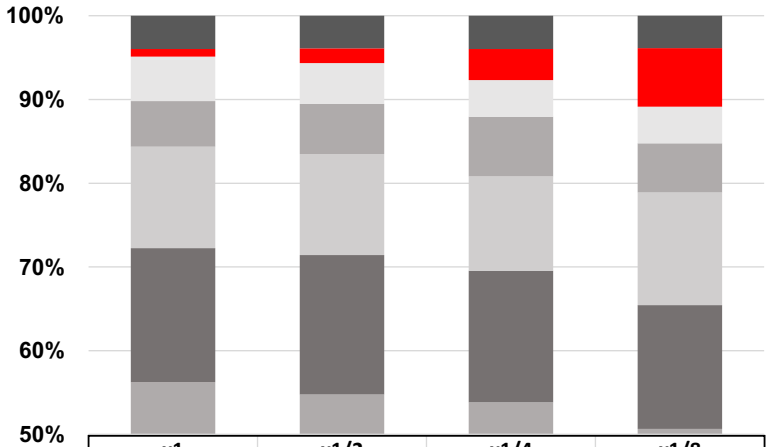


FIGURE 5 Principal coordinates analysis (PCoA) and permutational multivariate analysis of variance (PERMANOVA) of the microbiome in mice stool and cecum. Each dot represents a different sample. Red dots are stool samples and green dots are caecum samples



	x1	x1/2	x1/4	x1/8
Others	3.97	3.90	3.97	3.88
Thermaceae	0.91	1.77	3.72	6.97
Deferribacteraceae	5.33	4.87	4.36	4.37
Lactobacillaceae	5.43	5.98	7.13	5.89
Bacteroidaceae	12.10	12.07	11.26	13.45
Ruminococcaceae	16.02	16.65	15.70	14.77
Lachnospiraceae	56.24	54.77	53.85	50.66

FIGURE 6 The microbial composition of the serially diluted (twofold) stool suspension at the family level

hyperthermophile, is a sequence suitable for metabarcoding because it is registered in all databases but is not usually detected in animals or most environments. The main benefit of our approach is that yeast (eukaryote) lacks a 16S rRNA gene and will not interfere with the analysis. Yeasts are easier to cultivate than hyperthermophiles or halophiles. Furthermore, the copy number of the *T. aquaticus* 16S rRNA gene in all yeast cells is the same over time because all yeast cells used in this technique originated from a single recombination clone. We confirmed that a single copy of the 16S rRNA gene was inserted into the yeast genome and the 16S rRNA gene amount per yeast arginosuccinase gene was unchanged over three consecutive rounds of yeast cultivation (Figure A1a).

Using an existing 16S rRNA gene rather than a synthetic sequence probably makes it easier to analyze metagenomic bioinformatics in ordinary laboratories. Since the bacterial 16S rRNA gene (such as the 16S rRNA gene of *T. aquaticus*) is registered in most databases, it can be used as an internal standard in any preexisting pipeline without methodological modification or even be directly applied to a web-based metagenomics pipeline. Although hyperthermophiles are believed to exist only in extreme environments, a few studies have reported the presence of *Thermus* spp. in some gut and lung microbiomes (Jones et al., 2018; Yu et al., 2016). However, we believe that *T. aquaticus* is very suitable if the bacterial 16S rRNA gene should be used as an internal standard. Other studies have also studied the use of the 16S rRNA gene of *Thermus* spp. as an internal standard (Janes et al., 2020; Lin et al., 2019; Smets et al., 2016). However, when *Thermus* sp. is present in the sample to be studied, using a synthetic sequence as an internal standard can be a solution. In this study, no *Thermus* spp. other than the internal standard *T. aquaticus* were found. The data in Figures 3 and 6 are presented at the level of the family for visual convenience. In these figures, Thermaceae is composed exclusively of *T. aquaticus*.

In addition, it may be difficult to add the same number of newly cultured yeast cells to a sample, as in previous experiments. The yeast we developed did not show the same cell count per culture batch. (Figure A1b). Therefore, we recommend cultivating large quantities of yeast, making aliquots, storing them and using them for each experiment that requires an internal standard.

In this study, the average relative abundances of *T. aquaticus* 16S rRNA gene in five stool samples were 0.95% and at 95% confidence interval, [0.79%, 1.11%] the fluctuating abundances centered on 1% (Table 2). In a pilot study using quantitative PCR, we defined that one unit of yeast to prepare the *T. aquaticus* 16S rRNA gene could represent 1% of the total microbial 16S rRNA gene in one stool sample. When 50,000 total bacterial reads are obtained in one sample, 500 reads of *T. aquaticus* are required to reach 1% of the total reads. To calculate the differences between the samples while minimizing the effects on the remaining 99% of bacteria, 500 reads (1%) were believed to be sufficient. However, because one unit is set as 1% in one stool sample, it may be necessary to adjust the unit when studying samples other than stools. When the DNA density of the sample is too low, it might be effective to use artificial synthetic sequences or qPCR to measure the absolute abundances of the samples.

The study provided the necessary formula for comparing the absolute abundances of bacteria between the two sample groups, stool, and cecum, that is, five stool samples and five cecal samples. The average relative abundances of *T. aquaticus* in stool and cecal samples were calculated to 0.95% and 0.33%, respectively. The value obtained using the proposed formula revealed that a cecum sample contains 2.9 times more bacteria than a stool sample. This is likely due to the difference in the original amount of samples between the two groups.

In addition, the yeast we developed and the calculation used showed very accurate estimation ability in validation experiments using serial dilutions (Figure 6). It calculated the ratio of the total bacteria in twofold diluted samples (x1, x1/2, x1/4, and x1/8) as 8.15:4.20:1.96:1. A stool sample without internal standards was not included as a control because *T. aquaticus* is not expected to be present in fecal samples.

When comparing the absolute abundances of a specific bacterial taxon between two samples, the relative abundance of the target bacteria can be multiplied by the value derived using the formula. In the two sample groups, the relative abundances of bacteria belonging to the Bacteroidaceae family were similar, but the absolute abundance was 9.52 times higher in the cecum 2 sample than that in the stool 1 sample. However, it should be noted that the resulting values from our research method cannot be expressed in an exact unit (e.g., copies per g) and the efficiency of cell lysis for yeast may be different than that for bacteria, so the values have meaning mostly in a comparative sense.

In addition, the number of OTUs (richness) and Shannon index (alpha diversity analysis) revealed that cecal samples exhibit greater microbial diversity than the stool samples. PCoA and PERMANOVA confirmed the differences in microbial composition between the two sample groups. The observations of the current study are concordant with previous studies (Gu et al., 2013; Tanca et al., 2017).

However, the effect of adding the *T. aquaticus* 16S rRNA gene on the microbiome analyses, such as alpha and beta diversity cannot be excluded. This limitation can be overcome in various ways. One way is to add a trace amount of the *T. aquaticus* 16S rRNA gene to minimize its effect on the analysis of other bacterial taxa. The next method is to subtract the *T. aquaticus* reads from the taxonomy output before proceeding with the microbiome analysis. Lastly, HTS can be performed using the same samples with and without adding the *T. aquaticus* 16S rRNA gene.

5 | CONCLUSION

In this study, we developed a method to measure the absolute abundances of the microbial community when performing HTS and 16S rRNA gene amplicon analysis by adding *P. pastoris* harboring *T. aquaticus* 16S rRNA gene to the stool and cecal samples of mice.

ACKNOWLEDGEMENTS

This study was supported by a grant from the National Research Foundation of Korea (NRF) funded by the Korean Government (MEST; numbers NRF-2019R1A2B5B01069843 and 2020R1I1A2074562).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Ju Yeong Kim: Conceptualization (lead); Writing-original draft (lead). **Myung-hee Yi:** Investigation (lead); Methodology (lead). **Myungjun Kim:** Validation (lead). **Seogwon Lee:** Data curation (lead). **Hye Su Moon:** Resources (lead). **Dongeun Yong:** Formal analysis (lead). **Tai-Soon Yong:** Funding acquisition (lead); Supervision (lead).

ETHICS STATEMENT

None required.

DATA AVAILABILITY STATEMENT

Raw sequence data for microbiome analysis and WGS are available in NCBI GenBank under BioProject PRJNA688109: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA688109 and BioProject PRJNA718115: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA718115, respectively.

ORCID

Ju Yeong Kim https://orcid.org/0000-0003-2456-6298
Myung-hee Yi https://orcid.org/0000-0001-9537-5726

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brock, T. D., & Freeze, H. (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of Bacteriology*, *98*(1), 289–297.

Buchan, A., Hadden, M., & Suzuki, M. T. (2009). Development and application of quantitative-PCR tools for subgroups of the roseobacter clade. *Applied and Environmental Microbiology*, *75*(23), 7542–7547.

Chien, A., Edgar, D. B., & Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *Journal of Bacteriology*, *127*(3), 1550–1557. https://doi.org/10.1128/jb.127.3.1550-1557.1976

Deagle, B. E., Clarke, L. J., Kitchener, J. A., Polanowski, A. M., & Davidson, A. T. (2018). Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Molecular Ecology Resources*, *18*(3), 391–406. https://doi.org/10.1111/1755-0998.12740

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C., Hass, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, *53*, 325–338. https://doi.org/10.1093/biomet/53.3-4.325

Gu, S., Chen, D., Zhang, J. N., Lv, X., Wang, K., Duan, L. P., Nie, Y., & Wu, X. L. (2013). Bacterial community mapping of the mouse gastrointestinal tract. *PLoS One*, *8*(10), e74957. https://doi.org/10.1371/journal.pone.0074957

Hardwick, S. A., Chen, W. Y., Wong, T., Kanakamedala, B. S., Deveson, I. W., Ongley, S. E., Santini, N. S., Marcellin, E., Smith, M. A., Nielsen, L. K., Lovelock, C. E., Neilan, B. A., & Mercer, T. R. (2018). Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. *Nature Communications*, *9*(1), 3096. https://doi.org/10.1038/s41467-018-05555-0

Harrison, J. G., Clader, W. J., Shuman, B., & Buerkle, C. A. (2021). The quest for absolute abundance: The use of internal standards for DNA-based community ecology. *Molecular Ecology Resources*, *21*(1), 30–43. https://doi.org/10.1111/1755-0998.13247

Janes, V. A., van der Laan, J. S., Matamoros, S., Mende, D. R., de Jong, M. D., & Schultsz, C. (2020). Thermus thermophilus DNA can be used as internal control for process monitoring of clinical metagenomic next-generation sequencing of urine samples. *Journal of Microbiolcal Methods*, *176*, 106005. https://doi.org/10.1016/j.mimet.2020.106005

Ji, B. W., Sheth, R. U., Dixit, P. D., Huang, Y., Kaufman, A., Wang, H. H., & Vitkup, D. (2019). Quantifying spatiotemporal variability and noise in absolute microbiota abundances using replicate sampling. *Nature Methods*, *16*(8), 731–736. https://doi.org/10.1038/s41592-019-0467-y

Ji, Y., Huotari, T., Roslin, T., Schmidt, N. M., Wang, J., Yu, D. W., & Ovaskainen, O. (2020). SPIKEPIPE: A metagenomic pipeline for the accurate quantification of eukaryotic species occurrences and intraspecific abundance change using DNA barcodes or mitogenomes. *Molecular Ecology Resources*, *20*(1), 256–267. https://doi.org/10.1111/1755-0998.13057

Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., Gingeras, T. R., & Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research*, *21*(9), 1543–1551. https://doi.org/10.1101/gr.121095.111

Jones, M. B., Highlander, S. K., Anderson, E. L., Li, W., Dayrit, M., Klitgord, N., Fabani, M. M., Seguritan, V., Green, J., Pride, D. T., Yooseph, S., Biggs, W., Nelson, K. E., & Venter, J. C. (2015). Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 14024–14029. https://doi.org/10.1073/pnas.1519288112

Jones, R. B., Zhu, X., Moan, E., Murff, H. J., Ness, R. M., Seidner, D. L., Sun, S., Yu, C., Dai, Q., Fodor, A. A., Azcarate-Peril, M. A., & Shrubsole, M. J. (2018). Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples. *Scientific Reports*, *8*(1), 4139. https://doi.org/10.1038/s41598-018-22408-4

Kim, J. Y., Kim, E. M., Yi, M. H., Lee, J. Y., Lee, S. W., Hwang, Y. J., Yong, D. E., Sohn, W. M., & Yong, T. S. (2019). Chinese liver fluke Clonorchis sinensis infection changes the gut microbiome and increases probiotic Lactobacillus in mice. *Parasitology Research*, *118*(2), 693–699. https://doi.org/10.1007/s00436-018-6179-x

Krainer, F. W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O., & Glieder, A. (2012). Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway. *Microbial Cell Factories*, *11*, 22. https://doi.org/10.1186/1475-2859-11-22

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

Lin, Y., Gifford, S., Ducklow, H., Schofield, O., & Cassar, N. (2019). Towards quantitative microbiome community profiling using internal standards. *Applied and Environmental Microbiology*, *85*(5), e02634–e2718. https://doi.org/10.1128/AEM.02634-18

Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005

Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G., & Neufeld, J. D. (2012). PANDAseq: Paired-end assembler for illumina sequences. *BMC Bioinformatics*, *14*, 31. https://doi.org/10.1186/1471-2105-13-31.

Myers, E. W., & Miller, W. (1988). Optimal alignments in linear space. *Computer Applications in the Biosciences*, *4*(1), 11–17. https://doi.org/10.1093/bioinformatics/4.1.11

Piwosz, K., Shabarova, T., Tomasch, J., Šimek, K., Kopejtka, K., Kahl, S., Pieper, D. H., & Koblížek, M. (2018). Determining lineage-specific bacterial growth curves with a novel approach based on amplicon reads normalization using internal standard (ARNIS). *The ISME Journal*, *12*(11), 2640–2654. https://doi.org/10.1038/s41396-018-0213-y

Poretsky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PLoS One*, *9*(4), e93827. https://doi.org/10.1371/journal.pone.0093827

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*(4839), 487–491.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Horn, D. J. V., & Weber, C. F. (2009). Introducing mothur: Open-source, platform-independent, community supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), R60. https://doi.org/10.1186/gb-2011-12-6-r60

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Smets, W., Leff, J. W., Bradford, M. A., McCulley, R. L., Lebeer, S., & Fierer, N. (2016). A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biology and Biochemistry*, *96*, 145–151. https://doi.org/10.1016/j.soilbio.2016.02.003

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P. J., Gessner, A., & Spang, R. (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome*, *4*(1), 28. https://doi.org/10.1186/s40168-016-0175-0

Tanca, A., Manghina, V., Fraumene, C., Palomba, A., Abbondio, M., Deligios, M., Silverman, M., & Uzzau, S. (2017). Metaproteogenomics reveals taxonomic and functional changes between cecal and fecal microbiota in mouse. *Frontiers in Microbiology*, *8*, 391. https://doi.org/10.3389/fmicb.2017.00391

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Tkacz, A., Hortala, M., & Poole, P. S. (2018). Absolute quantitation of microbiota abundance in environmental samples. *Microbiome*, *6*(1), 110. https://doi.org/10.1186/s40168-018-0491-7

Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., & Sekiguchi, Y. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research*, *45*(4), e23.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R. Y., Commer, L. D., Darzi, Y., Vermeire, S., Falony, G., & Raes, J. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, *551*(7681), 507–511.

Venkataraman, A., Parlov, M., Hu, P., Schnell, D., Wei, X., & Tiesman, J. P. (2018). Spike-in genomic DNA for validating performance of metagenomics workflows. *BioTechniques*, *65*(6), 315–321. https://doi.org/10.2144/btn-2018-0089

Yoon, S. H., Ha, S. M., Kwon, S. J., Lim, J. M., Kim, Y. S., Seo, H. S., & Chun, J. S. (2017). Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology*, *67*(5), 1613–1617. https://doi.org/10.1099/ijsem.0.001755

Yu, G., Gail, M. H., Consonni, D., Carugno, M., Humphrys, M., Pesatori, A. C., Caporaso, N. E., Goedert, J. J., Ravel, J., & Landi, M. T. (2016). Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features. *Genomic Biology*, *17*(1), 163. https://doi.org/10.1186/s13059-016-1021-1

Zemb, O., Achard, C. S., Hamelin, J., De Almeida, M.-L., Gabinaud, B., Cauquil, L., Verschuren, L. M. G., & Godon, J. J. (2020). Absolute quantitation of microbes using 16S rRNA gene metabarcoding: A rapid normalization of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in standard. *MicrobiologyOpen*, *9*(3), e977. https://doi.org/10.1002/mbo3.977

How to cite this article: Kim, J. Y., Yi, M.-H., Kim, M., Lee, S., Moon, H. S., Yong, D., & Yong, T.-S. (2021). Measuring the absolute abundance of the microbiome by adding yeast containing 16S rRNA gene from a hyperthermophile. *MicrobiologyOpen*, 10, e1220. https://doi.org/10.1002/mbo3.1220

APPENDIX

TABLE A1 Linear discriminant analysis effect size (LEfSe) analysis of differentially abundant bacterial taxa between stool and cecum samples. Only taxa (species) meeting an LDA significant threshold of >2 are shown

Taxon name	Taxonomy	LDA effect size	Relative abundance in Stool	Relative abundance in Cecum
Mucispirillum schaedleri	Deferribacteres: Deferribacteres_c: Deferribacterales: Deferribacteraceae: *Mucispirillum*	4.63034	8.96960	0.43160
KE159600_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159600_g	3.59561	0.11520	0.90320
Thermus aquaticus	Deinococcus-Thermus: Deinococci: Thermales: Thermaceae: *Thermus*	3.48987	0.95320	0.33560
KE159605_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159605_g	3.39010	0.04400	0.53480
AB626924_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001228_g	3.35775	0.47080	0.01520
PAC001360_g_uc	Firmicutes: Clostridia: Clostridiales: Christensenellaceae: PAC001360_g	3.28489	0.21920	0.60440
PAC001125_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Kineothrix*	3.19065	0.34520	0.03520
PAC001757_s group	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g21	3.14647	0.03840	0.31840
PAC001782_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: PAC000661_g	3.13769	0.30400	0.02960
PAC001104_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC000664_g	3.09174	0.04800	0.29480
PAC002512_s group	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g21	3.00307	0.02440	0.22560
PAC001294_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g21	2.99883	0.06840	0.26760
PAC001706_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Oscillibacter*	2.99829	0.22080	0.02200
PAC001092_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001092_g	2.95863	0.01600	0.19760
PAC001727_s group	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001385_g	2.92810	0.01320	0.18240
PAC001684_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Oscillibacter*	2.82737	0.16000	0.02600
PAC001638_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Oscillibacter*	2.81236	0.17840	0.04880
PAC001557_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001043_g	2.75242	0.01120	0.12400
PAC001374_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g24	2.72112	0.01400	0.11880
PAC001361_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001103_g	2.71064	0.00600	0.10840
KE159781_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g21	2.68962	0.04400	0.14160
PAC001382_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001092_g	2.66791	0.01640	0.10920
PAC001772_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g21	2.65644	0.01600	0.10640
KE159714_s group	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Oscillibacter*	2.64527	0.11760	0.02960

(Continues)

TABLE A1 (Continued)

Taxon name	Taxonomy	LDA effect size	Relative abundance in Stool	Relative abundance in Cecum
PAC001542_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Pseudoflavonifractor*	2.63569	0.11080	0.02480
PAC002401_s group	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159605_g	2.62727	0.01320	0.09760
PAC001668_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001092_g	2.61322	0.01800	0.09960
PAC000668_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Monoglobus*	2.60557	0.00120	0.08160
PAC001744_s group	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Frisingicoccus*	2.60481	0.05280	0.13280
AB606336_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Agathobaculum*	2.60178	0.02840	0.10800
Adlercreutzia equolifaciens group	ActinoCoriobacteriia: Coriobacteriales: Coriobacteriaceae: *Adlercreutzia*	2.59700	0.03080	0.10960
Flintibacter butyricus group	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Pseudoflavonifractor*	2.59454	0.08120	0.15960
*Clostridium*_g24_uc	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g24	2.58135	0.00320	0.07920
PAC001088_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Pseudoflavonifractor*	2.58021	0.09440	0.01880
PAC002042_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC002042_g	2.51002	0.01200	0.07640
PAC001476_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Agathobaculum*	2.50193	0.01120	0.07440
PAC002397_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Pseudoflavonifractor*	2.45536	0.06200	0.00520
PAC000692_g_uc	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC000692_g	2.42353	0.00600	0.05880
PAC001535_s group	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC000664_g	2.42156	0.07000	0.01760
JX095379_g_uc	Firmicutes: Clostridia: Clostridiales: Christensenellaceae: JX095379_g	2.41006	0.05840	0.00720
PAC001360_s	Firmicutes: Clostridia: Clostridiales: Christensenellaceae: PAC001360_g	2.38939	0.06560	0.01680
PAC001087_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Pseudoflavonifractor*	2.35129	0.08960	0.04520
PAC002147_s	Firmicutes: Clostridia: Clostridiales: Christensenellaceae: PAC002147_g	2.32864	0.04320	0.00080
EF098562_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159600_g	2.32697	0.00360	0.04560
PAC002354_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Acutalibacter*	2.32197	0.01360	0.05480
PAC002159_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: PAC000661_g	2.32060	0.04680	0.00520
KE159600_g_uc	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159600_g	2.30840	0.00000	0.04040
PAC001096_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Sporobacter*	2.29992	0.04720	0.00760
PAC001092_g_uc	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001092_g	2.25716	0.00920	0.04440

(Continues)

TABLE A1 (Continued)

Taxon name	Taxonomy	LDA effect size	Relative abundance in Stool	Relative abundance in Cecum
PAC001043_g_uc	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001043_g	2.25524	0.00080	0.03640
PAC002555_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: PAC002555_g	2.23590	0.03480	0.00080
PAC001536_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: PAC000661_g	2.23329	0.04240	0.00840
EU511112_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159600_g	2.22109	0.01200	0.04400
PAC001767_s	Firmicutes: Clostridia: Clostridiales: Mogibacterium_f: PAC001236_g	2.20526	0.01760	0.04920
PAC001396_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159605_g	2.18935	0.01000	0.04040
PAC000184_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g24	2.18879	0.00160	0.03160
PAC002471_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC002471_g	2.17289	0.03320	0.00400
AB626948_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Oscillibacter*	2.12901	0.03760	0.01200
PAC001551_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Oscillibacter*	2.12092	0.04480	0.01880
PAC001473_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC000671_g	2.10215	0.00360	0.02800
AF371672_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: LLKB_g	2.07844	0.00000	0.02360
PAC001222_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: KE159797_g	2.07604	0.00200	0.02560
PAC001543_s	Firmicutes: Clostridia: Clostridiales: Ruminococcaceae: *Pseudoflavonifractor*	2.05036	0.01680	0.03840
PAC001085_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Clostridium*_g21	2.04864	0.00560	0.02720
PAC001681_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001681_g	2.04810	0.01120	0.03320
PAC002350_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001372_g	2.04045	0.00440	0.02520
PAC001722_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: *Coprococcus*_g2	2.04004	0.00400	0.02560
PAC001296_s	Firmicutes: Clostridia: Clostridiales: Lachnospiraceae: PAC001296_g	2.00101	0.00000	0.00120

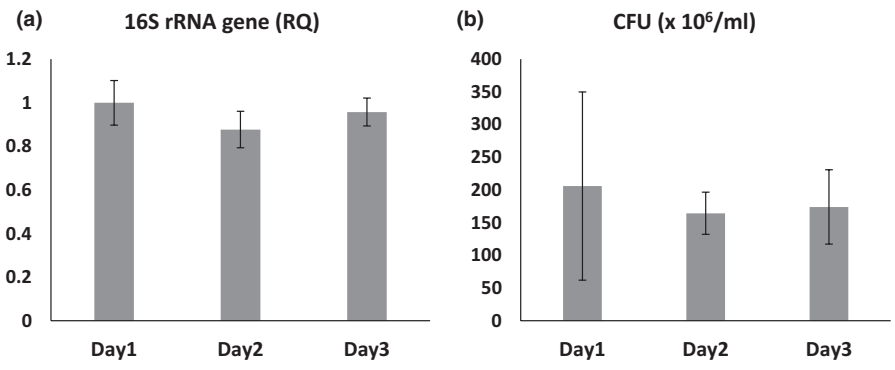


FIGURE A1 (a) The relative quantity (RQ) of the 16S rRNA gene and (b) colony-forming units (CFU) of the recombinant yeast over three consecutive rounds of yeast cultivation (n = 5 for each day)