







Impact of Preservation Method and 16S rRNA Hypervariable Region on Gut Microbiota Profiling

 Zigui Chen,^{a,b} Pak Chun Hui,^b  Mamie Hui,^{a,b} Yun Kit Yeoh,^{a,b} Po Yee Wong,^b  Martin C. W. Chan,^b Martin C. S. Wong,^{a,c} Siew C. Ng,^{a,d,e} Francis K. L. Chan,^{a,d}  Paul K. S. Chan^{a,b}

^aCentre for Gut Microbiota Research, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

^bDepartment of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

^cJockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

^dDepartment of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

^eLKS Institute of Health Science, State Key Laboratory of Digestive Disease, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China

ABSTRACT Proper preservation of stool samples to minimize microbial community shifts and inactivate infectious agents is important for self-collected specimens requiring shipment to laboratories when cold chain transport is not feasible. In this study, we evaluated the performance of six preservation solutions (Norgen, OMNI, RNAlater, CURNA, HEMA, and Shield) for these aspects. Following storage of human stool samples with these preservatives at room temperature for 7 days, three hypervariable regions of the bacterial 16S rRNA gene (V1-V2, V3-V4, and V4) were amplicon sequenced. We found that samples collected in two preservatives, Norgen and OMNI, showed the least shift in community composition relative to -80°C standards compared with other storage conditions, and both efficiently inhibited the growth of aerobic and anaerobic bacteria. RNAlater did not prevent bacterial activity and exhibited relatively larger community shift. Although the effect of preservation solution was small compared to intersubject variation, notable changes in microbiota composition were observed, which could create biases in downstream data analysis. When community profiles inferred from different 16S rRNA gene hypervariable regions were compared, we found differential sensitivity of primer sets in identifying overall microbial community and certain bacterial taxa. For example, reads generated by the V4 primer pair showed a higher alpha diversity of the gut microbial community. The degenerate 27f-YM primer failed to detect the majority of *Bifidobacteriales*. Our data indicate that choice of preservation solution and 16S rRNA gene primer pair are critical determinants affecting gut microbiota profiling.

IMPORTANCE Large-scale human microbiota studies require specimens collected from multiple sites and/or time points to maximize detection of the small effects in microbe-host interactions. However, batch biases caused by experimental protocols, such as sample collection, massively parallel sequencing, and bioinformatics analyses, remain critical and should be minimized. This work evaluated the effects of preservation solutions and bacterial 16S rRNA gene primer pairs in revealing human gut microbiota composition. Since notable changes in detecting bacterial composition and abundance were observed among choice of preservatives and primer pairs, a consistent methodology is essential in minimizing their effects to facilitate comparisons between data sets.

KEYWORDS 16S rRNA gene, amplicon sequencing, bacterial culture, gut microbiota, preservative

Humans harbor an enormous variety of bacteria, archaea, fungi, and viruses that form a community playing important roles in our metabolism and immune system (1, 2). Surveys of microbiomes inside the gastrointestinal tract, where most host-

Citation Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, Wong MCS, Ng SC, Chan FKL, Chan PKS. 2019. Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. *mSystems* 4:e00271-18. <https://doi.org/10.1128/mSystems.00271-18>.

Editor J. Gregory Caporaso, Northern Arizona University

Copyright © 2019 Chen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Zigui Chen, zigui.chen@cuhk.edu.hk, or Paul K. S. Chan, paulkschan@cuhk.edu.hk.

Received 28 October 2018

Accepted 1 February 2019

Published 26 February 2019

associated microbes reside, indicate that the disruption of microorganisms and their abundances could be linked to a number of diseases, such as carcinogenesis (3), cardiovascular disease (4), diabetes (5), hypertension (6), inflammatory bowel disease (7), and obesity (8). The study of gut microbiome has been empowered in a culture-independent manner through high-throughput sequencing and bioinformatics analyses (9–11). Because composition of gut microbiota differs greatly between individuals, microbiome studies have expanded into larger populations at multiple sites and/or time points to maximize detection of the small effects of microbe-host interactions. However, biases caused by methods of sample collection and choice of 16S rRNA gene hypervariable region remain critical and should be minimized.

Immediately freezing stool samples at -20°C or below is considered the “gold standard” to prevent shifting of microbial community composition (12). However, this approach is not feasible for sample collection in remote areas without reliable cold chain transport, or in studies where subjects are requested to send self-collected household samples to laboratories at ambient temperature. The effects of several storage conditions in revealing microbiota composition have been compared to immediate freezing at -80°C . For example, stool samples stored at room temperature for at most 1 day, or at 4°C and -20°C for up to 2 weeks, had little effect on microbiota composition shift (13–15). When samples are exposed to room temperature or higher other than ultralow freezing, however, use of preservation methods becomes essential and critical. Ethanol in concentrations of 95% or higher preserved bacterial DNA for long-term storage well (15), but another study has reported low DNA yield as a potential drawback of using ethanol (16). In addition, self-collected household samples stabilized with ethanol (70% and above) might require special shipping, depending upon local transportation regulations. RNAlater, commonly used as a general preservative for RNA samples, has been shown to have decreased DNA purity and to perform relatively poorly in maintaining microbiota composition derived by 16S rRNA gene amplicon sequencing (16, 17). Buffers containing EDTA can inhibit the growth of certain bacteria (18), but they significantly influenced gut microbiota profiles, such as an increased abundance of *Bacteroides* and *Proteobacteria* and a reduction of *Firmicutes* and *Actinobacteria* (13). OMNIgene.Gut, a commercially available approach to preserve stool samples at ambient temperature, has recently been reported to generate community profiles that diverge the least from -80°C standards (13, 19). However, its effectiveness relative to other approaches needs to be independently assessed over longer time scales. Moreover, these prior studies have generally not evaluated the bactericidal ability of preservatives in rendering microbiota samples noninfectious.

Since microbiome research is quickly shifting toward determination of small effect size variations relevant to host health, understanding the effects of preservative solutions is crucial to minimize bias. As part of the protocol development for a multiple-site, large-scale longitudinal microbiota study, we evaluated the effects of six preservative solutions (Norgen, OMNI, RNAlater, CURNA, HEMA, and Shield) in retaining gut microbiota composition compared to immediate freezing at -80°C . Stool samples from five healthy individuals were collected and subjected to storage in different preservation solutions for 7 days, a time period typically sufficient for delivery to the laboratory by post. Three hypervariable regions of the bacterial 16S rRNA gene (V1-V2, V3-V4, and V4) were amplicon sequenced. The bactericidal ability of these preservatives was assessed by standard aerobic and anaerobic cultures.

RESULTS

Study subjects. Five healthy individuals provided fresh stool samples (Fig. 1; see also Table S1 in the supplemental material). For each subject, one subset of sample was immediately frozen at -80°C ; eight aliquots were stored in different solutions, including six preservatives (Norgen, OMNI, RNAlater, CURNA, HEMA, and Shield), phosphate-buffered saline (PBS) solution, and Cary-Blair (CB) transport medium (Table 1). Following 7 days of preservation at room temperature, all aliquots were stored at -80°C without a freeze-thaw cycle until DNA extraction. Total DNA, including -80°C “gold standards,”

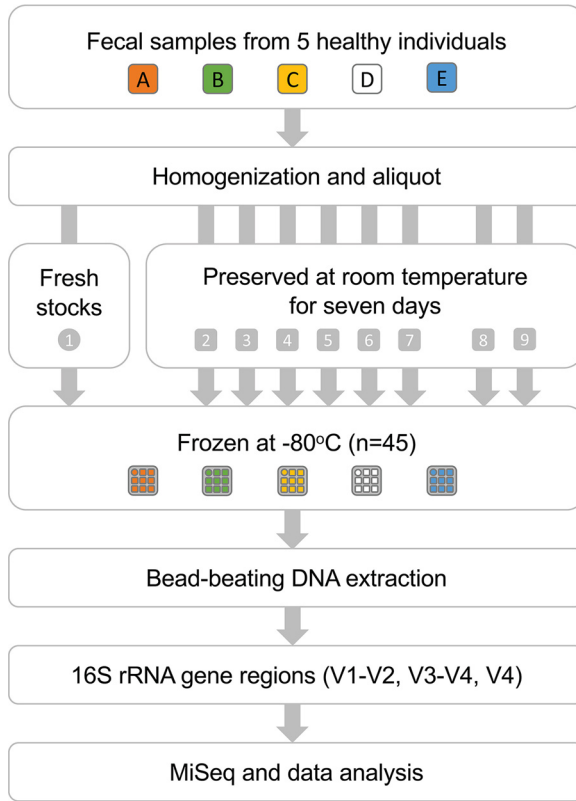


FIG 1 Workflow of the study design. Nine stool sample aliquots from each of the five donors were subjected to different preservatives and incubated at room temperature for 7 days to compare shifts in gut microbial community composition.

was extracted with a bead-beating method and amplicon sequenced on an Illumina MiSeq platform by targeting three hypervariable regions of the bacterial 16S rRNA gene (V1-V2, V3-V4, and V4) (Table 2). No obvious difference in quality of extracted DNA for PCR amplification was observed between subjects or preservatives, as visualized on the agarose gel with similar band brightness (data not shown).

Variation in gut microbial communities without preservative at room temperature. We first used sequences representing the V3-V4 region of the 16S rRNA gene to assess gut microbiota profile since this region provides the longest read length (Table 2). A total of 121,532 high-quality sequence reads for 45 samples (mean ± SD of 2,700 ± 667) was obtained, clustering into 308 amplicon sequence variants (ASVs). These ASVs represented 105 bacterial taxa at genera or higher levels, while 64 were detected at ≥1% relative abundance in at least one sample. The gut microbial community assigned at order level is displayed in Fig. 2A. As expected, the nonpreserved

TABLE 1 List of DNA/RNA preservation methods used in this study

Name ^a	Description	Producer	Catalogue no.
1_Standard	Stock immediately frozen at −80°C		
2_Norgen	Stool nucleic acid preservative	Norgen Biotek	63700
3_OMNI	OMNIGene.GUT	DNA Genotek	OMR-200
4_RNAlater	RNAlater RNA stabilization reagent	Qiagen	76106
5_CURNA	Homemade RNA preservative		
6_HEMA	HEMAGene.BUFFY COAT	DNA Genotek	HG-BCD-50
7_Shield	DNA/RNA shield transport medium	Zymo Research	R1100-50
8_PBS	Phosphate-buffered saline	Thermo Fisher	10010023
9_CB	Cary-Blair transport medium	Puritan	CB-206

^aThe original stocks (1_Standard) were immediately frozen at −80°C. All samples preserved in different solutions (samples 2 to 9) were incubated at room temperature for 7 days and then frozen at −80°C.

TABLE 2 16S rRNA gene PCR primer sequences used in this study

Region	Amplicon size (bp) ^a	Name	Direction	Primer sequence (5'–3')
V1-V2	349	27F-YM	Forward	AGAGTTTGATYMTGGCTCAG
		338R	Reverse	TGCTGCCTCCCGTAGGAGT
V3-V4	466	341F	Forward	CCTACGGGNGGCWGCAG
		806RB	Reverse	GGACTACNVGGGTWTCTAAT
V4	292	515F	Forward	GTGYCAGCMGCCGCGGTA
		806RB	Reverse	GGACTACNVGGGTWTCTAAT

^aMean amplicon size shown for hypervariable region amplified plus PCR primers.

stool samples stored in PBS (8_PBS) and Cary-Blair transport medium (9_CB) at room temperature for 7 days had significant changes in microbiota composition compared to -80°C standards and preserved stocks (samples 1 to 7), as discriminated by a principal-coordinate analysis (PCoA) using either weighted or unweighted UniFrac distances (Fig. 2B and C). Using the Adonis function in R's package Vegan, however, a permutational multivariate analysis of variance (PERMANOVA) based on weighted UniFrac distances ($df = 2$, $R^2 = 0.504$, pseudo- $F = 21.346$, $P < 0.001$) found a higher variability in microbial community composition than unweighted UniFrac distances ($df = 2$, $R^2 = 0.093$, pseudo- $F = 2.160$, $P < 0.01$) in separating nonpreserved samples from preserved ones, indicating that the sample clustering was driven more by the proportion of microbial community members than the presence/absence of bacterial taxa. For example, a significant increase of the relative abundance of *Enterobacteriales* (Fig. 2D) and *Fusobacteriales* (Fig. 2E) and a reduction of *Clostridiales* (Fig. 2F) and *Betaproteobacteriales* (Fig. 2G) were observed in nonpreserved gut samples (Table S2).

Variation in gut microbial communities stored in preservatives. Next, we refined the ASV table by excluding samples in PBS solution and Cary-Blair transport medium to measure the variation of six preservatives in retaining gut microbiota composition with reference to immediate freezing at -80°C as standards. Overall, the microbial community of the surveyed stool samples was mainly dominated by *Faecalibacterium* (mean relative abundance of 23.4%) and *Bacteroides* (22.2%), followed by *Roseburia* (7.3%) (Table S3). The alpha diversity analyses at either ASV or genus level did not differ significantly between -80°C standards and preserved samples (Fig. 3A and B), except that aliquots stored in Shield appeared relatively lower in Shannon diversity (mean of 3.0 and 2.2, respectively) and Simpson evenness (mean of 0.9 and 0.8, respectively). We found significant changes of the alpha diversities of gut microbial community across subjects (Kruskal-Wallis test, $P < 0.029$) (Fig. S1).

A PERMANOVA using weighted UniFrac distances indicated that both subject groups ($df = 4$, $R^2 = 0.632$, pseudo- $F = 26.134$, $P < 0.001$) and preservation conditions ($df = 6$, $R^2 = 0.223$, pseudo- $F = 6.157$, $P < 0.001$) contributed to the differences in microbiota composition, although the majority of variability was from differences between subjects rather than the lower impact of storage conditions (63.2% versus 22.3%). The variation was statistically attributed to subjects ($df = 4$, $R^2 = 0.779$, pseudo- $F = 28.034$, $P < 0.001$) but not preservatives ($df = 6$, $R^2 = 0.055$, pseudo- $F = 1.309$, $P = 0.118$) based on unweighted UniFrac distances. In line with these observations, the principal-coordinate analysis supported that microbial variation was mainly associated with differences between individuals (Fig. 3C and D). In order to examine the small effects of variation by preservation methods, we subsequently calculated the pairwise distances between -80°C standards and preserved samples collected from the same subjects using weighted UniFrac distances, since unweighted UniFrac distances did not reveal statistical significance ($P = 0.997$) (Fig. 3C). As shown in Fig. 3D (weighted), samples stored in various preservatives exhibited diversified microbial community shifts relative to -80°C standards (intrasubject distance to standard of 0.06 to 0.26) ($P = 0.027$), suggesting that all preservation methods contributed impacts on the variation in microbiota profiling. We found, however, that two preservatives (Norgen and OMNI) showed the least shift in community composition (mean distance to the standards of 0.09 and 0.09, respectively) while RNAlater had the highest impact (mean

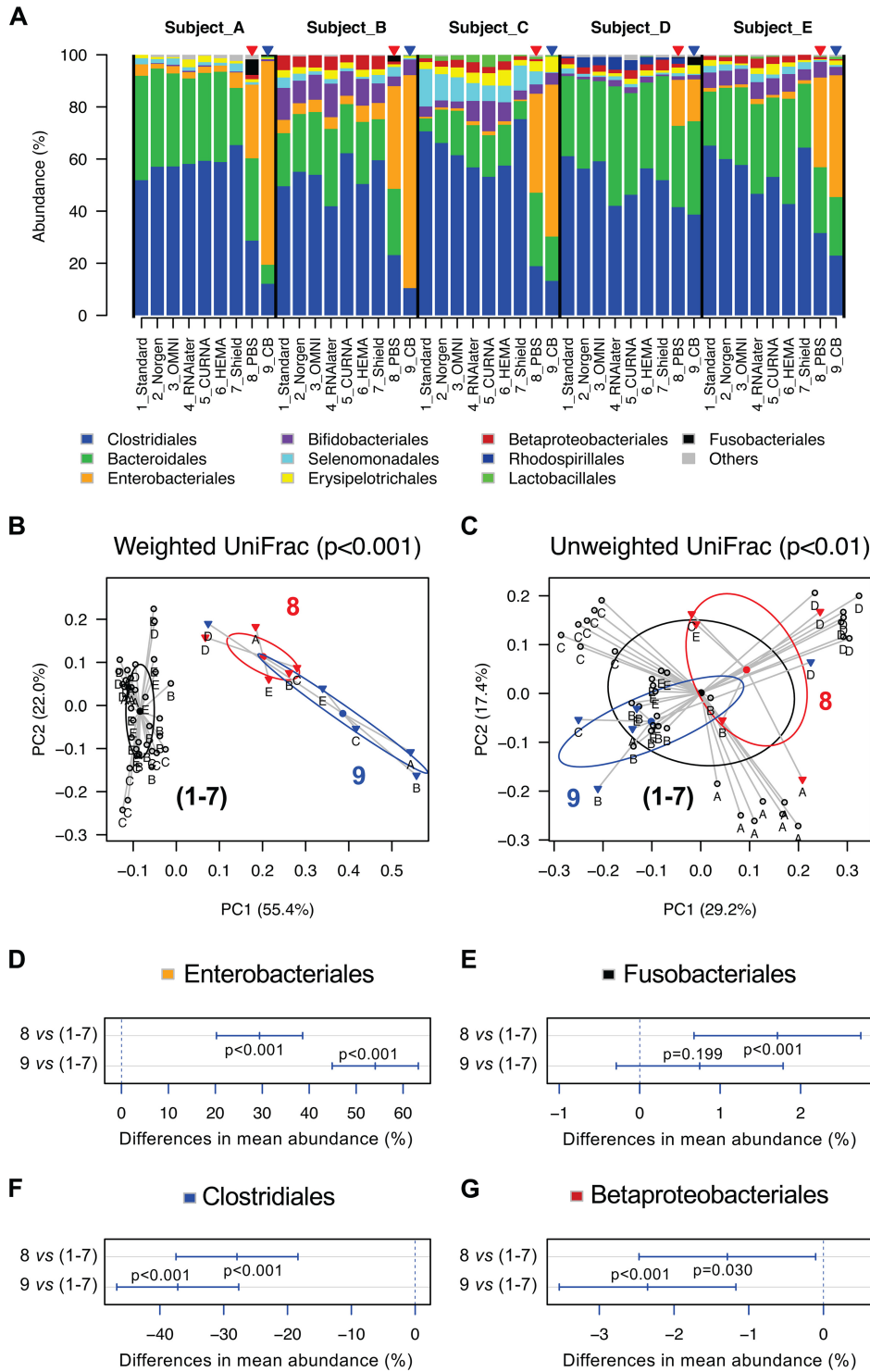


FIG 2 Community shift in nonpreserved stool samples in PBS or Cary-Blair transport medium observed by 16S rRNA gene V3-V4 amplicon sequencing. (A) Bar chart showing the proportion of the bacterial order. Filled triangles in red and blue represent nonpreserved samples in PBS (samples 8) and Cary-Blair transport medium (samples 9), respectively. The bacterial orders grouped into “Others” include *Coriobacteriales*, *Desulfovibrionales*, *Pasteurellales*, *Synergistales*, *Victivallales*, and *Bacillales*. (B) Principal-coordinate plot of weighted UniFrac distances discriminated nonpreserved stool samples from other aliquots in different preservatives (Table 1 gives detailed preservation methods for samples 1 to 7). (C) Principal-coordinate plot of unweighted UniFrac distances discriminated nonpreserved stool samples from other aliquots in different preservatives. (D to G) A pairwise Tukey *post hoc* test shows an overgrowth of *Enterobacteriales* and *Fusobacteriales* and a reduction of relative abundance of *Clostridiales* and *Betaproteobacteriales* in nonpreserved stool samples.

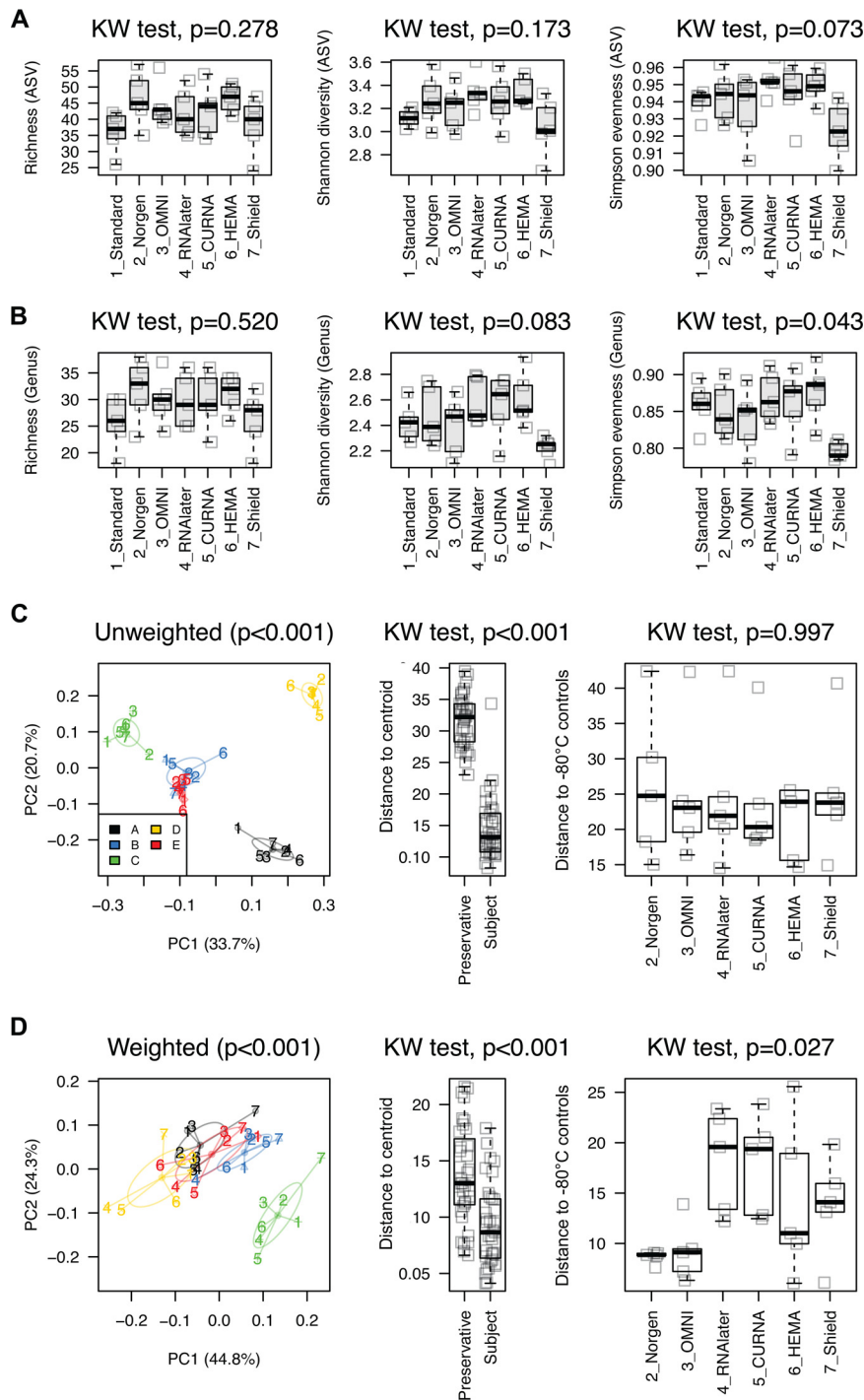


FIG 3 Comparison of gut microbiota alpha diversity in different preservatives based on 16S rRNA gene V3-V4 amplicon sequencing. Alpha diversity analysis on amplicon sequence variant (ASV) (A) and genus (B) levels. Principal-coordinate plot using unweighted (C) and weighted (D) UniFrac distances, with sample clustering by individual subjects. The PC1 and PC2 axes represent the first two principal coordinates. KW test, Kruskal-Wallis test.

distance to the standards of 0.18). Consistent with the pairwise distance comparison, the relative abundance of *Bacteroides* and *Ruminococcaceae* UCG-002 increased significantly in samples stored in RNAlater compared to -80°C standards (Tukey honest significant difference [HSD] *post hoc* test, $P = 0.0499$ and $P = 0.0408$, respectively) (Table S3). Similarly, an increase of the relative abundance of *Faecalibacterium* from

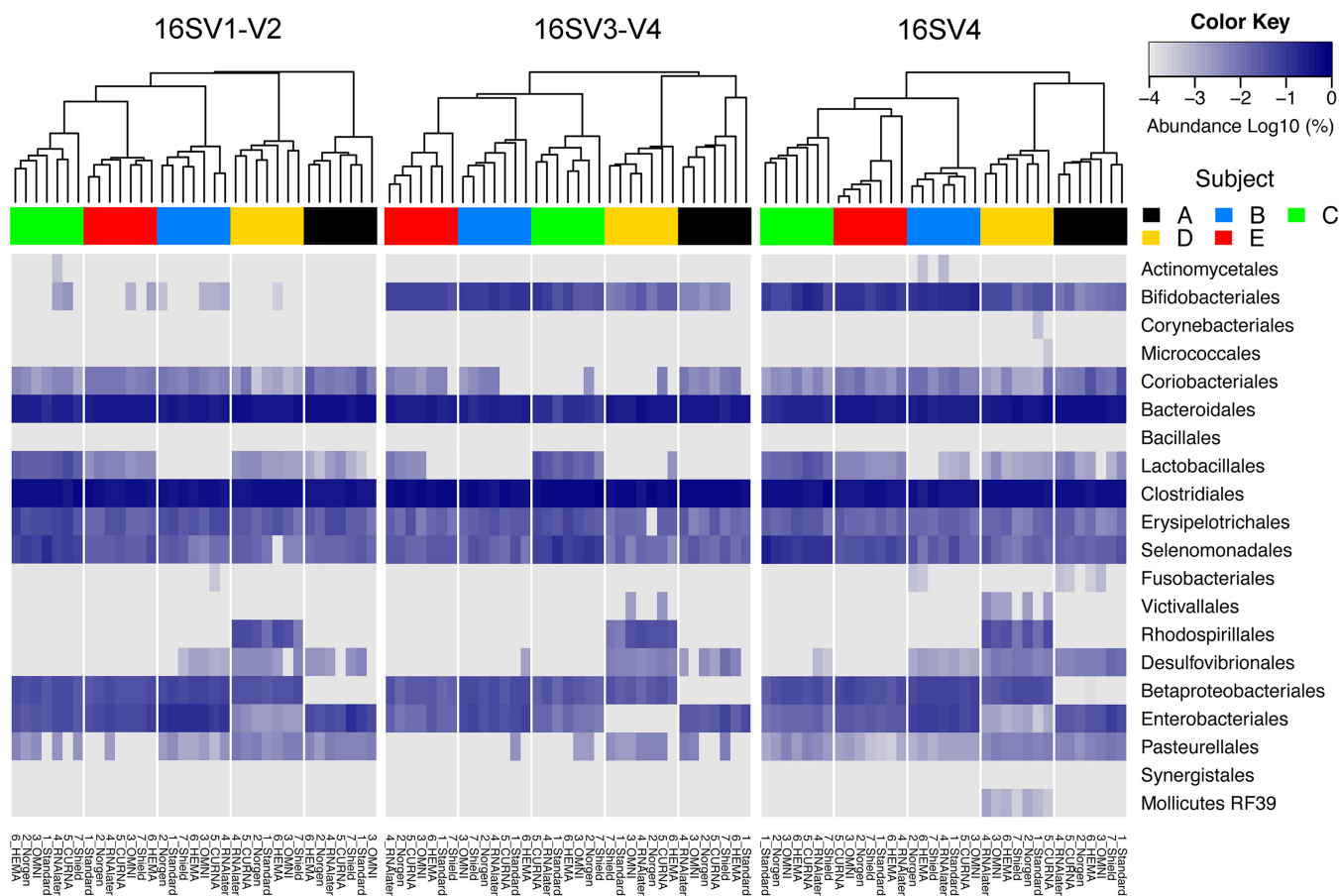


FIG 4 Heat map comparing the proportional abundances of gut microbiota summarized at the order level between 16S rRNA gene primer pairs. Each column represents individuals with the proportional abundance on a base 10 logarithmic scale, with 0 (a log₁₀ of 100%) in black and -4 (a log₁₀ of 0.01%) in light silver. Each row represents the bacterial order. The clustering of subjects at the top is based on the weighted GUniFrac distance.

samples stored in Shield ($P = 0.0239$) and a decrease of the relative abundance of *Lachnospiraceae* (*Eubacterium eligens* group) in OMNI ($P = 0.0461$) relative to -80°C standards were observed.

Impact of targeting different regions of the 16S rRNA gene in profiling gut microbiota. Previous work has suggested that targeting different variable regions of the 16S rRNA gene may generate different microbial community profiles (20). In this study, we compared the different performances of primer pairs targeting three hypervariable regions of bacterial 16S rRNA gene (V1-V2, V3-V4, and V4) in profiling gut microbiota composition. Overall, 254,363 and 420,622 high-quality sequence reads (mean \pm SD of $7,267 \pm 1,139$ and $12,017 \pm 3,799$) were obtained with V1-V2 and V4 primer pairs, respectively. When analyses on the divergences of preservation methods in revealing gut microbial community were repeated, the sequences generated by V1-V2 and V4 primer pairs (Fig. 4; Fig. S2 and S3) showed similar results as that of V3-V4 reads. The surveyed stool samples mainly clustered by subjects. Although variations in microbiota composition were observed for samples in all preservatives tested, the effects were relatively small (PERMANOVAs using weighted UniFrac distances: V1-V2, $R^2 = 0.163$, $P < 0.001$; V4, $R^2 = 0.167$, $P < 0.001$) compared to the high heterogeneity between individuals (V1-V2, $R^2 = 0.709$, $P < 0.001$; V4, $R^2 = 0.709$, $P < 0.001$). Samples in Norgen preservative consistently showed the least shift in community composition relative to -80°C standards. Meantime, the relative abundance of *Faecalibacterium* in samples stored in Shield was significantly higher, as revealed by all three 16S rRNA gene region primer pairs (Table S3).

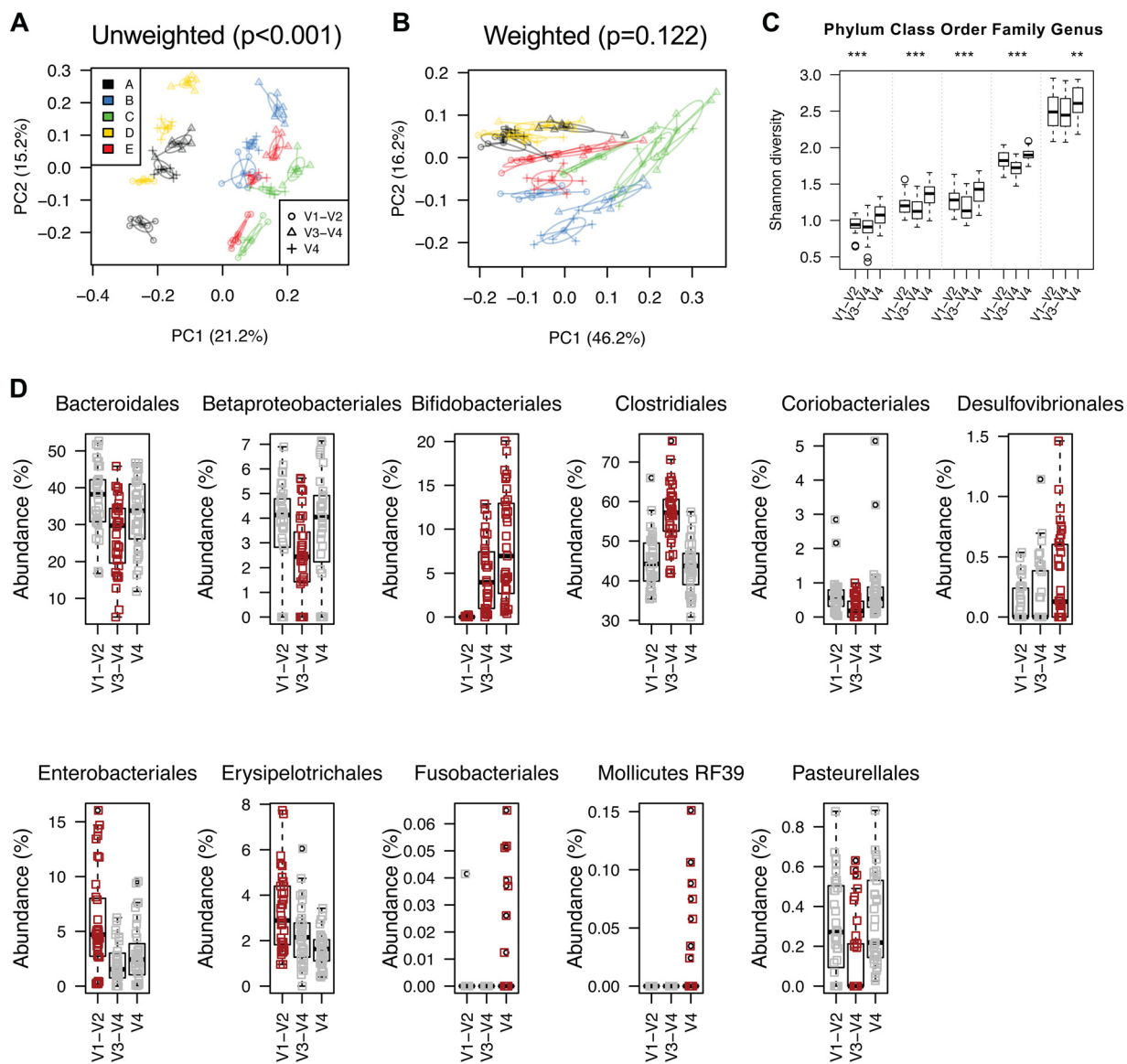


FIG 5 (A and B) Unweighted (A) and weighted (B) UniFrac clustering of three 16S rRNA hypervariable regions using fragment insertion SATé-enabled phylogenetic placement (SEPP) technique. (C) Box plot of the Shannon diversity indices between 16S rRNA V1-V2, V3-V4, and V4 regions. Statistical significance was calculated using Kruskal-Wallis test. (D) Relative abundances (at the order level) of gut microbiota showing statistically significant differences between 16S rRNA V1-V2, V3-V4, and V4 regions. Taxa in red show statistical significance according to nonparametric Wilcoxon signed-rank test ($P < 0.05$) (see details in Table S4).

However, the choice of primers targeting different hypervariable regions of the bacterial 16S rRNA gene may specifically impact microbial community profile and bacterial taxon assessment. When all three region reads were combined for principal-coordinate analyses and PERMANOVAs using fragment insertion SATé-enabled phylogenetic placement (SEPP) technique, we found strong bias of gene regions in sample clustering based on either unweighted ($df = 2$, $R^2 = 0.218$, $\text{pseudo-}F = 42.239$, $P < 0.001$) or weighted ($df = 2$, $R^2 = 0.197$, $\text{pseudo-}F = 51.940$, $P < 0.001$) UniFrac distances, although the majority of variability was consistently from differences between subjects (unweighted, $df = 4$, $R^2 = 0.527$, $\text{pseudo-}F = 51.073$, $P < 0.001$; weighted, $df = 4$, $R^2 = 0.507$, $\text{pseudo-}F = 66.863$, $P < 0.001$) (Fig. 5A and B). The impact of preservation methods on microbial community shifts was relatively small (unweighted, $df = 6$, $R^2 = 0.018$, $\text{pseudo-}F = 1.156$, $P = 0.198$; weighted, $df = 6$, $R^2 = 0.122$, $\text{pseudo-}F = 10.707$, $P < 0.001$). We also found that the V4 prime pair showed a higher Shannon

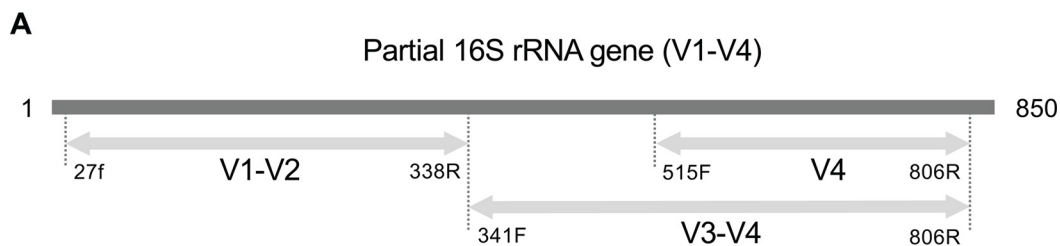
diversity for gut microbiota profile than did other primer sets (Fig. 5C), in line with the higher levels of alpha diversity with rarefied V4 region reads (at genus level and above) (Fig. S4). The relative abundance of 11 bacterial orders was significantly variable among reads generated by different primers (Fig. 5D; Table S4). For example, the commonly used V1-V2 primer pair failed to detect the majority of *Bifidobacteriales* (mean relative abundance of 0.03% versus 6.40%, $P < 0.001$) but revealed more *Enterobacteriales* (5.69% versus 2.57%, $P < 0.001$) and *Erysipelotrichales* (3.2% versus 2.0%, $P < 0.001$) compared to another two primer sets. The V3-V4 primers showed the highest abundance of *Clostridiales* (56.69% versus 44.12%, $P < 0.001$), which likely led to the apparent decrease of *Bacteroidales* (26.91% versus 34.97%, $P < 0.001$), *Betaproteobacteriales* (2.45% versus 3.58%, $P < 0.001$), *Coriobacteriales* (0.26% versus 0.71%, $P < 0.001$), and *Pasteurellales* (0.13% versus 0.31%, $P < 0.001$). *Mollicutes* RF39 was amplified only by the V4 primer set, although the relative abundance of this bacterial group was extremely low in the surveyed stool samples.

Sequence variation of primer-binding sites. In order to characterize the occurrences of sequence variations at the primer-binding sites used in this study, we refined an NCBI bacterial 16S rRNA reference database (a total of 18,773 complete or near-complete sequences) using a threshold of 99% similarity and extracted sequence sets matching each primer-binding site. The most common sequence variations are shown in Fig. 6, with the corresponding dominant bacterial group(s) in which the sequence variations are observed. It is noted that the V1-V2 forward primer-binding site sequences in most *Bifidobacteriales* and *Chlamydiales* 16S rRNA genes were not precisely accommodated by the commonly used degenerate 27f primers, differing at two and three positions from the best-matching 27f-YM primer, respectively. Similarly, the nondegenerate form of the V1-V2 reverse primer used in this study (338R) may have low sensitivity in detecting *Planctomycetales*, for which three mispairings were identified. The V1-V2 reverse primer (338R) and V3-V4 forward primer (341F) showed mismatches with *Mollicutes* RF39 16S rRNA genes in two positions, which could confer a low efficiency in detecting this group of bacteria (Fig. 5D). Optimization of primer pools to minimize the mispairing at primer-binding sites is proposed (Fig. 6B).

Bacterial culture of stool samples in different preservatives. To test whether the surveyed preservatives are able to render microbes noninfectious, stool suspensions preserved in each solution over time were cultured for bacteria under both aerobic and anaerobic conditions (Fig. S5). We found that Norgen, OMNI, HEMA, and Shield preservatives immediately and completely prevented the growth of gut bacteria, allowing the resulting samples to be handled safely. In contrast, RNAlater and CURNA, two preservation solutions designed for RNA samples, as well as the PBS and Cary-Blair transport medium, showed no effect in inactivating bacteria.

DISCUSSION

Microbiota studies have been quickly switched to large-scale sample collection to characterize the subtle differences in microbial communities in relation to human health and diseases. Although immediate freezing at -20°C or below has been considered the “gold standard” for microbiota preservation, this approach is not feasible for many studies, especially field sampling from remote areas or self-collected samples requiring shipment to the laboratory without cold chain transport. In such cases, a proper preservation method to minimize microbiota composition shift to reduce bias and to inactivate infectious agents for samples to be handled safely becomes critical. Hence, we evaluated the effects of multiple preservation solutions and 16S rRNA gene hypervariable regions in revealing gut microbiota composition. Our data showed that the profile of gut microbiota can change dramatically over a course of 7 days at room temperature in the absence of protection of preservation. The use of a proper preservative without cold chain, in contrast, provides satisfactory prevention of community shifts. The effects of these preservatives were small compared to intersubject variations; however, notable changes in microbiota composition were observed. Nevertheless, we found that two preservatives, Norgen and OMNI, showed



B

	Primer binding (5'-3')	Occurrences	Bacteria group(s)	
V1-V2 Forward	AGAGTTTGATYMTGGCTCAG		27f-YM	
	AGGGTTCGATCTGGCTCAG		27f-Bif	
	AGAATTGATCTTGGTTCAG		27f-Chl	
	AGAGTTTGATCCTGGCTCAG	2957	many Bacteria	
A.....	540	many Bacteria	
	T.....	116	some Bacteria	
T.....	53	some Bacteria	
	..G.....	40	some Thermotogales and Planctomycetales	
	..G...C...T.....	23	Bifidobacteriales	
TA.....	17	some Proteobacteria	
Variation	T.....T.....	14	some Streptomycetales	
	..A.....T...T...	13	Chlamydiales	
C.....	11	some Clostridiales	
	T.....A.....	8	many Bacteria	
	V1-V2 Reverse	ACTCCTACGGGAGGCAGCA		338R
		ACACCTACGGGTGGCWGCA		338R-ex1
		ACTCCTRCGGGRGGCWGCA		338R-ex2
		ACTTCTACGGAAGGCAGCA		338R-ex3
	V3-V4 Forward	CCTACGGGNGGCWGCAG		341F
		CCTGCGGGAGGCAGCAG		341F-ex1
TCTACGGAAGGCAGCAG			341F-ex2	
ACTCCTACGGGAGGCAGCAG		13133	many Bacteria	
VariationG.....	81	Lentisphaerales	
	..A.....T.....	68	Anaerolineales and Verrucomicrobia	
T.....	62	Halobacteroidaceae and many Leuconostocaceae	
G.....	38	Ktedonobacteria and Caldilineae	
	..T.....A.....	29	some Clostridiales and Mollicutes	
	..A.....T...T...	23	Planctomycetales	
	V4 Forward	GTGYCAGCMGCCGCGGTA		515F
GTGCCAGCAGYYGCGGTA			515F-ex1	
GTGCCAGCAGCCGCGGTG			515F-ex2	
GTGCCAGCAGCCGCGGTA		13436	many Bacteria	
VariationG.....	87	some Actinobacteria and Chlorobiales	
T.....	20	Chlamydia	
T.....	10	Mycoplasmatales	
C.....	8	Selenomonadales	
V3-V4 / V4 Reverse	ATTAGAWACCCBNGTAGTCC		806RB	
	ATTAGATACCCBGGTAGTCC		806R-ex1	
	ATTAGATACCCYDGTAGTCC		806R-ex2	
	ATTAGAKACCCGGTARTCC		806R-ex3	
	ATTAGATACCCCTGGTAGTCC	12468	many Bacteria	
C.....	378	many Bacteria	
G.....	211	many Bacteria	
	C.....	188	Propionibacteriaceae and many Microbacteriaceae	
A.....	92	many Bacteria	
T.....	84	many Bacteria	
VariationCT.....	74	some Actinobacteria and Verrucomicrobiales	
CA.....	58	many Cyanobacteria	
C...A...	25	Spirochaetia	
C.....T	17	some Flavobacteriales	
G...C.....	14	Anaerolineales	
T...T...	5	Lentisphaerales	
A...T...	5	Entomoplasmatales	
CT...T...	5	some Cyanobacteria	

FIG 6 Sequence variation of 16S rRNA gene primer-binding sites. (A) Schematic of the 16S rRNA gene regions and primer pairs. (B) Occurrences of the most commonly observed sequences of primer-binding sites and the primer formulations that they may match. Nucleotide sites identical to the most common consensus sequence (the first listed) are represented as dots. The main nucleotide changes of primer sequences from the most common consensus sequence are highlighted in red.

the smallest microbial community shifts, and both efficiently inhibited the growth of aerobic and anaerobic bacteria, suggesting they may serve as useful solutions for stool sample transportation and storage. Recent studies have reported that OMNIgene.GUT (here called OMNI) is able to stabilize DNA at room temperature for up to 14 days, with little difference in microbiota composition by 16S rRNA gene sequencing (13, 15, 19, 21). RNAlater, a commonly used preservative for RNA samples, on the other hand, exhibited larger community shift in bacterial relative abundance at room temperature for 7 days, consistent with previous reports that the microbial community in RNAlater may lose stability at longer time frames when maintained at ambient temperatures (12, 15, 16, 22). However, under cold conditions (e.g., 4°C or –20°C) or for a short period of room-temperature storage (e.g., within same or another day), it performed equally well as other storage conditions, such as 95% ethanol, FTA cards, and OMNIgene.GUT (21, 23, 24). Interestingly, RNAlater and CURNA (with similar components as RNAlater) do not prevent bacterial activity, highlighting themselves as potential alternatives for microbial sample collection when live bacterial culture is necessary. According to the manufacturer's instructions, HEMA was initially developed for the stabilization of buffy coat samples and there was no previous report evaluating its use as preservative for samples intended for microbiota studies. Shield performed well in microbiota preservation and disinfection but consistently showed increased relative abundance of *Faecalibacterium* regardless of the choice of 16S rRNA gene primer pairs. We recommend the use of Norgen or OMNI for microbial sample preservation and caution against RNAlater for long-term storage at ambient temperatures.

Choosing primers to target certain regions of the 16S rRNA gene for bacterial taxonomic profiling is another difficult challenge, and the comparability of primers with microbes from various communities requires attention (25, 26). The V1 to V4 region, for example, is more divergent than V4 to V7 and can provide a higher resolution in bacterial clustering for stool samples (20, 27). In this work, we compared the performance of three commonly used primer pairs targeting the 16S rRNA gene V1-V2, V3-V4, and V4 regions in profiling gut microbial community and found a higher alpha diversity and richness revealed by V4 primers. Our results also indicate an immense impact of primers with regard to amplification of bacterial taxa. Close examination of primer-binding sites revealed several sequence variations within cohesive phylogenetic groups (e.g., *Bifidobacteriales*) that are not accommodated by the commonly used 27f primer sequence. In such cases, the use of pooled primers to minimize mismatches with this bacterial group will provide an effective approach for revealing the true bacterial diversity (26, 28, 29). Knowledge of how primer pairs differentially amplify bacterial taxa is important for study design and relating results to prior studies. Alternatively, the quickly developing next-generation sequencing platforms, such as PacBio and Nanopore, may facilitate the characterization of full-length bacterial 16S rRNA genes and allow global comparisons of microbiome studies across sample types and gene regions (30–32).

This study has its strengths and limitations. We expand the current understanding of the performance of several DNA and RNA preservation solutions, including CURNA and HEMA, to our knowledge being evaluated for the first time, in retaining gut microbiota composition. The data provide a useful reference for knowing the biases introduced by storage conditions. We have expanded the knowledge of bactericidal ability of the surveyed preservatives that may provide alternatives in study design considering the potential of infectious agents and bacterial culture. We have also identified the significant difference of 16S rRNA gene primer pairs in characterizing bacterial taxa and emphasized the importance of a consistent methodology for large epidemiological studies. Nevertheless, one of the shortcomings of this study is the small sample size and the conserved sample source (5 subjects who were all male, young, and healthy), which may underestimate the impact of preservation solutions and primer pairs in characterizing the diversity of gut microbiota. Further methodological comparison is needed to expand to other settings, including patient groups and population-based epidemiological studies. Only sequencing-based evaluation was per-

formed in this work, while the concordance of preservation solutions in other omics studies, such as proteomics and metabolomics, may be largely variable. For example, RNAlater was reported to be not feasible for metabolomics measures probably due to its high sodium sulfate content, making the preserved samples incompatible with metabolomics platforms (22, 33). Last, more robust qualitative and quantitative testing under variable conditions, such as extreme temperatures, increased freeze-thaw cycles, or longer storage periods, may provide a more comprehensive evaluation of the effectiveness and stabilization of storage methods at preserving microbial DNA quality and yield.

Conclusions. It is important for large-scale microbiome studies to accurately and consistently reveal microbial communities while minimizing external effects. Choices of preservation solution for transportation of stool samples at ambient temperature and of gene region for 16S rRNA gene amplicon sequencing are critical determinants impacting the observed microbial community profiling. We recommend the use of a proper preservative, such as Norgen or OMNI, that can minimize microbial community shift and efficiently inactivate bacterial growth, allowing the resulting samples to be handled and shipped safely and stably. Amplicon sequencing-based human microbiota studies are highly dependent on choice of primer pairs targeting different hypervariable regions of the bacterial 16S rRNA gene. Nevertheless, a consistent methodology is essential in minimizing batch biases to facilitate comparisons between data sets.

MATERIALS AND METHODS

Ethics approval. This study was approved by the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee. All subjects recruited in this study were older than 18 years of age, and samples were anonymized without individual identifying information. Written informed consent was obtained from each participant.

Stool specimen collection, processing, and DNA extraction. Five healthy male subjects, aged between 23 and 28 years old, were recruited to provide fresh stool samples using a stool self-collection kit (see Table S1 in the supplemental material). Stools from these five donors were collected on different days and were transported to the laboratory in an icebox cooler at 4°C within 2 h without a freeze-thaw cycle. Upon receipt, approximately 5 g of sample was immediately mixed with 200 μ l of sterile phosphate-buffered saline (PBS) and thoroughly homogenized for 1 min using a sterile cotton stick. One gram of the homogenized stool, serving as a standard, was transferred into a 2-ml sterile collection tube and immediately stored at -80°C until further DNA extraction. In parallel, eight stool aliquots, each weighing approximately 0.2 g, were transferred into 2-ml collection tubes preloaded with 1 ml of preservative (Norgen, OMNI, RNAlater, CURNA, HEMA, and Shield), PBS, or Cary-Blair transport medium, which is intended for preserving the viability of enteric bacteria (Table 1). Tubes were pulse vortexed for 5 s and placed on a laboratory bench at room temperature for 7 days. Then, stool aliquots were stored at -80°C without a freeze-thaw cycle for 28 to 73 days until DNA extraction (Fig. 1). The procedures of stool homogenization and aliquoting were performed in a biosafety cabinet and finished within 30 min.

DNA extraction was performed in accordance with the Earth Microbiome Project protocol (34), with the bead-beating QIAamp PowerFecal DNA kit (Qiagen, USA). In detail, 100 μ l of stool suspension from aliquots stored at -80°C was added to the PowerBead tubes for total DNA extraction according to the manufacturer's protocol. An MP FastPrep-24 5G homogenizer, at a speed of 6.0 m/s for 45 s, was used for mechanical lysis. Approximately 0.2 g of stool standards was suspended in 1 ml of sterile PBS prior to DNA extraction. The DNA samples were eluted in 100 μ l elution buffer (pH 8.0) and stored at -20°C . DNA concentrations were measured using a NanoDrop 2000 (Thermo Scientific, USA).

16S rRNA gene PCR amplification and sequencing. We used three PCR primer pairs targeting the 16S rRNA gene V1-V2, V3-V4, and V4 regions, respectively, to profile gut microbiota compositions (Table 2) (35). A pair of dual 12-bp barcodes was indexed to each amplicon set through the forward and reverse primers modified from the Earth Microbiome Project protocol (34). Successful amplicons were pooled at approximately equal molar DNA concentrations, purified using the QIAquick gel extraction kit (Qiagen, USA), and sequenced on an Illumina MiSeq (Illumina, USA) at the Weill Cornell Medicine Genomics Resources Core Facility, New York, NY, USA, using paired-end 300-bp reads.

Bioinformatics and biostatistics analysis. Following demultiplexing, the QIIME2 (2018.8) package (<https://qiime2.org>) (36) was applied to assign Illumina short reads into amplicon sequence variant (ASV) and bacterial taxa tables. The detailed pipelines, including quality control, paired-end short reads merging, and dada2 denoising and clustering, can be found in the supplemental pipeline (Text S1). In order to combine reads from different 16S rRNA hypervariable regions for principal coordinates analysis, the fragment insertion SATé-enabled phylogenetic placement (SEPP) technique with default parameters was used (37). The SILVA 132 99% 16S rRNA gene reference database (<https://www.arb-silva.de/download/archive/qiime>) was used to assign bacterial taxonomic classification (38). Singleton reads were removed for statistical analysis. Compositions of microbiota communities were summarized by proportion at different taxonomy levels, including genus, family, order, class, and phylum ranks.

In order to describe the diversity of observed bacterial taxa taking into consideration richness and abundance, Shannon and Simpson diversity indexes (39) were calculated. We assessed the alpha diversity using species richness and Shannon diversity at various rarefaction depths ranging from 1 to 6,001 reads (with steps of 250 reads) and found no significant difference of alpha diversity using rarefied data sets at depth of 1,000 reads per sample or higher (Fig. S4). In order to retain all samples for diversity analysis, we set 1,500 reads per sample as rarefaction depth to normalize the data for differences in sequence count. We used a UniFrac algorithm in the GUniFrac R package (40), with unweight or weight ($\alpha = 1$) on abundant lineages to calculate pairwise Kantorovich-Rubinstein (KR) distances between samples. Differences in community composition were assessed using permutational multivariate analysis of variance (PERMANOVA) in the Vegan R package. Principal-coordinate analysis was performed to visualize associations between community composition and experimental factors. Comparisons of the relative abundances of characteristic ASVs between defined groups were performed using nonparametric Mann-Whitney Wilcoxon rank sum test (MWU), Wilcoxon signed rank test (WSR), Kruskal-Wallis test (KW), or Tukey's honest significant difference (Tukey HSD) *post hoc* test. Statistical analyses and plotting were performed in R (3.4.0). A two-sided *P* value of <0.05 was considered statistically significant.

Bacterial aerobic and anaerobic culture. Approximately 0.2 g of homogenized fresh stool samples was transferred into 2-ml collection tubes preloaded with 1 ml of preservative (Norgen, OMNI, RNAlater, CURNA, HEMA, and Shield), PBS, or Cary-Blair transport medium. For each sample, 100- μ l replicate suspensions were incubated at room temperature for 0 min, 10 min, 1 h, 5 h, 24 h, 48 h, and 96 h, respectively. Approximately 100 μ l of 1:10-diluted stool suspensions in 0.85% saline solution was spread on a blood agar plate and a K1 agar plate for aerobic and anaerobic culture, respectively. Plates were incubated at 37°C for 48 h in a duplicate setting, and the number of CFU was counted immediately afterwards.

Data availability. All sequence data generated from this study were deposited in the Sequence Read Archive with accession number [PRJNA470603](https://doi.org/10.1128/PRJNA470603).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00271-18>.

TEXT S1, PDF file, 0.1 MB.

FIG S1, TIF file, 0.6 MB.

FIG S2, TIF file, 1.3 MB.

FIG S3, TIF file, 1.3 MB.

FIG S4, TIF file, 1.3 MB.

FIG S5, TIF file, 0.9 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.01 MB.

TABLE S3, XLSX file, 0.2 MB.

TABLE S4, XLSX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank the anonymous participants providing stool samples for this study.

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. CUHK 14161017) (Z.C.), the Direct Grant (project no. CUHK 4054350) (Z.C.), and the Seed Fund for Gut Microbiota Research (P.K.S.C.) by the Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China. The funder had no role in study design, data collection, analysis, interpretation, or writing of the report.

Author contributions were as follows: study design, Z.C., S.C.N., F.K.L.C., and P.K.S.C.; conduct of experiment, P.C.H. and P.Y.W.; data interpretation and analysis, M.H., Z.C., Y.K.Y., and M.C.W.C.; manuscript writing, all authors.

We declare that we have no competing interests.

REFERENCES

1. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486:207–214. <https://doi.org/10.1038/nature11234>.
2. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, Brady A, Creasy HH, McCracken C, Giglio MG, McDonald D, Franzosa EA, Knight R, White O, Huttenhower C. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature* 550:61–66. <https://doi.org/10.1038/nature23889>.
3. Garrett WS. 2015. Cancer and the microbiota. *Science* 348:80–86. <https://doi.org/10.1126/science.aaa4972>.
4. Jie Z, Xia H, Zhong SL, Feng Q, Li S, Liang S, Zhong H, Liu Z, Gao Y, Zhao H, Zhang D, Su Z, Fang Z, Lan Z, Li J, Xiao L, Li J, Li R, Li X, Li F, Ren H,

- Huang Y, Peng Y, Li G, Wen B, Dong B, Chen JY, Geng QS, Zhang ZW, Yang H, Wang J, Wang J, Zhang X, Madsen L, Brix S, Ning G, Xu X, Liu X, Hou Y, Jia H, He K, Kristiansen K. 2017. The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun* 8:845. <https://doi.org/10.1038/s41467-017-00900-1>.
5. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto JM, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490:55–60. <https://doi.org/10.1038/nature11450>.
 6. Li J, Zhao F, Wang Y, Chen J, Tao J, Tian G, Wu S, Liu W, Cui Q, Geng B, Zhang W, Weldon R, Auguste K, Yang L, Liu X, Chen L, Yang X, Zhu B, Cai J. 2017. Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* 5:14. <https://doi.org/10.1186/s40168-016-0222-x>.
 7. Pickard JM, Zeng MY, Caruso R, Nunez G. 2017. Gut microbiota: role in pathogen colonization, immune responses, and inflammatory disease. *Immunol Rev* 279:70–89. <https://doi.org/10.1111/imr.12567>.
 8. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henriksen B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JL. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341:1241214. <https://doi.org/10.1126/science.1241214>.
 9. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9:811–814. <https://doi.org/10.1038/nmeth.2066>.
 10. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108(Suppl 1):4516–4522. <https://doi.org/10.1073/pnas.1000080107>.
 11. Fraher MH, O'Toole PW, Quigley EM. 2012. Techniques used to characterize the gut microbiota: a guide for the clinician. *Nat Rev Gastroenterol Hepatol* 9:312–322. <https://doi.org/10.1038/nrgastro.2012.44>.
 12. Gorzelak MA, Gill SK, Tasnim N, Ahmadi-Vand Z, Jay M, Gibson DL. 2015. Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. *PLoS One* 10:e0134802. <https://doi.org/10.1371/journal.pone.0134802>.
 13. Choo JM, Leong LE, Rogers GB. 2015. Sample storage conditions significantly influence faecal microbiome profiles. *Sci Rep* 5:16350. <https://doi.org/10.1038/srep16350>.
 14. Tedjo DI, Jonkers DM, Savelkoul PH, Masclee AA, van Best N, Pierik MJ, Penders J. 2015. The effect of sampling and storage on the fecal microbiota composition in healthy and diseased subjects. *PLoS One* 10:e0126685. <https://doi.org/10.1371/journal.pone.0126685>.
 15. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, Knight R. 2016. Preservation methods differ in fecal microbiome stability affecting suitability for field studies. *mSystems* 1:e00021-16. <https://doi.org/10.1128/mSystems.00021-16>.
 16. Hale VL, Tan CL, Knight R, Amato KR. 2015. Effect of preservation method on spider monkey (*Ateles geoffroyi*) fecal microbiota over 8 weeks. *J Microbiol Methods* 113:16–26. <https://doi.org/10.1016/j.mimet.2015.03.021>.
 17. Dominianni C, Wu J, Hayes RB, Ahn J. 2014. Comparison of methods for fecal microbiome biospecimen collection. *BMC Microbiol* 14:103. <https://doi.org/10.1186/1471-2180-14-103>.
 18. Reidmiller JS, Smith WL, Sawyer MM, Osburn BI, Stott JL, Cullor JS. 2006. Antimicrobial properties of the chelating agent EDTA on streptococcal bovine mastitis isolates. *J Food Prot* 69:1460–1462. <https://doi.org/10.4315/0362-028X-69.6.1460>.
 19. Penington JS, Penno MAS, Ngui KM, Ajami NJ, Roth-Schulze AJ, Wilcox SA, Bandala-Sanchez E, Wentworth JM, Barry SC, Brown CY, Couper JJ, Petrosino JF, Pappenfuss AT, Harrison LC, ENDIA Study Group. 2018. Influence of fecal collection conditions and 16S rRNA gene sequencing at two centers on human gut microbiota analysis. *Sci Rep* 8:4386. <https://doi.org/10.1038/s41598-018-22491-7>.
 20. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. 2012. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS One* 7:e39315. <https://doi.org/10.1371/journal.pone.0039315>.
 21. Wang Z, Zolnik CP, Qiu Y, Usyk M, Wang T, Strickler HD, Isasi CR, Kaplan RC, Kurland IJ, Qi Q, Burk RD. 2018. Comparison of fecal collection methods for microbiome and metabolomics studies. *Front Cell Infect Microbiol* 8:301. <https://doi.org/10.3389/fcimb.2018.00301>.
 22. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, Flores R, Sampson J, Knight R, Chia N. 2016. Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol Biomarkers Prev* 25:407–416. <https://doi.org/10.1158/1055-9965.EPI-15-0951>.
 23. Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, Knight R, Chia N, Sinha R. 2017. Comparison of collection methods for fecal samples in microbiome studies. *Am J Epidemiol* 185:115–123. <https://doi.org/10.1093/aje/kww177>.
 24. Blekhnman R, Tang K, Archie EA, Barreiro LB, Johnson ZP, Wilson ME, Kohn J, Yuan ML, Gesquiere L, Grieneisen LE, Tung J. 2016. Common methods for fecal sample storage in field studies yield consistent signatures of individual identity in microbiome sequencing data. *Sci Rep* 6:31519. <https://doi.org/10.1038/srep31519>.
 25. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glockner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1. <https://doi.org/10.1093/nar/gks808>.
 26. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 2015. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* 3:26. <https://doi.org/10.1186/s40168-015-0087-4>.
 27. Kim M, Morrison M, Yu Z. 2011. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Methods* 84:81–87. <https://doi.org/10.1016/j.mimet.2010.10.020>.
 28. Rintala A, Pietila S, Munukka E, Eerola E, Pursiheimo JP, Laiho A, Pekkala S, Huovinen P. 2017. Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. *J Biomol Tech* 28:19–30. <https://doi.org/10.7171/jbt.17-2801-003>.
 29. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. 2008. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl Environ Microbiol* 74:2461–2470. <https://doi.org/10.1128/AEM.02272-07>.
 30. Wagner J, Coupland P, Browne HP, Lawley TD, Francis SC, Parkhill J. 2016. Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC Microbiol* 16:274. <https://doi.org/10.1186/s12866-016-0891-4>.
 31. Myer PR, Kim M, Freetly HC, Smith TP. 2016. Evaluation of 16S rRNA amplicon sequencing using two next-generation sequencing technologies for phylogenetic analysis of the rumen bacterial community in steers. *J Microbiol Methods* 127:132–140. <https://doi.org/10.1016/j.mimet.2016.06.004>.
 32. Shin J, Lee S, Go MJ, Lee SY, Kim SC, Lee CH, Cho BK. 2016. Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Sci Rep* 6:29681. <https://doi.org/10.1038/srep29681>.
 33. Loftfield E, Vogtmann E, Sampson JN, Moore SC, Nelson H, Knight R, Chia N, Sinha R. 2016. Comparison of collection methods for fecal samples for discovery metabolomics in epidemiologic studies. *Cancer Epidemiol Biomarkers Prev* 25:1483–1490. <https://doi.org/10.1158/1055-9965.EPI-16-0409>.
 34. Gilbert JA, Jansson JK, Knight R. 2014. The Earth Microbiome project: successes and aspirations. *BMC Biol* 12:69. <https://doi.org/10.1186/s12915-014-0069-1>.
 35. Wang Y, Qian PY. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* 4:e7401. <https://doi.org/10.1371/journal.pone.0007401>.
 36. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciulek T, Kreps J, et al. 2018. QIIME 2: reproducible, interactive, scalable, and

- extensible microbiome data science. PeerJ Preprints 6:e27295v2. <https://doi.org/10.7287/peerj.preprints.27295v2>.
37. Janssen S, McDonald D, Gonzalez A, Navas-Molina JA, Jiang L, Xu ZZ, Winker K, Kado DM, Orwoll E, Manary M, Mirarab S, Knight R. 2018. Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems* 3:e00021-18. <https://doi.org/10.1128/mSystems.00021-18>.
 38. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
 39. Lozupone CA, Knight R. 2008. Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* 32:557–578. <https://doi.org/10.1111/j.1574-6976.2008.00111.x>.
 40. Hamady M, Lozupone C, Knight R. 2010. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 4:17–27. <https://doi.org/10.1038/ismej.2009.97>.