



# Impact of inter- and intra-individual variation, sample storage and sampling fraction on human stool microbial community profiles

Yun Kit Yeoh<sup>1,2</sup>, Zigui Chen<sup>1,2</sup>, Mamie Hui<sup>1,2</sup>, Martin C.S. Wong<sup>2,3</sup>, Wendy C.S. Ho<sup>1</sup>, Miu Ling Chin<sup>1</sup>, Siew C. Ng<sup>2,4,5</sup>, Francis K.L. Chan<sup>2,4</sup> and Paul K.S. Chan<sup>1,2</sup>

<sup>1</sup> Department of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>2</sup> Centre for Gut Microbiota Research, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>3</sup> Jockey Club School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>4</sup> Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>5</sup> State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

## ABSTRACT

Stools are commonly used as proxies for studying human gut microbial communities as sample collection is straightforward, cheap and non-invasive. In large-scale human population surveys, however, sample integrity becomes an issue as it is not logistically feasible for researchers to personally collect stools from every participant. Instead, participants are usually given guidelines on sample packaging and storage, and asked to deliver their stools to a centralised facility. Here, we tested a number of delivery conditions (temperature, duration and addition of preservative medium) and assessed their effects on stool microbial community composition using 16S rRNA gene amplicon sequencing. The largest source of variability in stool community composition was attributable to inter-individual differences regardless of delivery condition. Although the relative effect of delivery condition on community composition was small compared to inter-individual variability (1.6% vs. 60.5%, permutational multivariate analysis of variance [PERMANOVA]) and temporal variation within subjects over 10 weeks (5.2%), shifts in microbial taxa associated with delivery conditions were non-systematic and subject-specific. These findings indicated that it is not possible to model or accurately predict shifts in stool community composition associated with sampling logistics. Based on our findings, we recommend delivery of fresh, preservative-free stool samples to laboratories within 2 hr either at ambient or chilled temperatures to minimise perturbations to microbial community composition. In addition, subsamples from different fractions of the same stool displayed a small (3.3% vs. 72.6% inter-individual variation, PERMANOVA) but significant effect on community composition. Collection of larger sample volumes for homogenisation is recommended.

Submitted 28 August 2018  
Accepted 28 November 2018  
Published 11 January 2019

Corresponding author  
Paul K.S. Chan,  
paulkschan@cuhk.edu.hk

Academic editor  
Lesley Hoyles

Additional Information and  
Declarations can be found on  
page 12

DOI 10.7717/peerj.6172

© Copyright  
2019 Yeoh et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Microbiology, Gastroenterology and Hepatology, Public Health

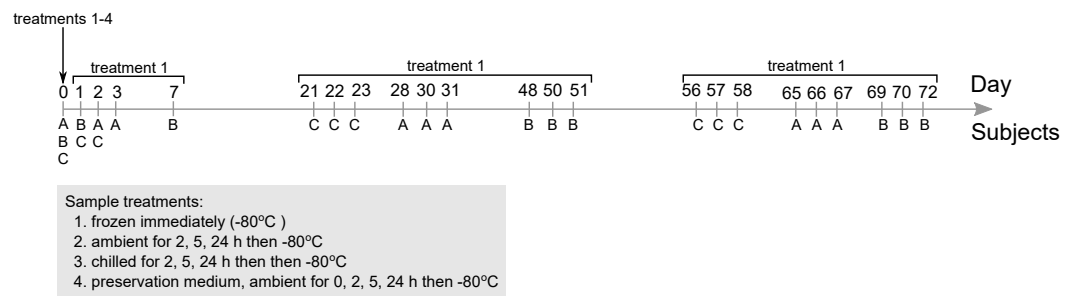
**Keywords** Gut microbiota, Faecal material, Microbial community profiling, 16S rRNA gene, Sample storage, Preservative media, Temperature

## INTRODUCTION

The human gut microbiome has gained unprecedented appreciation for its many vital roles in health and disease. In recent years, studies have linked dysbiosis in gut microbial communities to conditions such as obesity ([Turnbaugh et al., 2006](#)), diabetes ([Qin et al., 2012](#)) and even neurodegenerative diseases such as Parkinson's ([Bedarf et al., 2017](#)) and Alzheimer's disease ([Vogt et al., 2017](#)), underscoring their roles in development, maturation and maintenance of health. Owing to this intimate association between gut microorganisms and their hosts, the gut microbiome has become an attractive target for healthcare practices that aim to restore and promote health by manipulating microbial communities of the gut. This undertaking requires an underlying knowledge of the "healthy" baseline composition and function of gut microbial communities, and how they respond to host factors and environmental perturbations. To that end, various studies have examined the influence of demographics, environment and lifestyle habits over ecology of gut microbial communities in cohorts from various biogeographical backgrounds ([Zeevi et al., 2015](#); [Zhernakova et al., 2016](#); [Falony et al., 2016](#); [Rothschild et al., 2018](#)), and generally agree that environment is a stronger determinant of gut microbiota composition relative to host factors such as ancestry and genetics.

Currently, there are no reference gut microbiota profiles for the South Asian population in Hong Kong. Since environment is commonly the strongest predictor of gut microbiota profiles ([Rothschild et al., 2018](#)), it is necessary to assess whether the gut microbiota of distinct populations are comparable to previously surveyed cohorts and validate whether trends observed from international studies apply to a local population while considering differences in demographics and lifestyles. To address this lack of a local gut microbiome dataset, we are initiating a gut microbial community survey targeted at the general Hong Kong public. One important consideration is that gut microbiome surveys face a challenge of balancing sampling logistics and study power ([Sze & Schloss, 2016](#)) where hundreds to thousands of stool samples are often needed to account for unique population niches and discern potential confounding factors. Because of the large number of participants a population gut microbiome survey would entail, we wanted to formulate simple instructions for participants to deliver stools to a centralised laboratory while also preserving sample integrity as much as possible during transit.

Here, we tested the impact of commonly used sample delivery conditions on stool microbial community profiles. Parameters evaluated included delivery temperature, duration and use of a preservative medium. The impact on microbial community composition was assessed by 16S ribosomal RNA gene (16S) amplicon profiling, and results were interpreted to determine the relative contributions of the evaluated parameters towards compositional differences in community composition. In addition, shifts in the relative abundances of specific microbial taxa associated with the tested parameters were assessed to determine whether perturbations in community composition due to sampling logistics could be generalised across samples and predicted in future sampling efforts.



**Figure 1** Overview of the stool microbial community survey and main experimental factors investigated. Three subjects were recruited to provide stools three times over the span of three to four days (with the exception of subject B at day 0, 1 and 7) thrice over the duration of 72 days. Stools collected at day 0 were subjected to all four treatments to assess their influence on microbial community composition. Community profiles from stools immediately frozen at  $-80^{\circ}\text{C}$  used as references for comparison. Subsequent stools collected from day 1 onwards (sampling time points indicated in figure) were immediately frozen at  $-80^{\circ}\text{C}$  only treatment 1).

Full-size DOI: [10.7717/peerj.6172/fig-1](https://doi.org/10.7717/peerj.6172/fig-1)

## MATERIAL AND METHODS

### Ethics statement and consent

This study has been approved by the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (reference number 2016.707). Written consent was obtained from all participants prior to collecting stool samples.

### Sample collection

Stool samples were collected from three subjects on nine separate occasions. Collection was scheduled in three batches—stools were collected thrice over a period up to seven days to study short-term temporal variation in stool microbial communities, and again three to four weeks and eight to ten weeks after the first collection to study medium-term temporal variability (Fig. 1). Subjects defecated on site and the whole stool was collected in a sterile container. The first stool collected from all three subjects at day 0 were aliquoted and assigned to four treatments intended to simulate how various delivery conditions could impact outcomes of 16S-based gut microbial community surveys: (i) frozen at  $-80^{\circ}\text{C}$  within 5 min of defecation, (ii) ambient temperature for 2, 5 and 24 h then frozen at  $-80^{\circ}\text{C}$ , (iii) chilled with cool packs for 2, 5 and 24 h in an insulated container then frozen at  $-80^{\circ}\text{C}$ , and (iv) stored in a stool nucleic acid collection and preservation medium (catalogue number 63700; Norgen Biotek Corp., Canada) and frozen at  $-80^{\circ}\text{C}$  immediately or after 2, 5 and 24 h at ambient temperature. Subsequent stool samples collected from day 1 onwards were immediately frozen at  $-80^{\circ}\text{C}$ .

### DNA extraction and 16S amplicon sequencing

DNA was extracted in triplicate from 0.1 g homogenised fractions of stool using the QIAGEN DNeasy PowerSoil Kit following manufacturer's instructions. Following extraction, the V3 and V4 variable regions of the 16S gene were amplified by polymerase chain reaction (PCR) in 25  $\mu\text{l}$  reactions with the following master mix recipe: 1X

PCR buffer, 0.2 mM of each dNTP, 1 mM MgCl<sub>2</sub>, 0.2 μM of each forward (341F: CCTACGGGNGGCWGCAG) and reverse (806R: GGACTACNVGGGTWTCTAAT) primers (Klindworth *et al.*, 2013; Apprill *et al.*, 2015), 0.625 U Hot Start plus DNA Taq polymerase and molecular biology grade water. Thermocycler settings were as follows: 95 °C for 3 min, followed by 15 cycles of 94 °C for 45 s, 55 °C for 30 s and 72 °C for 3 min, followed by 10 cycles of 94 °C for 45 s, 60 °C for 30 s and 72 °C for 3 min, and finally 72 °C for 10 min. Four DNA-free negative controls using water were included in the PCR. Successful PCR amplification was verified by visualising products on an agarose gel. Sequencing adapters and multiplex indices were added to the PCR products in a second 25-μl PCR reaction with the following master mix recipe: 1X PCR buffer, 0.2 mM of each dNTP, 1 mM MgCl<sub>2</sub>, 0.2 mM of each Nextera index 1 and 2 primers, 0.1 μM of each forward and reverse Goyal index, 0.625 U Hot Start plus DNA Taq polymerase and molecular biology grade water. Thermocycler settings were as follows: 95 °C for 3 min, followed by 10 cycles of 94 °C for 45 s, 63 °C for 30 s and 72 °C for 3 min, and finally 72 °C for 10 min. Amplicons were purified using the QIAquick Gel Extraction Kit. Purified final products including the four negative controls were sent to the Genomics Resource Core Facility at Cornell University for 2 × 300 bp sequencing on an Illumina MiSeq using the v3 MiSeq Reagent Kit.

### 16S sequence data processing

Demultiplexed raw sequence data were imported into QIIME 2 v2017.12 (<https://qiime2.org/>). Using the DADA2 workflow (Callahan *et al.*, 2016) in QIIME 2, primer and low quality sequences were trimmed, and remaining reads subsequently denoised and merged. Alpha diversity metrics were calculated based on the counts table produced by DADA2, normalised to 1,086 sequences per sample (Fig. S1D). To assign taxonomy to the sequences, a classifier was first trained on sequences extracted from the SILVA 16S database release 128 (Quast *et al.*, 2013) using the 16S gene V3-4 universal primer sequences. This classifier was then run on the representative sequences produced by DADA2 to assign probable taxonomies to the corresponding sequences. The final counts table based on exact sequence variants (ESVs) was exported from QIIME2, chloroplast sequences removed, and used as input in R for statistical analysis.

### 16S amplicon-based microbial community composition analyses

The resulting ESV counts table from QIIME2 was imported together with sample metadata into R v3.4.1. A centered log-ratio transformation was applied to ESV counts before downstream analyses to ensure that the counts fulfilled assumptions of independence between predictor variables for statistical analyses (explained in Lê Cao *et al.*, 2016). Permutational multivariate analysis of variance (PERMANOVA) was used to assess whether factors such as subject-specificity, sampling time point and the simulated storage conditions significantly influenced community composition, as well as to determine the amount of variation in community composition attributable to each of these experimental factors. Principal component analysis (PCA) was used to visualise the clustering of samples based on their compositional similarities. Association of ESVs to experimental factors

were tested using generalised linear models (GLMs) and sparse partial least squares discriminant analysis (SPLSDA). PERMANOVA, PCA and GLMs are implemented in the vegan R package v2.4-6 (Oksanen et al., 2013), and SPLSDA implemented in the mixOmics R package v6.3.1 (Rohart et al., 2017). Statistical power was assessed using the HMP R package (La Rosa et al., 2012). Figures were edited in Inkscape v0.92 for clarity. All data are provided as supplemental files 1 to 4.

### Data Availability

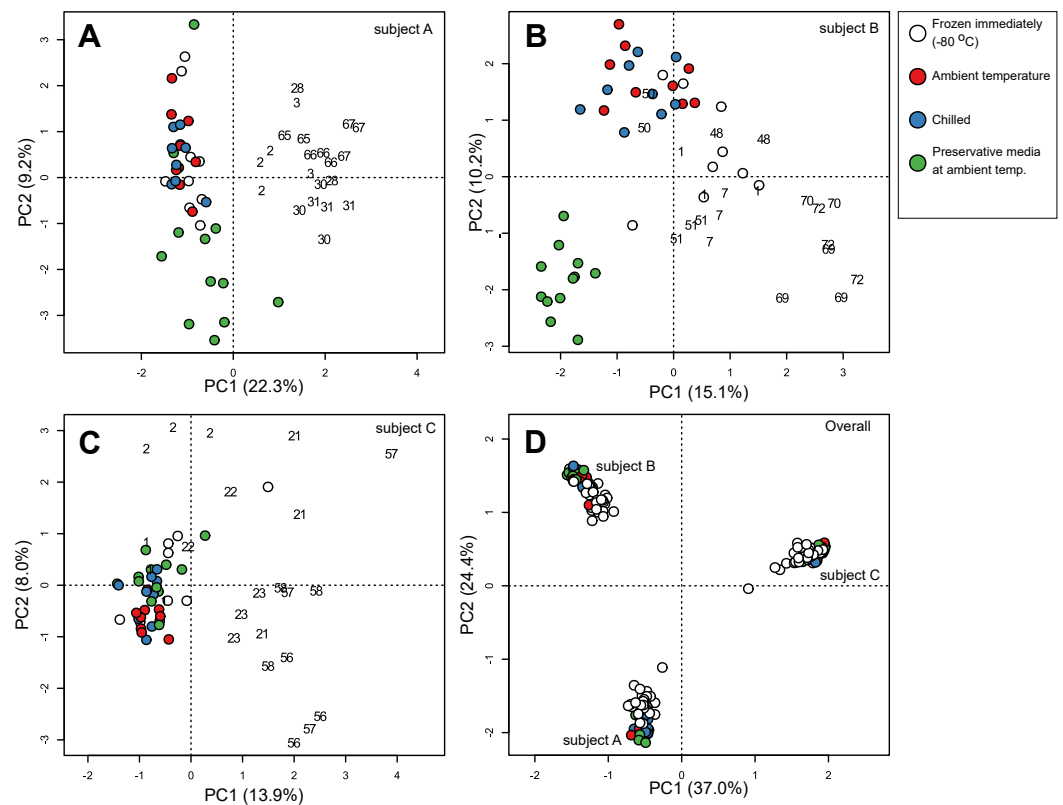
Raw sequence data generated for this study are available in the Sequence Read Archive under BioProject accession [PRJNA450690](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA450690).

## RESULTS

### Effect of storage conditions relative to inter-individual differences and intra-individual temporal variation

PCR amplicons from a total of 189 stool samples (Table S1) were pair-end sequenced on an Illumina MiSeq platform producing 886,252 reads, with a median of 4,556 reads per sample. Reads were quality-filtered to remove adaptor, primer, low-quality and chloroplast sequences, producing a final count of 475,361 reads. Of four DNA-free negative controls included in the PCR and sequencing workflow, two contained zero reads while the other two had three and four reads each. These controls were excluded from downstream statistical analyses. A total of seven samples each containing less than 1,000 reads were also excluded, leaving 182 samples.

To determine which parameters were most strongly associated with community composition, we first applied a centered log-ratio transformation on 16S counts data and then performed a PCA. Samples clustered according to their respective subjects based on community composition regardless of storage conditions or sampling time point (Fig. 2D), indicating that inter-individual variation had an overarching influence when comparing stool microbial communities among multiple subjects. Subject-to-subject stool microbial community specificity was also partially reflected in alpha diversity metrics including observed species, Shannon and Faith's phylogenetic diversity indices, in which samples from one subject (subject B) were consistently more species rich and phylogenetically diverse compared to samples from the other two subjects across all nine sampling time points ( $p < 0.001$ , Kruskal–Wallis test) (Fig. S1). Next, we performed PCA on samples from each subject separately and observed secondary clustering by sampling time point and transit storage conditions in all three subjects (Figs. 2A–2C). Stool communities collected across multiple days tended to cluster together relative to samples from weeks apart, reflecting shifts in community composition over short and medium-term durations (Fig. S2). This overall sequence of sample clustering beginning with between-subject variability, followed by within-subject temporal variability and sample storage conditions was supported by a permutational multivariate analysis of variance (PERMANOVA) indicating that subjects (accounting for 60.5% of the total variance in community composition), sampling time point (5.2%) and storage conditions (1.6%) were significantly associated with community composition in descending size of



**Figure 2** Principal component ordination of stool microbial community composition collected from three subjects over 72 days. Community composition was determined by sequencing 16S rRNA gene amplicons. Ordination of (A) samples from subject A; (B) samples from subject B; (C) samples from subject C; (D) samples from all three subjects. Each circle represents one sample, and colour of circle represents storage condition: white, immediately frozen at  $-80^{\circ}\text{C}$  red, ambient temperature; blue, chilled; green, preservative medium at ambient temperature. Numbers in (A–C) represent number of days subsequent samples were collected following initial sampling at day 0 (day 0 samples immediately frozen are represented by the white circles). All samples collected after day 0 were immediately frozen at  $-80^{\circ}\text{C}$ .

Full-size [DOI: 10.7717/peerj.6172/fig-2](https://doi.org/10.7717/peerj.6172/fig-2)

effect (Table 1). Storage duration (2, 5 or 24 h) before transferring samples to  $-80^{\circ}\text{C}$  were also significantly associated with shifts in community compositions in the three storage conditions (ambient, chilled and preservative) (PERMANOVA, Table 1).

### Effect of preservation medium on stool community composition

To specifically investigate the effects of sample storage conditions during transport to the laboratory on community composition, we performed PCA on community composition of stools that were subjected to all three storage conditions (i.e., stools from the first sampling time point on day 0). Samples were analysed separately by subject to remove the large influence of inter-individual variability. In all three subjects, PCA ordination of stool community composition revealed that samples stored in preservative medium often clustered to the exclusion of fresh frozen, ambient temperature or chilled samples (Figs. 3A–3C). In comparison, community compositions of samples without preservation buffer were comparatively more similar to each other irrespective of storage condition

**Table 1** PERMANOVA of all samples.

	Degrees of freedom	Sums of squares	F model	R <sup>2</sup>	P value
subject	2	58,888	198.1	0.605	<0.001
sampling time point	8	5088	4.3	0.052	<0.001
storage temperature	3	1597	3.6	0.016	<0.001
subject:sampling time point	16	9071	3.8	0.093	<0.001
storage temperature:duration	7	1184	1.1	0.012	0.04
Residuals	145	21555		0.221	

**Notes.**

**subject:** three subjects A, B and C.

**sampling time point:** days stools were collected.

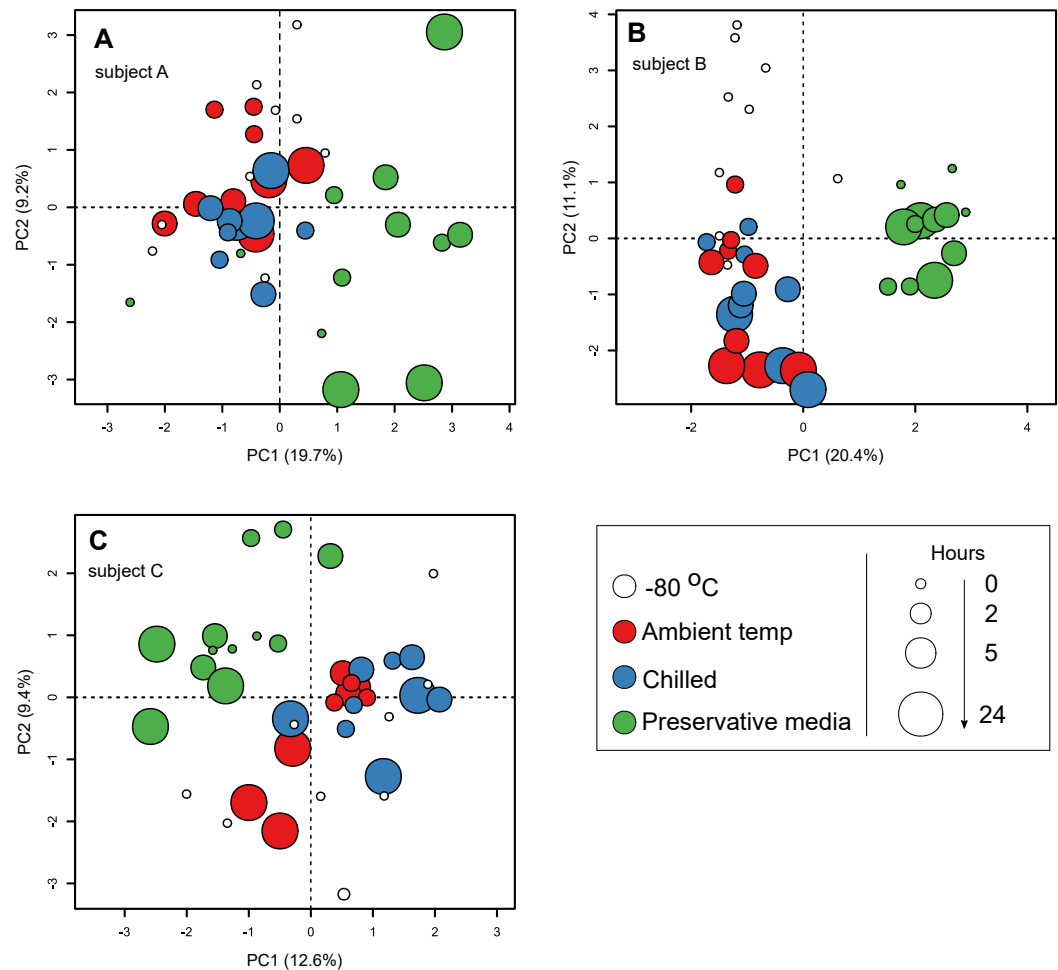
**storage temperature** before long-term freezing: immediately frozen, ambient, chilled, preservative at ambient.

**duration** of storage before long-term freezing: 0, 2, 5 or 24 h.

(Fig. 3A–3C). Interestingly, effects of the preservation medium on alpha diversity was only significant in one subject (subject B, Kruskal–Wallis tests,  $p < 0.05$ ) (Fig. S3), indicating that the preservation medium tested here introduced inconsistent biases into the resident stool community across samples. To examine which ESVs were enriched in the preserved samples, we used a GLM to predict ESV associations with the use of preservative medium based on centered log-ratio transformed 16S counts. Using data from all three subjects, the model predicted 51 ESVs significantly associated with preservation medium treatment (false discovery rate-adjusted  $p < 0.05$ ), 44 of which were enriched in preserved samples (Table S2). These associations were supported by an SPLSDA implemented in the mixOmics R package, in which 37 ESVs were overlapping between the two methods (GLM and SPLSDA, Table S2). In total, 26 of the 37 ESVs were clostridial. When communities were analysed separately by subject, there was little overlap in the ESVs associated with preservation medium among subjects (Table S3) which was most likely due to the strong inter-individual variability in stool communities. Nevertheless, there was a strong representation of clostridial ESVs associated with the use of preservation medium in all three subjects (Table S3). These results suggest that the preservation medium tested here may be enriching for certain clostridial taxa in stools. Thus, its effects need to be further evaluated in other stool and non-stool samples to determine whether the observed enrichment is specific to this study. Since these findings indicated that community composition in samples treated with preservation medium was altered relative to preservative-free options, we revisited the initial analysis of all 182 samples and excluded those stored in preservation medium. When the PERMANOVA was restricted to non-preserved samples, sample storage (ambient vs. chilled temperatures) now accounted for a smaller but still significant amount of variability (0.6% vs 1.6% when preserved samples were included), and duration of storage was no longer associated with community composition when analysing stools from multiple subjects combined (Table 2).

### Effect of storage duration on community composition

To more specifically identify at which time points significant differences could be detected in the community composition of stools stored at ambient and chilled temperatures relative



**Figure 3** Principal component ordination of stool microbial community composition by storage condition and duration. Community composition was determined by sequencing 16S rRNA gene amplicons. (A) samples from subject A; (B) samples from subject B; (C) samples from subject C. Each circle represents one sample. Colour of circle represents storage condition: white, immediately frozen at  $-80^{\circ}\text{C}$  red, ambient temperature; blue, chilled; green, preservative medium at ambient temperature. Size of circle represents duration samples were subject to their respective conditions before frozen long-term at  $-80^{\circ}\text{C}$ .

Full-size [DOI: 10.7717/peerj.6172/fig-3](https://doi.org/10.7717/peerj.6172/fig-3)

to fresh frozen samples, we separately compared profiles of samples from 2, 5, and 24 h to the 0 h reference using PERMANOVA. At 2 h, stools from all three subjects did not differ in community composition from their respective fresh frozen fractions regardless of ambient or chilled storage. At 5 and 24 h, only stools from subject B showed significant alterations in both ambient and chilled conditions whereas stool profiles of the other two subject did not differ from their respective 0 h fractions (Table S4). Similar to the non-systematic effects of the tested preservation medium on community composition, these observations further indicate that storage conditions can produce varying outcomes on stool community composition among individuals. We then assessed changes in the centered log-ratio transformed counts of each ESV across the 0, 2, 5 and 24 h time points using GLMs to



**Table 2** PERMANOVA excluding preserved samples.

	Degrees of freedom	Sums of squares	F model	R <sup>2</sup>	P value
subject	2	45,861	170.1	0.604	<0.001
sampling time point	8	4895	4.5	0.064	<0.001
storage temperature	2	476	1.8	0.006	<0.001
subject:sampling time point	16	8923	4.1	0.117	<0.001
storage temperature:duration	4	573	1.1	0.008	0.27
residuals	113	15237		0.2	

**Notes.**

**subject:** three subjects A, B and C.

**sampling time point:** days stools were collected.

**storage temperature** before long-term freezing: immediately frozen, ambient, chilled, preservative at ambient.

**duration** of storage before long-term freezing: 0, 2, 5 or 24 h.

identify taxa enriched or depleted over time. A total of 86 ESVs were implicated, of which 74 were enriched, eight depleted, and four both enriched and depleted among the three subjects (Table S5). While the number of ESVs enriched or depleted could be influenced by the relative nature of microbial community data, their taxonomic identities were specific to subjects and included a range of common gut taxa (Table S5). These observations indicate that there is likely no generalizable pattern in how stool community composition from multiple individuals changes over a 24 h storage duration. Within subjects, many ESVs enriched or depleted over time in the ambient and chilled conditions overlapped, indicating that temperature may not radically influence enrichment/depletions at least in the first 24 h. Taken together, these results show that responses in stool community composition to storage conditions is specific to individuals and may not be easily accounted for in downstream analyses. While these shifts were small compared to inter-individual variability detected in stool community compositions (Table 2), their potential impacts on gut communities should be considered when interpreting survey results.

### Effect of sampling different stool fractions on community composition

For each of the three subjects, we produced replicate community profiles from three separate fractions of the whole stool (from day 0) taken at least 5 cm apart to assess variability in community composition attributable to heterogeneity in microbial niches within stools. Subsampling of different fractions was significantly associated with distinct community compositions in two of three subject ( $p < 0.05$ ; PERMANOVA) (Table S6, Fig. S4). We then combined the samples from all three subjects and performed a PERMANOVA to assess the relative influences of inter-individual variability and intra-sample heterogeneity on stool community composition. The PERMANOVA indicated that in the absence of other experimental variables, differences in community composition linked to sampling different fractions of the same stools were subtle but still significant despite the stronger influence of inter-individual variation (72.6% vs. 3.3% R<sup>2</sup>, PERMANOVA) (Table S6). This finding indicates that distinct fractions of the same stool may not necessarily be true biological replicates, therefore collecting a larger sample for homogenisation before storage

may be necessary. Taken together, the variables investigated in this study ranked according to their effect on stool microbial community composition are: inter-individual variation >intra-individual temporal variation >use of preservative media >storage temperature >storage duration (up to 24 h) before long term freezing.

## DISCUSSION

Large-scale population gut microbiota surveys often face a logistical issue of procuring hundreds of stools while maintaining sample integrity. Typically, the ideal laboratory procedure for storing stools is often immediate freezing at  $-80^{\circ}\text{C}$  following defecation (*Carroll et al., 2012; Fouhy et al., 2015*); however, this workflow is not feasible in surveys involving large numbers of participants from the general public. A major concern with collection of stool and other biological samples intended for microbial community surveys is the issue of prolonged exposure to ambient temperatures during transit, leading to shifts in community composition (*Cardona et al., 2012; Carroll et al., 2012; Guo et al., 2016*). Biological samples are usually chilled or mixed with preservative media to suppress opportunistic microorganisms from flourishing and dominating community profiles (*Vandeputte et al., 2017*), although this problem may be somewhat alleviated by setting a maximum allowed transit duration. A few studies have compared the use of preservation buffers in maintaining microbial community profiles and generally recommend its use especially when samples cannot be immediately frozen due to logistical constraints (*Menke et al., 2017; Flores et al., 2015; Voigt et al., 2015*). Our results, however, show that preservation media may not be necessary for human stool samples for short transit durations of up to 24 h as the use of a medium was associated with larger shifts in microbial community composition compared to samples stored at ambient and lowered temperatures (*Fig. 3*). While there is a possibility that the three samples collected in this study contained less appreciable abundances of fast-growing microorganisms favoured at ambient temperatures, our findings echo recommendations relating to storage options and durations made by Vandeputte and colleagues, in which they rated chilled buffer-free samples for up to 24 h as the best option other than immediate freezing (Table 1 in *Vandeputte et al., 2017*). Nevertheless, we only tested one preservative medium on stools from three subjects and there may be alternatives that produce community profiles more comparable to buffer-free storage conditions. Song and colleagues previously tested five sample preservative media on human stool samples and reported a general increase in Firmicutes taxa in samples preserved using ethanol (70% and 95%), RNALater or Whatman FTA cards but not OMNIgene Gut (*Song et al., 2016*). Our observations of increased relative abundances of clostridial ESVs (members of Firmicutes) in preserved samples (Table S2) was consistent with their findings. Although our data set presented here provided sufficient statistical power for discriminating preserved and non-preserved samples based on community profiles (98.5% power at 0.05 alpha level, Monte-Carlo permutation of Dirichlet multinomial likelihood ratio implemented in HMP R package), additional tests inclusive of a larger number of subjects are required to determine whether composition and/or diversity of the resident stool microbial community influence performance of preservative media.

As highlighted above, our findings indicated that preservative media may not be necessary for stools collected and frozen within 2 h (Fig. 3, Table S4). Stools stored at ambient and chilled temperatures produced community profiles more closely resembling those from fresh frozen samples, however, we noted that these outcomes were variable between subjects (Table S4). A few published studies have reported that storing stools at 4 °C for 24 h did not significantly shift community profiles compared to fresh frozen –80 °C controls (Bassis *et al.*, 2017; Choo, Leong & Rogers, 2015; Tedjo *et al.*, 2015; Cardona *et al.*, 2012). While profiles from two of our three subjects did not significantly change over 24 h at ambient and chilled temperatures, significant shifts were detected in the stools of the third subject (subject B) at five hours post collection. Furthermore, when we attempted to identify taxa associated with these shifts, we observed that the relatively enriched and depleted ESVs were specific to subject and encompassed generic human gut taxa including various *Bacteroidales* and clostridial members (Table S5). In some instances, the same ESV present in multiple subjects showed enrichment in one but depletion or no change in relative abundance over time in other subjects. Community shifts over time in non-frozen samples are often associated with reductions in community diversity as profiles become dominated by a limited number of taxa (Choo, Leong & Rogers, 2015) due to factors such as altered temperatures relative to the normal human body or exposure to oxygen favouring aerobes and/or facultative anaerobes (Chu *et al.*, 2017). However, subject B's profiles were increased in diversity (Fig. S3) due to depletion of a dominant *Prevotella* ESV from an average 26.0% relative abundance to 11.4% and 14.6% under ambient and chilled temperatures, respectively (Table S5). These results indicate that how stool communities respond to non-frozen storage varies by individual due to their specific configurations in stool microbial communities, and generalisations as to how stool community profiles respond over time after collection have to be made with caution. As *Prevotella* constitutes a major component of human gut communities and is a common feature in a large proportion of the human population (enterotype 2 described in Arumugam *et al.*, 2011), changes as seen in subject B's profiles could severely impact the outcome of human population-based microbial community surveys.

The various human stool microbial community profiling studies mentioned here, like most other gut microbiome studies, involve homogenising samples to potentially reduce variability associated with subsampling (Vandeputte *et al.*, 2017; Sinha *et al.*, 2016; Choo, Leong & Rogers, 2015; Tedjo *et al.*, 2015; Carroll *et al.*, 2012). The issue with subsampling small portions of a stool for DNA isolation is that microenvironments can harbour specific taxa as shown by a quantitative PCR study measuring abundances of common human gut taxa in non-homogenised stools (Gorzalak *et al.*, 2015). Using 16S-based community profiling, we also detected significant differences in overall community composition in two of three subjects when subsamples derived from different fractions of the same stool were compared. Therefore, we recommend based on these observations collection of a larger stool sample and inclusion of a homogenisation step before samples are stored frozen. That said, differences in community composition linked to within-sample variability is relatively minor compared to other experimental factors such as inter-individual and temporal

variation ([Voigt et al., 2015](#)) and is unlikely to influence interpretation of a multi-subject stool community survey.

The biggest limitation in this study is the limited number of subjects involved as we subsequently showed that inter-individual variability largely influences how profiles respond to the storage conditions tested. Future surveys should include a larger population of subjects to capture greater variability in stool community composition, which may help discern general patterns in how stool taxa respond to storage conditions. Another issue not addressed here is the influence of oxygen on stool community composition. One study demonstrated that exposure to oxygen primarily lowered relative abundances of *Faecalibacterium* and *Megamonas*, while increasing that of *Bacteroides* ([Chu et al., 2017](#)). These alterations could be minimised by storing samples in air-tight containers with oxygen scavengers, but we have not incorporated such elements in our sample collection workflow as it introduces additional inconveniences to participants of stool community surveys during self-sampling. Furthermore, Chu and colleagues showed that using standard 16S sequencing, community profiles from stools processed under anaerobic and aerobic conditions equally varied from control samples (anaerobic with cysteine as reducing agent to remove oxygen from solution) (Fig. 1A in [Chu et al., 2017](#)). Lastly, when comparing the findings presented here with other studies attention should be paid to the primers used as primer choice is widely known to influence 16S-based community compositional profiles ([Walker et al., 2015](#); [Engelbrektsen et al., 2010](#)).

## CONCLUSION

Stool sample delivery conditions and subsampling biases are known to alter microbial community composition. The use of preservative media, while meant to lessen such alterations, resulted in non-systematic, non-taxa specific but subject-specific changes in stool community profiles. However, when compared to inter-individual differences and community variations within an individual over time, effects due to delivery conditions and biases in sample fractions are small and unlikely to obscure inter-individual variability in surveys involving multiple subjects. Nevertheless, the effects of sample collection strategies should be given consideration especially when dealing with sensitive applications such as faecal microbiota transplants, which are reliant on exact resident microbial species.

## ACKNOWLEDGEMENTS

We wish to thank the stool donors for participating in this study, and reviewers for providing excellent feedback and suggestions.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study was supported by a seed fund for gut microbiota research provided by the Faculty of Medicine, The Chinese University of Hong Kong. The funders had no role

in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### **Grant Disclosures**

The following grant information was disclosed by the authors:  
Faculty of Medicine, The Chinese University of Hong Kong.

### **Competing Interests**

The authors declare there are no competing interests.

### **Author Contributions**

- Yun Kit Yeoh conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Zigui Chen analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Mamie Hui analyzed the data, approved the final draft.
- Martin C.S. Wong, Wendy C.S. Ho and Miu Ling Chin performed the experiments, contributed reagents/materials/analysis tools, approved the final draft.
- Siew C. Ng and Francis K.L. Chan conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.
- Paul K.S. Chan conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper, approved the final draft.

### **Human Ethics**

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

This study has been approved by the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (reference number 2016.707).

### **DNA Deposition**

The following information was supplied regarding the deposition of DNA sequences:

Raw sequence data generated for this study are available in the Sequence Read Archive under BioProject accession [PRJNA450690](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA450690).

### **Data Availability**

The following information was supplied regarding data availability:

Raw data is provided in the [Supplemental Files](#).

### **Supplemental Information**

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.6172#supplemental-information>.

## REFERENCES

- Apprill A, McNally S, Parsons R, Weber L. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* 75:129–137 DOI 10.3354/ame01753.
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto JM, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, De Vos WM, Brunak S, Doré J, MetaHIT Consortium, Antolin M, Artiguenave F, Blottiere HM, Almeida M, Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariac G, Dervyn R, Foerstner KU, Friss C, Van de Guchte M, Guedon E, Haimet F, Huber W, Van Hylckama-Vlieg J, Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A, Melo Minardi R, M’rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y, Zeller G, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174–180 DOI 10.1038/nature09944.
- Bassis CM, Moore NM, Lolans K, Seekatz AM, Weinstein RA, Young VB, Hayden MK, CDC Prevention Epicenters Program. 2017. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC Microbiology* 17:78 DOI 10.1186/s12866-017-0983-9.
- Bedarf JR, Hildebrand F, Coelho LP, Sunagawa S, Bahram M, Goeser F, Bork P, Wüllner U. 2017. Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson’s disease patients. *Genome Medicine* 9:Article 39 DOI 10.1186/s13073-017-0428-y.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods* 13:581–583 DOI 10.1038/nmeth.3869.
- Cardona S, Eck A, Cassellas M, Gallart M, Alastrue C, Dore J, Azpiroz F, Roca J, Guarner F, Manichanh C. 2012. Storage conditions of intestinal microbiota matter in metagenomic analysis. *BMC Microbiology* 12:158 DOI 10.1186/1471-2180-12-158.
- Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y. 2012. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. *PLOS ONE* 7:e46953 DOI 10.1371/journal.pone.0046953.
- Choo JM, Leong LE, Rogers GB. 2015. Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports* 5:16350 DOI 10.1038/srep16350.
- Chu ND, Smith MB, Perrotta AR, Kassam Z, Alm EJ. 2017. Profiling living bacteria informs preparation of fecal microbiota transplantations. *PLOS ONE* 12:e0170922 DOI 10.1371/journal.pone.0170922.

- Engelbrekton A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P. 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME Journal* 4:642–647 DOI 10.1038/ismej.2009.153.
- Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D’hoë K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaut L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. 2016. Population-level analysis of gut microbiome variation. *Science* 352:560–564 DOI 10.1126/science.aad3503.
- Flores R, Shi J, Yu G, Ma B, Ravel J, Goedert JJ, Sinha R. 2015. Collection media and delayed freezing effects on microbial composition of human stool. *Microbiome* 3:Article 33 DOI 10.1186/s40168-015-0092-7.
- Fouhy F, Deane J, Rea MC, O’Sullivan Ó, Ross RP, O’Callaghan G, Plant BJ, Stanton C. 2015. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. *PLOS ONE* 10:e0119355 DOI 10.1371/journal.pone.0119355.
- 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 DOI 10.1371/journal.pone.0134802.
- Guo Y, Li SH, Kuang YS, He JR, Lu JH, Luo BJ, Jiang FJ, Liu YZ, Papasian CJ, Xia HM, Deng HW, Qiu X. 2016. Effect of short-term room temperature storage on the microbial community in infant fecal samples. *Scientific Reports* 6:26648 DOI 10.1038/srep26648.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41:e1 DOI 10.1093/nar/gks808.
- La Rosa PS, Brooks JP, Deych E, Boone EL, Edwards DJ, Wang Q, Sodergren E, Weinstock G, Shannon WD. 2012. Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PLOS ONE* 7:e52078 DOI 10.1371/journal.pone.0052078.
- Lê Cao KA, Costello M-E, Lakis VA, Bartolo F, Chua XY, Brazeilles R, Rondeau P. 2016. MixMC: a multivariate statistical framework to gain insight into microbial communities. *PLOS ONE* 11:e0160169 DOI 10.1371/journal.pone.0160169.
- Menke S, Gillingham MAF, Wilhelm K, Sommer S. 2017. Home-made cost effective preservation buffer is a better alternative to commercial preservation methods for microbiome research. *Frontiers in Microbiology* 8:Article 102 DOI 10.3389/fmicb.2017.00102.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O’hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2013. Package ‘vegan.’ Community ecology package. Version 2. Available at <https://CRAN.R-project.org/package=vegan>.

- 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 J-M, 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 DOI 10.1038/nature11450.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590–D596 DOI 10.1093/nar/gks1219.
- Rohart F, Gautier B, Singh A, Lê Cao KA. 2017. mixOmics: an R package for ‘omics feature selection and multiple data integration. *PLOS Computational Biology* 13:e1005752 DOI 10.1371/journal.pcbi.1005752.
- Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, Costea PI, Godneva A, Kalka IN, Bar N, Shilo S, Lador D, Vila AV, Zmora N, Pevsner-Fischer M, Israeli D, Kosower N, Malka G, Wolf BC, Avnit-Sagi T, Lotan-Pompan M, Weinberger A, Halpern Z, Carmi S, Fu J, Wijmenga C, Zhernakova A, Elinav E, Segal E. 2018. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555:210–215 DOI 10.1038/nature25973.
- 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 Epidemiology, Biomarkers & Prevention* 25:407–416 DOI 10.1158/1055-9965.EPI-15-0951.
- 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 DOI 10.1128/mSystems.00021-16.
- Sze MA, Schloss PD. 2016. Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio* 7:e01018–16 DOI 10.1128/mBio.01018-16.
- Tedjo DI, Jonkers DMAE, 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 DOI 10.1371/journal.pone.0126685.
- Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444:1027–1031 DOI 10.1038/nature05414.
- Vandeputte D, Tito RY, Vanleeuwen R, Falony G, Raes J. 2017. Practical considerations for large-scale gut microbiome studies. *FEMS Microbiology Reviews* 41:S154–S167 DOI 10.1093/femsre/fux027.
- Vogt NM, Kerby RL, Dill-McFarland KA, Harding SJ, Merluzzi AP, Johnson SC, Carlsson CM, Asthana S, Zetterberg H, Blennow K, Bendlin BB, Rey FE. 2017.



- Gut microbiome alterations in Alzheimer's disease. *Scientific Reports* 7:13537 DOI 10.1038/s41598-017-13601-y.
- Voigt AY, Costea PI, Kultima JR, Li SS, Zeller G, Sunagawa S, Bork P. 2015. Temporal and technical variability of human gut metagenomes. *Genome Biology* 16:Article 73 DOI 10.1186/s13059-015-0639-8.
- 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:Article 26 DOI 10.1186/s40168-015-0087-4.
- Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosower N, Rein M, Zilberman-Schapira G, Dohnalová L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E. 2015. Personalized nutrition by prediction of glycemic responses. *Cell* 163:1079–1094 DOI 10.1016/j.cell.2015.11.001.
- Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, Wang J, Imhann F, Brandsma E, Jankipersadsing SA, Joossens M, Cenit MC, Deelen P, Swertz MA, Weersma RK, Feskens EJM, Netea MG, Gevers D, Jonkers D, Franke L, Aulchenko YS, Huttenhower C, Raes J, Hofker MH, Xavier RJ, Wijmenga C, Fu J. 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352:565–569 DOI 10.1126/science.aad3369.