

Video Article

Guided Protocol for Fecal Microbial Characterization by 16S rRNA-Amplicon Sequencing

Ayelet Di Segni¹, Tzipi Braun¹, Marina BenShoshan¹, Sarit Farage Barhom¹, Efrat Glick Saar¹, Karen Cesarkas¹, James E. Squires², Nathan Keller^{1,3}, Yael Haberman^{1,4}

¹Sheba Medical Center

²Children's Hospital of Pittsburgh of UPMC

³Ariel University

⁴Cincinnati Children's Hospital Medical Center

Correspondence to: Yael Haberman at Yael.Haberman@cchmc.org

URL: <https://www.jove.com/video/56845>

DOI: [doi:10.3791/56845](https://doi.org/10.3791/56845)

Keywords: Biology, Issue 133, Microbiome, 16S rRNA sequencing, Next Generation Sequencing, V4 variable region, fecal samples, library preparation

Date Published: 3/19/2018

Citation: Di Segni, A., Braun, T., BenShoshan, M., Farage Barhom, S., Glick Saar, E., Cesarkas, K., Squires, J.E., Keller, N., Haberman, Y. Guided Protocol for Fecal Microbial Characterization by 16S rRNA-Amplicon Sequencing. *J. Vis. Exp.* (133), e56845, doi:10.3791/56845 (2018).

Abstract

The human intestinal microbiome plays a central role in protecting cells from injury, in processing energy and nutrients, and in promoting immunity. Deviations from what is considered a healthy microbiota composition (dysbiosis) may impair vital functions leading to pathologic conditions. Recent and ongoing research efforts have been directed toward the characterization of associations between microbial composition and human health and disease.

Advances in high-throughput sequencing technologies enable characterization of the gut microbial composition. These methods include 16S rRNA-amplicon sequencing and shotgun sequencing. 16S rRNA-amplicon sequencing is used to profile taxonomical composition, while shotgun sequencing provides additional information about gene predictions and functional annotation. An advantage in using a targeted sequencing method of the 16S rRNA gene variable region is its substantially lower cost compared to shotgun sequencing. Sequence differences in the 16S rRNA gene are used as a microbial fingerprint to identify and quantify different taxa within an individual sample.

Major international efforts have enlisted standards for 16S rRNA-amplicon sequencing. However, several studies report a common source of variation caused by batch effect. To minimize this effect, uniformed protocols for sample collection, processing, and sequencing must be implemented. This protocol proposes the integration of broadly used protocols starting from fecal sample collection to data analyses. This protocol includes a column-free, direct-PCR approach that enables simultaneous handling and DNA extraction of large numbers of fecal samples, along with PCR amplification of the V4 region. In addition, the protocol describes the analysis pipeline and provides a script using the latest version of QIIME (QIIME 2 version 2017.7.0 and DADA2). This step-by-step protocol is aimed to guide those interested in initiating the use of 16S rRNA-amplicon sequencing in a robust, reproducible, easy to use, detailed way.

Video Link

The video component of this article can be found at <https://www.jove.com/video/56845/>

Introduction

Concentrated efforts have been made to better understand microbiome diversity and abundance, as another aspect of capturing difference and similarities between individuals in healthy and pathological conditions. Age^{2,3}, geography⁴, lifestyle^{5,6}, and illness⁵ were shown to be associated with the composition of the gut microbiome, but many conditions and populations have not yet been fully characterized. Recently it has been reported that the microbiome can be modified for therapeutic applications^{7,8,9}. Therefore, additional insight into the relationship between various physiological conditions and the microbial composition is the first step toward optimization of potential future modifications.

The traditional microbial culture methods are limited by low yields^{10,11}, and are conceptualized as a binary state where a bacteria is either present in the gut or not. High-throughput DNA-based sequencing has revolutionized microbial ecology, enabling the capture of all members of the microbial community. However, sequence read length and quality remain significant barriers to accurate taxonomy assignment¹². Furthermore, high-throughput based experiments may suffer from batch effects, where measurements are affected by non-biological or non-scientific variables¹³. In recent years, several programs have been established to study the human microbiome, including the American Gut project, the United States (US) Human Microbiome Project, and the United Kingdom (UK) MetaHIT project. These initiatives have generated vast amounts of data that are not easily comparable due to a lack of consistency in their approaches. A variety of international projects such as the International Human Microbiome Consortium, the International Human Microbiome Standards project, and the National Institute of Standards and Technology (NIST) attempted to address some of these issues¹⁴, and developed standards for microbiome measurements which should

enable the achievement of reliable reproductive results. Described here is an integrated protocol of several broadly used methods^{15,16} for 16S rRNA high-throughput sequencing (16S-seq) starting from fecal sample collection thru data analyses. The protocol describes a column-free PCR approach, originally designed for direct extraction of plant DNA¹⁶, to enable the simultaneous handling of large numbers of fecal samples in a relatively short time with high quality amplified DNA for targeted sequencing of the microbial variable V4 region on a common sequencing platform. This protocol aims to guide scientists interested in initiating the use of 16S rRNA-amplicon sequencing in a robust, reproductive, easy to use, detailed way, using important controls. Having a guided and detailed step-by step protocol may minimize batch effect and thus will allow more comparable sequencing results between labs.

Protocol

Ethical approval for the study was granted by the Sheba Local Research Ethics Committee and all methods were performed in accordance with the relevant guidelines and regulations. The protocol received a patient consent exception from the local Ethical Review Board, since the fecal material that were used were already submitted to the microbiology core as part of clinical workup and without identifiable patient information other than age, gender, and microbial results. Written, informed consent was obtained from healthy volunteers and the Institutional Review Board approved the study. Some of those samples have already been included in a previous analysis¹.

1. Sample Handling

1. Collect an approximately 5 mm² smear (roughly the size of a pencil eraser) from a fresh fecal sample with a sterile swab in a test tube (see **Table of Materials**). Store all swabs containing fecal samples at -80 °C within 24 h. The fecal swabs can remain there until further processing.

2. DNA Extraction

1. Thaw the Extraction and Dilution solutions at room temperature (see **Table of Materials**).
2. Transfer the fecal swab into an empty 2 mL collection tube (see **Table of Materials**). Adjust the swab stick size by cutting it using a clean scissors to enable tube closure with minimum cross-contamination. Add 250 µL of Extraction solution to each collection tube containing the fecal swab and vortex to mix.
3. Heat the samples for 10 min in a boiling water bath (95–100 °C). Add 250 µL of Dilution solution to each sample and vortex to mix.
4. Store the 2 mL tubes containing the extracted DNA and the swab at 4 °C.

3. PCR and Library Preparation

For steps 3.1 and 3.2, work in a PCR workstation that provides clean, template and amplicon free environment.

1. Label the primers (**Table 1**) according to their barcode. Dilute each primer in double distilled water (DDW) to a 50 µM concentration and store at -20 °C.
2. Use a 96-well plate for PCR reactions. Each plate can contain 32 different samples, which are tagged by 32 different index primers. Thaw the forward primer and the 32 reverse primers at room temperature and dilute them to 5 µM.
3. Prepare PCR reaction mixes for 100 reactions (final volume in each well will be 20 µl) by mixing 100 µL of 5 µM Forward primer, 1 mL 2X PCR Master mix, and 400 µL DDW. Put 15 µL of this PCR mix in each well (total of 96 wells). Add 1 µL of each 5 µM Reverse indexed primer to 3 different wells (32 different primers in triplicates gives a total of 96 wells).
4. In a pre-PCR dedicated zone, meaning a clean bench that is template and amplicon free, add 4 µL of each extracted DNA sample to reaction mixtures (each extracted DNA sample is amplified in triplicate — 32 samples per 96-well plate).
5. Run PCR with the following settings: initial denaturation of 94 °C for 3 min; followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 55 °C for 1 min, and extension at 72 °C for 1 min; and a final extension at 72 °C for 10 min.
6. On a different bench, which will be defined as a post-PCR dedicated zone, combine each triplicate PCR reaction into a single volume tube (60 µL per sample).
7. To assess the quality of PCR amplicons, run 4 µL of each combined-PCR reaction on an ethidium bromide-stained 1% agarose gel. Under UV wavelength of 260 nm, the positive amplicons will appear in an expected band size of 375–425 bp. Only these amplicons will be included in the subsequent steps.

NOTE: Each sample is amplified in triplicate, meaning that each sample is amplified in 3 different PCR reactions. Do not scale up.

4. Library Quantification and Cleaning

1. In order to get an equimolar concentration pool of all PCR samples, quantify each amplicon by a double stranded DNA (dsDNA quantify reagent) fluorescent nucleic acid stain (see **Table of Materials**), suitable for the simultaneous quantification of a large amount of samples. NOTE: Since the size range of the amplicon is equivocal, the actual amount in nanograms is used to load an equimolar concentration.
2. Combine 500 ng of each sample into a single, sterile tube. Vortex to mix.
3. Run 200 µL of the pooled library on an Ethidium Bromide-stained 1% agarose gel. Extract the 375–425 bp bands from the gel using a gel extraction kit according to the manufacturer's instructions (see **Table of Materials**) and elute the pooled library in 80 µL DDW. NOTE: The pooled library should be size selected to reduce non-specific amplification products from host DNA.
4. Measure the final library concentration using a highly sensitive dsDNA detecting kit according to the manufacturer's instructions (see **Table of Materials**). Measure the accurate library size using a highly sensitive separation and analysis kit for DNA libraries according to the manufacturer's instructions (see **Table of Materials**).

5. Sequencing

1. Dilute the pooled library to 7 pM with addition of 20% control library (see **Table of Materials**), according to the sequencing machine protocol.
2. For sequencing, use custom designed read primers that are complimentary to the V4 amplification primers (see **Table 1**).
3. Use a third custom designed sequencing primer that reads the barcode in an additional cycle (see **Table 1**) and generates paired-end reads of 175 bases in length in each direction, according to the manufacturer's specifications.
4. Run the sequencing machine and obtain FASTQ files according to the manufacturer's protocol.

6. Data Processing

1. Stitch together and process the overlapping paired-end FASTQ files in a data curation pipeline implemented in QIIME 2 version 2017.7.0¹⁷. Demultiplex the reads according to sample specific barcodes.
2. Use DADA2¹⁸ for quality control and sequence variant (SVs) detection. Truncate reads at 13 bases from the 3' end and 15 bases from the 5' end, discard reads with more than 2 expected errors. Identify and remove the chimeras using the consensus method — chimeras are detected in samples individually, and sequences found chimeric in a sufficient fraction of samples are removed.
3. Perform SVs taxonomic classification using a Naive Bayes fitted classifier, trained on the August 2013 99% identity Greengenes database⁶, for 175 long reads and the Forward/Reverse primer set.
4. Rarefy all samples to depth of 2,146 sequences.
5. Use Unweighted UniFrac for measurement of β -diversity (between sample diversity^{19,20}) on the rarefied samples, to avoid a sample size effect.
6. Use the resulting distance matrix to perform a principal coordinates analysis (PCoA).

NOTE: The data processing script and mapping files are provided as supplementary material (**Supplementary Material 1 and 2**).

Representative Results

A schematic illustration of the protocol is shown in **Figure 1**.

We have prospectively collected stool samples from hospitalized patients with suspected infectious diarrhea. Those samples were submitted to the Clinical Microbiology Lab at the Sheba Medical Center between February and May 2015, as was previously described¹. Stool samples were subjected to conventional microbiological culture performed at the Clinical Microbiology Lab and to broad range high-throughput 16S-seq in parallel. In addition, stool samples from healthy adults were sequenced for comparison.

In this analysis, the focus was on the quality control of the data, and show representative results of the output obtained from QIIME2¹⁷. Initially, sequencing results achieved in 6 negative control samples were reviewed for quality control including: i) PCR without template, ii) PCR on clean and sterile swabs in extraction and dilution solution, and iii) PCR on mixed extraction and dilution solutions. **Table 2** shows that all negative control samples had ≤ 118 total reads (average of 29 sequences), primarily from the Mycoplasma taxa, while all other samples analyzed showed mean total reads of 10622.57 (± 1211.34 standard error) and no Mycoplasma read passed the quality control of the main samples.

Sixteen samples that each originated from the same stool sample but went through different library preparation with different reverse barcoded primers were used as positive controls. Those 16 samples included 8 with positive cultures for *Campylobacter*, *Salmonella*, or *Shigella* that were reported by the clinical microbiology lab, and 8 that were reported as having negative cultures. An area plot at the phyla level is shown in **Figure 2**. Samples that originated from the same stool sample but went through different library preparation with different reverse barcoded primers show very consistent relative abundance (Mantel test between duplicates of phylum level taxonomic relative abundance values simulated p-value = 1×10^{-04} based on 9999 replicates).

Principal coordinates analysis (PCoA), which reduces dimensionality of the input data, was used on the UniFrac matrix to visually explore sample separation and similarity. In **Figure 3A**, sequenced samples that originated from the same original stool sample but went through different library preparation are colored the same. Indeed, those are located relatively closely in the PCoA plot (Mantel test between duplicates PC1 PC2 values simulated p-value = $1e-04$ based on 9999 replicates). In addition, the average unweighted unifrac beta diversity distance between samples in this study is 0.706, while the average distance between two duplicates is 0.235 (Wilcoxon rank sum test p-value = 4.205×10^{-11}). Interestingly, samples that were culture positive for *Campylobacter*, *Salmonella*, or *Shigella* enteropathogens showed significantly different PC1 values (Wilcoxon rank sum test p-value = 0.011) in comparison to samples with negative culture results. **Figure 3B** shows the same PCoA plot but now samples are colored by the relative abundance of Proteobacteria. Samples with high Proteobacteria abundance had significantly different PC1 values (Spearman correlation $r = -0.533$, p-value = 0.000975). Those results are consistent with previous analysis of this data, where hospitalized patients have profound increases in taxa from Proteobacteria phylum¹. Here we showed that PC1 is correlated with Proteobacteria abundance, with culture positive samples clustering mostly on the one side of the PC1 axis and culture negative samples clustering mostly on the other side, together with the healthy samples.

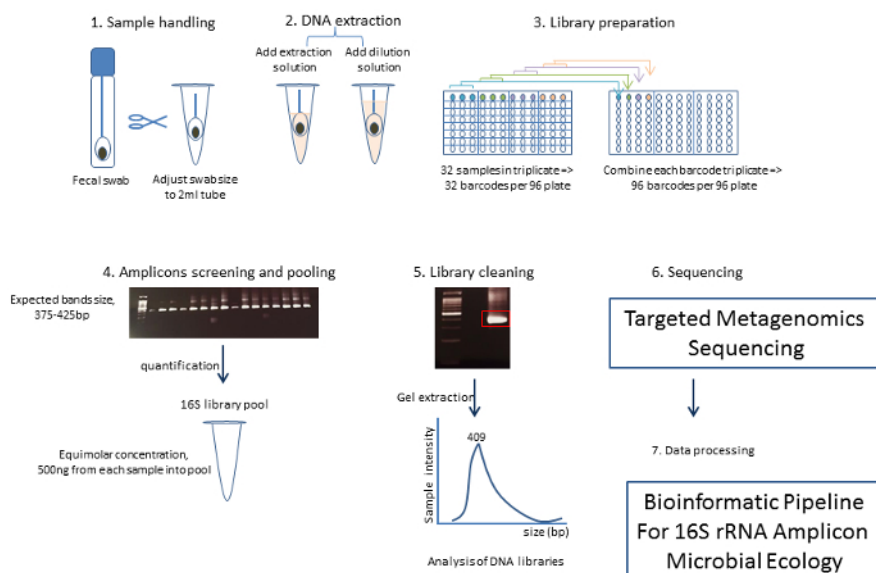


Figure 1: Flow chart from fecal samples to relative microbial abundance. **1)** Sample handling: Left, fecal samples are collected in a sterile swab tube. Right, the swab stick size is adjusted by cutting it to enable the closure of 2ml tube, and storage. **2)** DNA extraction: DNA is extracted by direct PCR columns-free approach. **3)** Library preparation: Left, 4 μ L of extracted DNA is amplified with Forward/Reverse-index primers. Each reverse primer contains a 12 base barcode. Each sample is amplified in triplicate. A 96-well plate contains 32 different barcodes. Right, combine each barcode triplicate into a single volume tube. A 96-well plate contains 96 different barcodes. **4)** Library quantification: Upper panel, screening for positive samples, amplicons were detected on Ethidium Bromide-stained 1% agarose gel. Lower panel, quantification of the positive amplicons to reach equimolar concentration of 500 ng amplicon from each sample into a single library pool. **5)** Library purification: Upper panel, the pooled library is gel extracted for size selection and purification. Lower panel, the library exact size and concentration is measured. **6)** Sequencing: high-throughput sequencing of the pooled library. **7)** The sequencing data is analyzed by a bioinformatic pipeline for 16S rRNA-amplicon microbial ecology. [Please click here to view a larger version of this figure.](#)

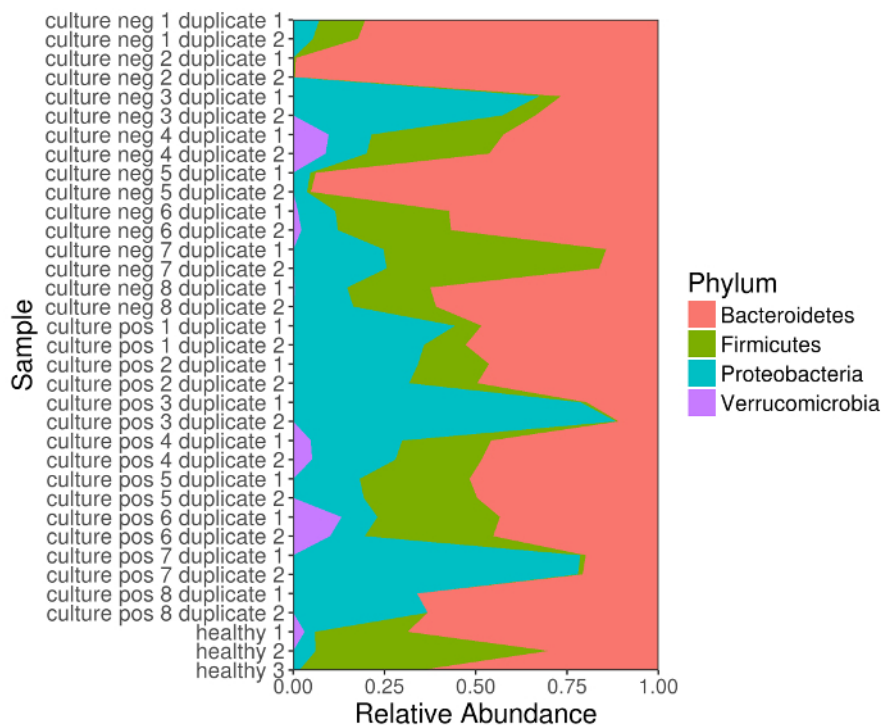


Figure 2: Consistent relative abundance between samples that originated from the same stool sample but went through different library preparation. Taxonomic relative abundance at the phylum level is shown per sample as indicated. Relative abundance is shown for 16 fecal samples obtained from hospitalized patients with suspected infectious diarrhea that each went through different library preparation (Duplicate 1 and 2) with different reverse barcoded primers and of 3 healthy controls. [Please click here to view a larger version of this figure.](#)

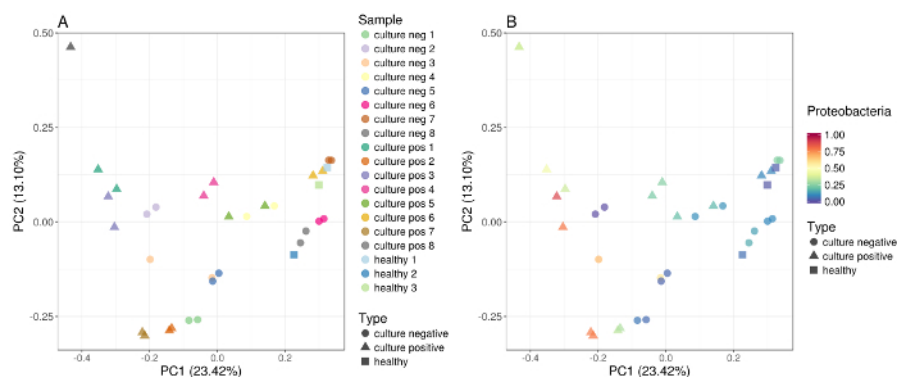


Figure 3: Phylogenetic diversity show consistent results for samples that originated from the same stool sample but went through different library preparation. Principal coordinates analysis (PCoA) that reduces dimensionality of the input data was used on the UniFrac matrix to visually explore sample separation and similarity. The distance between two samples represents the difference in their microbiome composition. (A) Unweighted UniFrac PCoA plot. Each point represents a single sequenced sample. Samples originated from the same original stool sample are colored the same. Samples from hospitalized patients with positive stool culture reported by the clinical microbiology¹ are marked by squares, hospitalized patients with culture negative in triangles¹, and healthy adults from a separate sequencing run¹ are marked with filled in circles. (B) Same PCoA as in A, but samples are now colored by their relative abundance of Proteobacteria, as indicated. [Please click here to view a larger version of this figure.](#)

| Primer name | Primer sequence | Primer description |
|--------------------------|---|---|
| Forward primer | AATGATACGGCGACCACCGAGATCTACAC - TATGGTAATT - GT - GTGCCAGCMGCCGCGGTAA | 5' adapter sequence - Forward primer pad - Forward primer linker - Forward primer |
| Reverse index primer | CAAGCAGAAGACGGCATACGAGAT - XXXXXXXXXXXX - AGTCAGTCAG - CC - GACTACHVGGGTWTCTAAT | Reverse complement of 3' adapter sequence - Golay barcode - Reverse primer pad - Reverse primer linker - Reverse primer |
| Read 1 sequencing primer | TATGGTAATT - GT - GTGCCAGCMGCCGCGGTAA | Forward primer pad - Forward primer linker - Forward primer |
| Read 2 sequencing primer | AGTCAGTCAG - CC - GGACTACHVGGGTWTCTAAT | Reverse primer pad - Reverse primer linker - Reverse primer |
| Index sequence primer | ATTAGAWACCCBDGTAGTCC - GG - CTGACTGACT | Reverse complement of reverse primer - Reverse complement of reverse primer linker - Reverse complement of reverse primer pad |

Table 1: Primer list.

| Sample | Number of reads post initial quality control in QIIME |
|---|---|
| PCR on clean swabs in extraction and dilution solution -1 | 118 |
| PCR on clean swabs in extraction and dilution solution -2 | 39 |
| PCR without template -1 | 0 |
| PCR without template -2 | 0 |
| PCR on extraction and dilution solutions mix - 1 | 16 |
| PCR on extraction and dilution solutions mix - 2 | 0 |
| Samples (mean ± standard error) | 10622.57 ± 1211.34 |

Table 2: Number of reads for negative controls and for fecal samples.

Supplementary Material 1: Data processing script. [Please click here to download this file.](#)

Supplementary Material 2: Mapping file. [Please click here to download this file.](#)

Supplementary Material 3: Primer scheme. [Please click here to download this file.](#)

Discussion

16S rRNA-amplicon and metagenomics shotgun sequencing have gained popularity in clinical microbiology applications^{21,22,23}. These techniques are advantageous in their increased ability to capture culturable and non-culturable taxa, providing data about the relative abundance of the

pathogenic inoculum, and their ability to identify more precisely a polymicrobial infectious fingerprint²⁴. The advances in the field of microbiome research have generated vast amounts of data that are not easily comparable due to a lack of consistency in the various approaches. In recent years, several consortiums have attempted to address some of these issues. This protocol follows similar library preparation as the earth microbiome project (EMP). However, added is a detailed step-by-step protocol for DNA extraction from fecal samples through analyses pipeline.

Previously, 16S rRNA sequencing has been used to characterize the microbial composition patterns of fecal samples from healthy non-hospitalized subjects and from hospitalized patients with suspected infectious diarrhea, and to traditional culture results. The results from that analysis showed that hospitalized patients have profound increases in taxa from Proteobacteria phylum¹. Described here is an integrated protocol of broadly used methods for 16S rRNA gene amplicon sequencing using the direct PCR columns-free approach that was originally designed for extraction of plant DNA¹⁶. This approach can efficiently and quickly extract good-quality amplified-DNA libraries targeting the V4 variable region of the 16S rRNA gene from a large number of samples suitable for high-throughput 16S rRNA sequencing.

As this is a DNA-based sequencing method, data on both aerobes and anaerobes microbial communities present in an individual fecal sample are obtained. The protocol adopted the primers that were originally designed¹⁵ against the V4 region of the 16S rRNA gene. Importantly, the reverse amplification primer contain a twelve base barcode sequence that supports pooling of up to 2,167 different samples in each lane¹⁵. The forward amplification primer includes nine extra bases in the adapter region that support paired-end sequencing on the sequencing platform²⁵ (**Table 1** and **Supplementary Material 3**). This approach enabled handling a library composed of 250 fecal samples that were sequenced on one lane with a running depth of 10622.57 ± 1211.34 (mean \pm standard error) reads per sample at a cost of ~30 dollars per sample.

To ensure quality control, we suggest using the following negative controls; i) PCR without DNA template, ii) PCR on DNA extracted from a clean sterile swab, and iii) PCR on mixed extraction and dilution solution. As for positive controls, we recommend including samples that originated from the same stool sample but went through different library preparation with different reverse barcoded primers and samples from previous runs as internal quality control. Of note, other optional positive controls to be used are those microbiome measurement standards provided by NIST. The average reads coverage for the negative controls is remarkably lower in comparison to the actual stool samples (**Table 1**) and was composed primarily from the Mycoplasma taxa. We further showed the consistency of the sequencing results obtained using samples that each originated from the same fecal material but that went through different library preparation (**Figure 3**). This result also confirms that, when using our integrated method, there is no cross-contamination between samples.

In this protocol we introduce an integrated uniformed, feasible protocol to analyze large numbers of samples. This protocol aims to guide scientists interested in initiating the use of 16S rRNA-amplicon sequencing in a robust, reproducible, easy to use, inexpensive, detailed way from fecal collection to data analyses, using important controls. Using this standardized detailed guided protocol, may minimize batch effects and allow more comparable sequencing results between different labs.

Disclosures

The authors have nothing to disclose.

Acknowledgements

This work was supported in part by the I-CORE program (grants No. 41/11), the Israel Science Foundation (grant No. 908/15), and the European Crohn's and Colitis Organization (ECCO).

References

- Braun, T. *et al.* Fecal microbial characterization of hospitalized patients with suspected infectious diarrhea shows significant dysbiosis. *Scientific Reports*. **7**, 1088 (2017).
- De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences of the United States of America*. **107**, 14691-14696 (2010).
- Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature*. **486**, 222-227 (2012).
- Rodriguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microbial Ecology in Health and Disease*. **26**, 26050 (2015).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*. **444**, 1027-1031 (2006).
- McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*. **6**, 610-618 (2012).
- Bakken, J. S. *et al.* Treating Clostridium difficile infection with fecal microbiota transplantation. *Clinical Gastroenterology and Hepatology*. **9**, 1044-1049 (2011).
- Khoruts, A., Dicksved, J., Jansson, J. K., & Sadowsky, M. J. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *Journal of Clinical Gastroenterology*. **44**, 354-360 (2010).
- Nood, E. V. *et al.* Duodenal infusion of donor feces for recurrent Clostridium difficile. *The New England Journal of Medicine*. **368**, 407-415 (2013).
- Koplan, J. P., Benfari Ferraro, M. J., Fineberg, H. V., & Rosenberg, M. L. Value of stool cultures. *The Lancet*. **316**, 413-416 (1980).
- Platts-Mills, J. A., Liu, J., & Houpt, E. R. New concepts in diagnostics for infectious diarrhea. *Mucosal Immunology*. **6**, 876-885 (2013).
- Bokulich, N. A. *et al.* Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*. **10**, 57-59 (2013).
- Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*. **11** (2010).
- Dubilier, N., McFall-Ngai, M., & Zhao, L. Create a global microbiome effort. *Nature*. **526**, 631-634 (2015).
- Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*. **108**, 4516-4522 (2011).

16. Flores, G. E., Henley, J. B., & Fierer, N. A direct PCR approach to accelerate analyses of human-associated microbial communities. *PLoS One*. **7** (2012).
17. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. **7**, 335-336 (2010).
18. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*. **13**, 581-583 (2016).
19. Lozupone, C., & Knight, R. UniFrac: A new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*. **71**, 8228-8235 (2005).
20. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. UniFrac: An effective distance metric for microbial community comparison. *The ISME Journal*. **5**, 169-172 (2011).
21. Srinivasan, R. *et al.* Use of 16S rRNA gene for identification of a broad range of clinically relevant bacterial pathogens. *PLoS One*. **10**, e0117617 (2015).
22. Hiergeist, A., Gläsner, J., Reischl, U., & Gessner, A. Analyses of intestinal microbiota: culture versus sequencing. *ILAR Journal*. **56**, 228-240 (2015).
23. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E. A., & Crook, D. W. Transforming clinical microbiology with bacterial genome sequencing. *Nature Reviews Genetics*. **13**, 601-612 (2012).
24. Salipante, S. J. *et al.* Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLoS One*. **8**, e65226 (2013).
25. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*. **6**, 1621-1624 (2012).