International Journal of Medical Microbiology 306 (2016) 334–342

Contents lists available at ScienceDirect

International Journal of Medical Microbiology

journal homepage: www.elsevier.com/locate/ijmm

Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability

Andreas Hiergeist, Udo Reischl, Priority Program 1656 Intestinal Microbiota Consortium/quality assessment participants, Andrè Gessner*,[1]

Institute of Clinical Microbiology and Hygiene, University of Regensburg, Regensburg, Germany

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form 3 March 2016
Accepted 7 March 2016

Keywords:
Microbiome
Next-generation sequencing
External quality assessment
16S rDNA
Proficiency testing

ABSTRACT

The composition of human as well as animal microbiota has increasingly gained in interest since metabolites and structural components of endogenous microorganisms fundamentally influence all aspects of host physiology. Since many of the bacteria are still unculturable, molecular techniques such as high-throughput sequencing have dramatically increased our knowledge of microbial communities. The majority of microbiome studies published thus far are based on bacterial 16S ribosomal RNA (rRNA) gene sequencing, so that they can, at least in principle, be compared to determine the role of the microbiome composition for host metabolism and physiology, developmental processes, as well as different diseases. However, differences in DNA preparation and purification, 16S rDNA PCR amplification, sequencing procedures and platforms, as well as bioinformatic analysis and quality control measures may strongly affect the microbiome composition results obtained in different laboratories. To systematically evaluate the comparability of results and identify the most influential methodological factors affecting these differences, identical human stool sample replicates spiked with quantified marker bacteria, and their subsequent DNA sequences were analyzed by nine different centers in an external quality assessment (EQA). While high intra-center reproducibility was observed in repetitive tests, significant inter-center differences of reported microbiota composition were obtained. All steps of the complex analysis workflow significantly influenced microbiome profiles, but the magnitude of variation caused by PCR primers for 16S rDNA amplification was clearly the largest. In order to advance microbiome research to a more standardized and routine medical diagnostic procedure, it is essential to establish uniform standard operating procedures throughout laboratories and to initiate regular proficiency testing.

1. Introduction

The analysis of human body microbial ecology by high-throughput sequencing has the potential to revolutionize nearly every field of medicine. In the past few years, the number of culture-independent, sequencing-based investigations and publications on the human and mouse microbiome has significantly increased, now representing one of the most studied and interesting fields in medicine, and one having the highest potential impact on clinical practice. A large spectrum of disease phenotypes has been linked to the composition of microbiota: chronic inflammatory diseases, obesity, diabetes, allergies, cardiovascular diseases, some cancer types, lung diseases and even psychiatric illnesses such as autism and depression have been reported to occur concomitantly with a distinct microbiome composition (Sekirov et al., 2010; Shreiner et al., 2015; Marsland and Gollwitzer, 2014). Although the knowledge to date of a causative or curative role for any of the microbial members detected in these approaches is still very limited, it can be expected that microbiome data obtained from gut, lung, mouth and other body sites will serve as an important diagnostic or prognostic biomarker for human diseases in the near future. However, microbiome studies published at an exponential rate over the past several years have often been difficult to reproduce across investigations. This may be due to both relevant variation in the methodology used for microbiome analysis, or significant differences in the patient cohorts studied. One of the most widely applied current methods for microbiota profiling is based on the sequencing of a very important and convenient bacterial gene, known as the 16S rRNA gene (Olsen et al., 1986). The complex

* Corresponding author at: Institute of Clinical Microbiology and Hygiene, University Hospital Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany.
E-mail address: andre.gessner@ukr.de (A. Gessner).
[1] For details of the collaboration please see Appendix A.

http://dx.doi.org/10.1016/j.ijmm.2016.03.005

multistep process of 16S rDNA-based microbiome analysis involves physical specimen collection, DNA extraction and purification, PCR and amplicon purification or enrichment, sequencing and bioinformatics. Addressing the sources of variation in each of these steps, and determining their effect on the final result, is critical in helping to standardize and optimize the entire process in the future.

In general, medical laboratory diagnostics have traditionally emphasized the achievement of between-laboratory consensus by external quality assessments, which play a unique role in providing objective data for tests performed in different laboratories under routine conditions, and can guide improvement in the use of various diagnostic reagents and laboratory practices. For microbiome profiling, no data on multicenter quality assessment have been published thus far. Therefore, with the help of INSTAND e.V., a German WHO-collaborating center for quality assurance and standardization in laboratory medicine, we organized the first microbiome quality assessments in 2014 and 2015 with participating laboratories from Austria, Germany and Norway.

2. Material and methods

2.1. Sample preparation and shipment

To properly address all of the steps of gut microbiome analysis in a realistic manner, we decided not to use defined bacterial mixtures or chemostat cultured microbiome cultures, but instead homogenized and spiked human fecal samples. A total of 1.5 g of fresh formed stool from a healthy human donor were collected in a sterile container and subsequently suspended in 4.5 ml of sterile PBS buffer. The stool suspension was homogenized by rigorous mixing and subsequently divided into 40 × 100 μl aliquots, corresponding to 33.3 mg of wet weight stool. In order to investigate the recovery performance of preset bacterial ratios by all participants, every sample was spiked with quantitative amounts of three exogenic eubacterial species *Salinibacter ruber* DSM 13855^T (1E + 06 16S rDNA copies per sample), *Rhizobium radiobacter* DSM 30147^T (3E + 08 16S rDNA copies per sample) and *Alicyclobacillus acidiphilus* DSM 14558^T (6E + 07 16S rDNA copies per sample). Total 16S rDNA copy numbers for spike-in bacteria were calculated from measured optical densities and genomic 16S rRNA gene copy numbers, which were obtained from the rrnDB 16S copy number database (Lee et al., 2009). All bacteria were purchased from the DSMZ (German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany). One additional sample served as a non-spiked control sample. Spiked stool suspensions were stored at −80 °C until shipment to the nine participating laboratories. Samples were shipped from the EQA center in Regensburg as duplicates on dry ice without thawing. The participating institutions received the duplicate stool samples together with a 100-ng aliquot of DNA extract (at a concentration of 10 ng/μl) originating from the same individual sample (deposited as sample number R02 in sequencing raw data analysis).

2.2. Extraction of nucleic acids

Four nonconsecutive, randomly picked stool samples from the entire EQA set were selected for DNA extraction and subsequent homogeneity testing. Samples were thawed on ice and DNA was extracted from the total material content of the sample tubes using Proteinase K pretreatment followed by mechanical disruption. Therefore, samples were subjected to five consecutive freeze/thaw cycles between liquid nitrogen and boiling water and repeated bead beating. DNA was subsequently isolated from the lysates using the MagNA Pure 96 system (Roche).

2.3. Amplification of V3–V6 16S rDNA hypervariable regions and 454-pyrosequencing

For homogeneity testing, the hypervariable V3–V6 regions of the 16S-rRNA gene were amplified from the DNA extracts of four randomly picked samples through 30 PCR cycles using *Platinum* Taq polymerase (Life Technologies) and barcoded fusion primers 341F/1061R containing Titanium/Lib-L adaptors. PCR products (790 bp) were purified from agarose gels (QIAquick PCR Purification Kit, Qiagen) and repurified with Agencourt AMPure XP Beads (Beckmann-Coulter). DNA copy numbers of the amplicons containing Titanium/Lib-L adaptors on both ends were determined using the KAPA Library Quant Real-time PCR Kit (KAPA Biosystems) and pooled at an equimolar ratio. The amplicon library was subjected to sequencing on a Roche 454 GS Junior+ and the GS FLX+ system in parallel using GS FLX Titanium XL+ chemistry with an acyclic flow pattern. Flowgrams were denoised and low quality reads were filtered using the FlowClus software package v1.1 with standard parameters (Gaspar and Thomas, 2015). Sequences longer than 400 and shorter than 800 bp after quality filtering were demultiplexed by barcode sequences. Quality filtered raw data reads were processed as described in the raw data analysis section below. Datasets were included as participant numbers P1 (GS Junior +) and P7 (GS FLX +) to the comparative dataset (P1 to P9). Both datasets were generated starting from individual 16S rDNA amplicon libraries.

2.4. Quantification of 16S rDNA copy numbers by qRT-PCR

Amounts of spike-in bacteria *A. acidiphilus*, *S. ruber*, and *R. radiobacter* were assayed in fecal DNA preparations by a 16S rDNA targeted qRT-PCR on a LightCycler 480 II Instrument (Roche). Species-specific primers and hydrolysis probes and LightCycler 480 Probes Master reagents were used for detection. Full length 16S rDNA amplicons cloned into pGEM-T.Easy vector (Invitrogen) served as the quantification standard.

Total 16S rRNA gene copy numbers were determined using universal eubacterial 16S rDNA primers 764F and 907R, and SYBR Green I Master (Roche) qPCR reagents. 16S rRNA gene copy numbers were calculated from plasmid standards of cloned full-length 16S rDNA genes into the pGEM T-Easy. Primer names and sequences are listed in Supplementary Table 4.

2.5. Reporting of results

Each of the nine participating centers analyzed the microbial communities according to their established protocols starting from DNA extraction to visualization of final results. Participating laboratories were asked to send their final analysis report together with their raw sequencing files in fastq- (Illumina), sff- (454) or BAM- (IonTorrent) file format. A questionnaire was provided to obtain methodological details for DNA extraction, PCR amplification including 16S rDNA amplification primer sequences, bioinformatic pipelines used for data analysis, as well as 16S rRNA reference databases.

2.6. Computational analyses of sequencing raw data

Illumina paired end reads were joined using fastq-join (Aronesty, 2011) and then filtered with QIIMEís split_libraries_fastq.py workflow script using default parameters, with the exception of applying a phred quality threshold of 19. IonTorrent (P2) data and 454 data generated with Titanium XLR70 chemistry (P9) were denoised with FlowClus using default parameters, despite removing reads shorter than 100 or 300 and longer than 300 or 600, respectively. For all raw data sets USEARCH v6.1 (Edgar, 2010) was used to remove chimeric 16S

rDNA sequences from filtered reads. Downstream analyses of quality and chimera filtered reads for participants P1 to P9 were performed using the QIIME 1.9.1 software package (Caporaso et al., 2010). Default parameters and software packages were used unless otherwise stated. Each of the quality filtered sequencing read datasets for participants P1 to P9 were separately assigned to operational taxonomic units (OTUs) with a threshold of 97% pairwise identity using QIIMEís reference-based workflow scripts and the SILVA release 119 16S-rRNA reference database. Afterwards, participant P1 to P9 OTU tables were merged. No additional filtering threshold for low abundant OTUs was applied in order not to circumvent the detection of very low abundant marker bacteria. The relative abundance of eubacterial sequences were calculated for each sample (R01 to R34) at each taxonomic level to identify bacterial taxa. R version 3.2.3 (R Developement Core Team, 2008) was used for subsequent analyses of OTU tables and dtables and data visualization. The vegan v2.3-2 (Oksanen et al., 2015) package was used for ordination analyses based on Bray-Curtis dissimilarities. Ordination ellipses were created by calculating the 95% confidence intervals around the centroid. Database coverage of primer pairs from each participant were evaluated using TestPrime 1.0 (Klindworth et al., 2013) against the SILVA SSU release 123 16S ribosomal RNA database, allowing no primer mismatches. We performed the Adonis test over the Bray-Curtis distances using the adonis function of vegan to evaluate sources of variations among cagegories: participants, 16S hypervariable V-regions, sequencing technology, DNA extraction method. The number of permutations was set to 999.

3. Results

3.1. Participant questionnaires

In order to compare the performance of nine different clinical and basic research laboratories in performing 16S rDNA amplicon-based microbiome analysis, we produced EQA sample sets comprised of duplicate homogenized human stool samples spiked with quantitative amounts of three defined marker bacteria. To further assess the level of methodological variance, identical aliquots of a DNA extract prepared from one individual EQA sample were also delivered to the participants, including a questionnaire focused on the methodological aspects of their standard operating procedures for microbiome profiling. Table 1 summarizes the recorded data. Participants who had different levels of practical experience in performing microbiome analyses reported a wide range of practices and approaches starting with initial DNA extraction to 16S rDNA-based PCR amplification to the final data analysis pipelines. Five participants applied pretreatment steps for DNA isolation, including repeated bead beating protocols or enzymatic pretreatment followed by a manual or automated DNA purification workflow. Six of the nine participants implemented data analysis pipelines based on mothur (Schloss et al., 2009) or QIIME (Caporaso et al., 2010) software packages, while two others used combinations of available open source and proprietary software tools, and one participant reported using solely proprietary tools for analysis of sequencing reads. Different releases of the SILVA (Quast et al., 2013), Greengenes (DeSantis et al., 2006) or Ribosomal Database Project (RDP, Cole et al., 2014) 16S ribosomal databases were used by all participants for sequence alignment or taxonomy assignment. With the GS Junior+, the 454 GS FLX+ system (both Roche/454), the Illumina MiSeq (Illumina) and IonTorrent PGM (Life Technologies) instruments, all major next-generation-sequencing (NGS-) technologies presently used for 16S rRNA gene based microbiome analyseswere covered.

Table 1
Participant questionnaire details. Methodological details for microbiome profiling as reported by participants P1 to P9. Length and number of reads were obtained from reanalyzed datasets. BB = Bead- Beating, E = Enzymatic, FT = Freeze/Thaw, C = Commercial Kit, PhCl = Phenol/Chloroform, M = Manual, A = Automated; PE = Paired End, SE = Single End; n.r. = not reported; QC = quality control.

Participant	P1	P2	P3	P4	P5	P6	P7	P8	P9	
DNA Extraction	BB, E, FT,C, A	C,A	C,A	n.r.	BB, C, M	BB,C, M	BB, E, FT, C, A	PhCl,M		*BB, E, C, M*
Amplified 16S rDNA Variable Region	V3-V6	V4	V3-V4	V4	V1-V2	V3-V4	V3-V6	V3-V4	V1-V3	
Universal 16S rDNA Primers	341f/1061r	533f/806r	341f/806r	518f/805r	27f/338r	341f/785r	341f/1061r	341f/805r	8f/541r	
No. of PCR Cycles	30	n.r.	25	n.r.	30	25	30	30	33	
Proofreading	No	n.r.	yes	n.r.	yes	yes	n.r.	yes	yes	
Sequencing Platform	454/GS Junior+/XL+ Kit	IonTorrent PGM	Illumina MiSeq	Illumina MiSeq	Illumina MiSeq	Illumina MiSeq	454/GS FLX+/XL+ Kit	Illumina MiSeq	454/GS FLX+/ Titanium XLR70 Kit	
Sequencing Mode	SE	SE	PE	PE	PE	PE	SE	PE	SE	
Mean Read Length after QC –(obtained from reanalyzed data)	666	215	445	252	320	405	621	405	404	
Mean reads per Sample–(obtained from reanalyzed data)	13,870	21,024	246,333	15,805	27,661	40,280	7820	58,852	25,606	
Data Analysis Pipeline	Qiime 1.9.1	Qiime 1.8.0	MiSeq Analysis Software (2.4.60.8)	Phyloseq	mothur 1.33.3	In-house pipeline based on Uparse	QIIME 1.9.1	Qiime 1.8.0	, usearch7.0.1090 and mothur 1.36.1	
16S rRNA Reference Databases	SILVA 119	Greengenes 13.08	Greengenes 13.08	Greengenes	mod. SILVA 102, RDP9	RDP 9	SILVA 119	SILVA 111	RDP classifier 2.10.1, default training set; SILVA LTPs 121 and modified RDP 11.2	

Each participant was asked to analyze at least one of the stool samples sent together with the already extracted DNA. Eight of nine participants reported results for the DNA sample, while participant P4 failed to generate PCR amplicons from the extracted DNA. Replicate testing of extracted DNA was performed by some of the participants. All reported data from repetitive testing was included in the subsequent data analysis. Assignment of analyzed samples to reported results can be found in Supplementary Table 2.

3.2. QRT-PCR analyses of spiked-in marker bacteria and sample homogeneity

Four nonconsecutive samples from the entire sample set were randomly chosen for homogeneity testing using qRT-PCR-based determination of 16S rRNA gene copy numbers for total and spiked marker bacteria. *A. acidiphilus and R. radiobacter* were included to assess the recovery of defined quantitative ratios by the participants' standard operating procedures for NGS-based microbiota profiling. *S. ruber* was spiked to a total of 1.0E+06 rRNA gene copies per sample to survey the capacity for low abundant species detection by the different sequencing platforms, since we expected a platform-inherent variation in sequencing depth. An average of 9.2E+09 ± 1.0E+09 (arithmetic mean ± standard deviation) total eubacterial 16S rRNA gene copies were determined per sample. Inter-sample variation in measured 16S rDNA copies was lower than 10% of the mean. An average of 5.7E+07 ± 1.2E+07 (*A. acidiphilus)*, 2.7E+08 ± 3.3E+07 (*R. radiobacter*) and 6.8E+05 ± 2.1E+05 (*S. ruber*) were measured for spiked-in bacterial marker species, while the non-spiked control sample tested negative for each of the spiked-in bacterial species. The relative abundance calculated from qRT-PCR-determined 16S rDNA copy numbers was 0.6, 2.9 and 0.007% for *A. acidiphilus*, *R. radiobacter* and *S. ruber*, respectively.

Homogeneity of microbiota profiles examined at the EQA center by sequencing of amplified V3–V6 variable 16S rDNA regions using two different sequencing platforms (included as participants P1 and P7 to the data set) showed a high degree of similarity. Microbiota profiles determined at the genus level are depicted in Fig. 1. Relative abundance profiles of individual taxa corresponded very well both within and between the GS Junior+ (P1) and the GS FLX+ (P2) datasets.

3.3. Microbial community profiling by 16S ribosomal RNA amplicon sequencing

Taxa distributions reported from the participants according to their standard operating procedures for data analysis and distributions reanalyzed from sequencing raw data were compared on the family level for the eleven most frequently occurring taxa. In order to eliminate sources of bias introduced by varying data analysis pipelines and post-sequencing steps, microbiota profiles were additionally studied by a joint analysis of raw sequencing data provided by all participating laboratories (P1 to P9). Analyzed raw data reads per sample (Table 1) varied from 7298 (R26, P7, GS FLX+) to 371.900 (R11, P3, Illumina MiSeq). An average of 1077 OTUs per sample were detected, with a maximum of 2.226 (R32, P9, GS FLX+/Titanium XLR70 chemistry) and a minimum of 464 (R25, P7, GS FLX+/XL+) identified OTUs. The number of OTUs analyzed within each participating group showed a high degree of similarity, indicating a high intra-center reproducibility. However, when compared between centers, distributions of taxonomically assigned OTUs varied tremendously from the phylum to genus level. Fig. 1 graphically displays the relative abundance of identified taxa on the genus level. While almost all taxa were affected by inter-center deviations, particularly bacterial species of the genus *Bifidobacterium* were detected at a level of high abundance (16%, P4) to complete absence in participating groups P5 and P9. *Prevotella*, which was on average the most prevalent genus in the P2 dataset, varied from highest abundance (40%, R09, P2) to a minimum abundance of 0.15% (R30, P8).

When comparing taxa profiles between datasets originating from stool or DNA samples (marked with asterisks in Fig. 1) within one participating group, differences between sample types were, in general, less pronounced than inter-center deviations. However, varying effects of different DNA extraction methods (Table 1) implemented by participants on obtained microbiota profiles were evident. Effects were analyzed in more detail in a later section.

Furthermore, wide discrepancies were reported for the relative abundance of individual classified taxa. Supplementary Table 1 comparatively summarizes the relative abundance obtained on the family level. Highly abundant families Ruminococcaceae and Lachnospiraceae, both members of phylum Firmicutes, were reported by all participants, similarly to taxonomic assignment after raw data analysis. Significant differences between reported and reanalyzed relative abundance were observed for the families Bacteroidaceae, Veillonellaceae, Acidaminococcaceae and Peptostreptococcaceae. Reads assigned to the family Enterococcaceae were reported only by participant P5 with a relative abundance of 9.25%. Surprisingly, the presence of reads classified as Enterococcaceae could not be verified from raw data reads, indicating marked differences in applied downstream bioinformatic analyses.

3.4. Detection of spiked eubacterial marker species

Homogenized stool samples were quantitatively spiked with grown cells of three eubacterial marker species *R. radiobacter, A. acidiphilus and S. ruber* at fixed 16S rDNA copy numbers. Ratios of *R. radiobacter*/*A. acidiphilus* 16S rRNA gene copies (RA ratio) as quantified by qRT-PCR was 4.7. To assess whether preset spike-in ratios were accurately reproduced by the participantís standard operating procedures, spike-in ratios were calculated from the relative abundance of taxonomically assigned OTUs clustered from reanalyzed raw sequencing data (Fig. 4A). DNA sequences of universal 16S rDNA targeted amplification primers provided by the participants were evaluated for the coverage of spiked bacterial species, and showed no mismatches when compared to full-length 16S rDNA sequences of *S. ruber, R. radiobacter* and *A. acidiphilus* obtained from the Silva 123 release.

The maximum measured ratio was 21.6 (stool sample R13, participant P4) while the minimum was 0.37 (stool sample R16, participant P5). Expected *Rhizobium* to *Alicyclobacillus* ratios (4.7) were best recovered by participants P1, P2, P6 and P7, who determined mean ratios of 4.8, 4.4, 6.8 and 4.8, respectively. In general, ratios lower than expected were observed for participant P5 (mean RA ratio of 1.4) due to underestimation of *R. radiobacter* reads, while P3 (RA ratio 13.1) and P4 (RA ratio 19.3) showed high ratios caused by depreciation of *A. acidiphilus* reads. To assess detection of low abundant species, cultivated cells of *S. ruber* were spiked into each sample at 1E+06 16S rRNA gene copies. This number of cells was estimated to be around the detection threshold of metagenomics-based methods described by Lagier et al. (2012). However, less than half of the participating laboratories (P1, P2, P3 and P5) were successful in detecting *S. ruber* by 16S rDNA-based community profiling (Fig. 4B). No significant correlation between sequencing depth and rare spike-in detection was observed.

3.5. Beta diversity between participants, sequencing technology and amplified hypervariable 16S rDNA regions

To investigate causes that led to potential inter-laboratory deviations between community profiles, beta diversity was analyzed by non-metric multidimensional scaling of Bray-Curtis dissimilari-

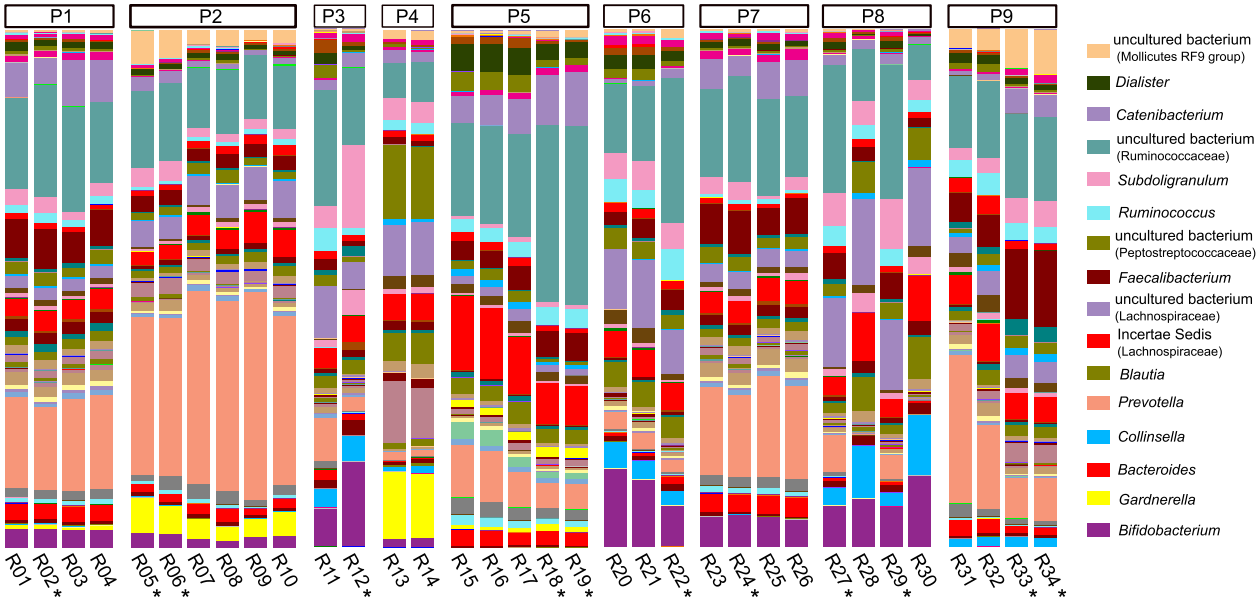

Fig. 1. Community composition reveals high inter-center variability. Relative genus-level distribution of assigned operational taxonomic units (OTUs) for shipped stool and DNA samples as revealed by reanalysis of the quality assessment raw data for each participant (P1 to P9). Closed reference OTUs were picked against the SILVA release 119 16 rRNA database. Results (R01-R34) generated from shipped DNA samples are indicated by asterisks (See full legend in Supplementary Fig. 1. The underlying OTU Table is deposited as Supplementary Table 3).

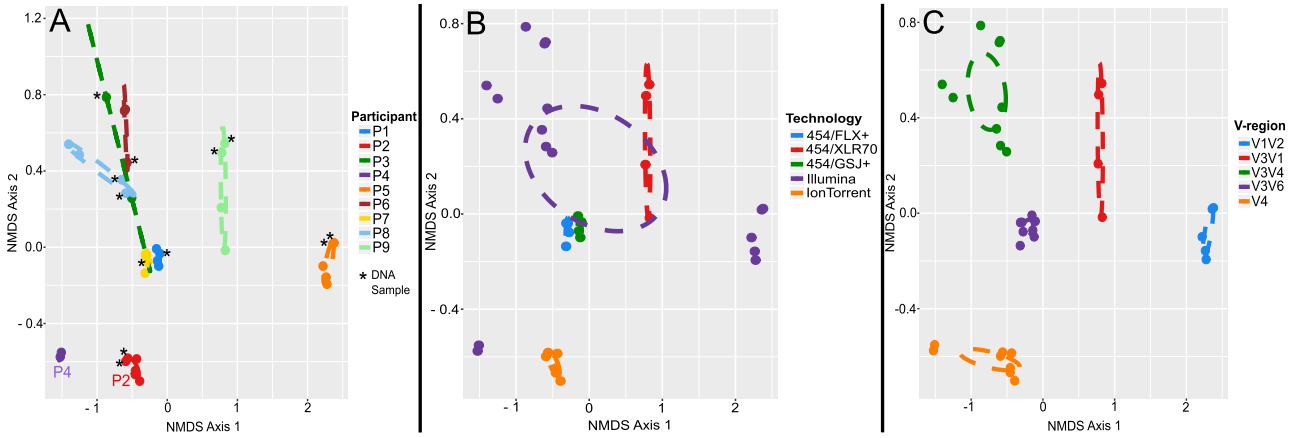

Fig. 2. Ordination analyses. Two dimensional nonmetric multidimensional scaling (NMDS) plots based on the Bray-Curtis measure for participants P1 to P9. Dissimilarities were calculated from the relative abundance of assigned operational taxonomic units (OTUs) based on the reanalyzed quality assessment raw data. Ellipses indicate the 95% confidence interval around cluster centroids. Ordination was performed based on groups: Participants (A), used sequencing technology (B) and amplified variable regions (Vregion) of the 16S rRNA gene (C). Asterisks denote data points originating from DNA samples. The stress value is 0.073 ($R^2 = 0.995$).

ties, which were calculated from the relative abundance of assigned OTUs. Ordination analyses showed that the communities clustered perfectly according to the participating laboratories (Fig. 2A). These findings again reflect a high degree of intra-center reproducibility (compare Fig. 1). Effects of DNA extraction methods on the ascertained community structure were visible by a greater ordination distance of DNA sample dissimilarities to the data points derived from stool samples within one participating group, while technical replicates from the same stool samples were clustered closely together (Fig. 2A). DNA samples are marked with asterisks). From this perspective, participants P3, P8 and P9 showed the largest variability. Compared to the DNA sample extraction method, participant P3 reported automated DNA extraction without any pretreatment step, while participant P8 used a manual phenol-chloroform extraction method. Participant P9 used repeated bead beating followed by enzymatic digestion but used manual DNA extraction. Sequencing technology itself does not seem to have a clear direct impact on the sample dissimilarities (Fig. 2B) since Illumina based results were irregularly distributed over the entire ordination plot. However, inter-laboratory variations were best explained by covered hypervariable (V-) regions and primers which were selected for 16S rDNA-based amplicon sequencing (Fig. 2C) according to their compatibility with sequencing platform requirements. This was further confirmed by the statistical analysis using an Adonis test. EQA participants ($R2 = 0.84$, $p = 0.001$) and amplified V-regions ($R2 = 0.67$, $p = 0.001$) were both factors which explained the highest proportion of variation (Table 2) in the ordination analysis causing significant changes in the analyzed microbial taxa.

3.6. Evaluation of SILVA release 123 16S ribosomal database primer coverage

Universal 16S rDNA targeted primers used in this study were reported by the participating laboratories (Table 1). Coverage of eubacterial database records varied significantly among the participants when 91.2 (P1, P7), 77.2 (P2), 86.0 (P3, P6), 73.9 (P4), 79.8

Table 2
Summary statistics of Adonis test measures among tested groups. The analyzed groups: EQA participants and amplified hypervariable 16S rDNA regions explained 84% and 67% of the data varation, respectively.

Test Category	R2 value	p-value
Participant	0.84	0.001
V-region	0.67	0.001
DNA extraction method	0.57	0.001
Sequencing technology	0.47	0.001

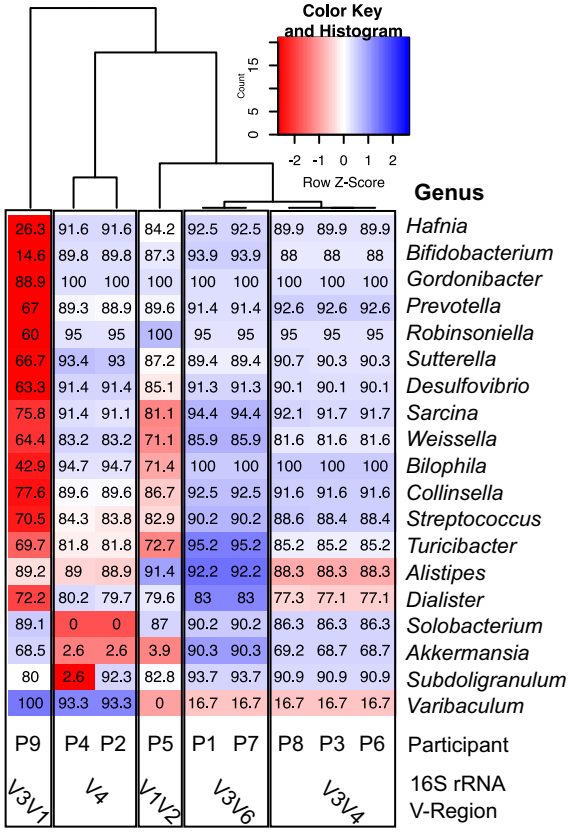

Fig. 3. Evaluating the SILVA 123 release coverage of universal 16S rDNA target primers. Heatmap and hierarchical clustering dendrogram based on the SILVA release 123 16S rRNA sequence database coverage of PCR amplification primers as used by the participants (P1 to P9). Numbers and row centered heatmap colors indicate the percentage of database coverage of PCR primers for nineteen selected high abundant genera evaluated using TestPrime 1.0. Black boxes indicate Dendrogram clusters.

(P5), 85.3 (P8) and 60.2 (P9) percent coverage of the current SILVA release 123 16S database was evaluated using TestPrime. However, overall coverage was significantly increased by allowing one or two primer mismatches. Only one pair of amplification primers allowed considerable coverage (64.8% database coverage, P6) of archaeal species. However, no reads were assigned to archaeal taxa in the entire dataset.

Considering the database coverage of the participants' primer sequences for most frequently occurring genera in the dataset, hierarchal cluster analyses from calculated Pearsonís correlation coefficients were performed to identify possible causes of inter-center variance based on the amplified variable 16S regions (Fig. 3). Generated clusters were in perfect accordance to cluster groups identified by NMDS ordination analyses (see Fig. 2C) based on hypervariable 16S regions amplified by the participants. Results from participating laboratories P3, P6 and P8 (V3–V4 variable 16S regions) as well as P2, P4 (V4) or P1 and P7 (V3–V6) are grouped according to their amplified V-regions, while P5 (V1–V2) and P9 (V3–V1) build separate clusters. Thus, coverage of amplification primers used to bacterial taxa present in the examined microbiota are conceivably a major source of inter-center variance in this study.

3.7. Variations of microbiota profiles on the phylum level using different DNA preparation protocols

Extracted nucleic acid from one individual stool sample from the EQA sample set was sent to the participating laboratories in order to investigate the effects of different DNA extraction methods on the microbiota community structures obtained. The various institutions used a wide variety of different methodological approaches for DNA extraction, with or without implemented mechanical and enzymatic pretreatment steps, followed by a manual or automated nucleic acid purification from crude extracts (Table 1). For both DNA and stool samples, means of the relative abundance of OTUs classified to the phylum level were compared from reanalyzed raw sequencing reads. Participants P3, P8 and P9 showed the greatest variation in relative phylum abundance between DNA and stool samples, while differences for other participants were less prominent. For participant P3, who performed automated DNA extraction without any pretreatment step, reanalysis of raw sequencing data revealed a reduced amount of Actinobacteria in DNA samples (13%) compared to extracted stool samples (24%), with a simultaneous increase in Bacteroidetes (14.8% compared to 6.8 between DNA and stool samples, respectively) and Firmicutes (70.3–66.2%, respectively). Participant P8, who carried out manual phenol-chloroform extraction of DNA from stool specimens, reported a greatly reduced relative reduction in phyla Bacteroidetes (1.0 compared to 8.7%) and Actinobacteria (24.6 compared to 12.43%) in stool samples when compared to DNA sample types, while Firmicutes was increased in DNA samples (73.9 and 77.3%, respectively). A decrease in the relative abundance of the genus *Prevotella,* a member of the phylum Bacteroidetes family, is mainly responsible for the observed approx. 8-fold reduction in this phylum. Participant P9 conducted similar pretreatment steps for DNA preparation compared to the EQA center, but used a manual vs. an automated purification procedure. In this case, the manual purification method possibly led to a greater intra-center variance since *R. radiobacter* to *A. acidiphilus* spike-in ratios varied on a large scale (Fig. 4A), as did the phylum abundance. Comparing means of the relative abundance in DNA and stool sample types revealed an enrichment of OTUs from the phylum Bacteroidetes (32.5% compared to 13.3%, respectively) between stool and DNA samples, while Firmicutes showed a significantly higher abundance in DNA (74.4%) compared to stool samples (61.0%) Table 3.

4. Discussion

In this study, we evaluated the performance of independent laboratories, which had developed and implemented individual operational procedures for microbial community profiling using 16S rDNA amplicon-based next-generation-sequencing techniques. Our primary aim was to assess the comparability of microbiome analyses by implementing an external quality assessment scheme. By sending homogenized and spiked stool samples, we ensured that every result obtained originated from only one uniform sample. In addition, we minimized bias introduced by sampling and storage procedures, which are known to have a decisive influence on the microbial composition in analyzed specimens (Dominianni et al., 2014; Bahl et al., 2012). We found that

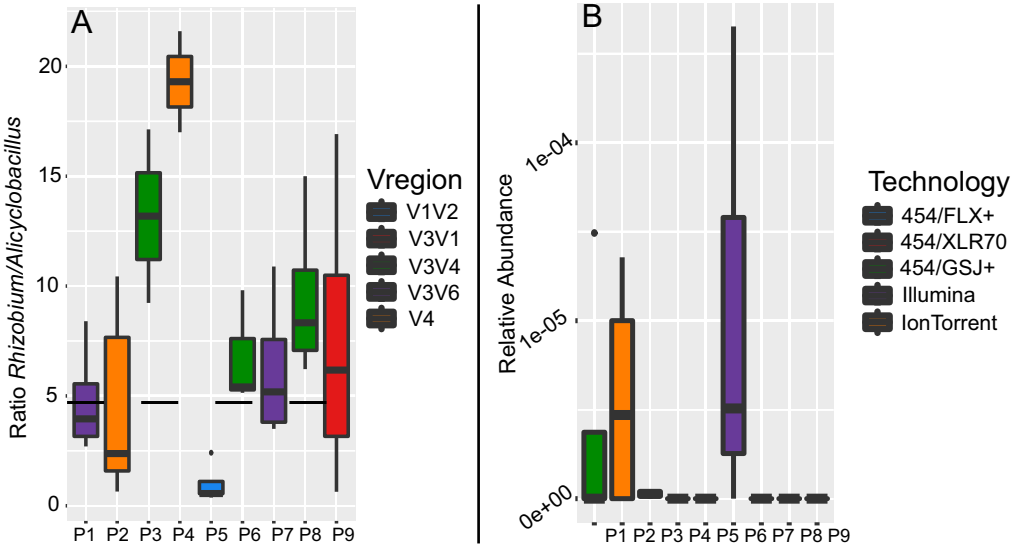

Fig. 4. Detection of spiked eubacterial marker species by EQA participants. Boxplot of *R. radiobacter* to *A. acidiphilus* ratios (A) per participant calculated from the relative abundance of taxonomically assigned OTUs on the genus level. The black dashed line indicates the expected *R. radiobacter/A. acidiphilus* ratio as quantified by qRT-PCR. Boxplot of taxonomically assigned OTUs to genus *Salinibacter (B) per participant*. Data was obtained from reanalyzed raw sequencing data of participating institutions (P1 to P9).

Table 3
Variation of the relative abundance between analyzed DNA and stool samples. Calculated relative abundance from assigned OTUs at the taxonomic level of phylum for analyzed DNA and stool samples. Nucleic acids from DNA samples were extracted at the EQA center, while DNA from stool samples were extracted at the participating institute. n.d. = not detected.

Phylum	Actinobacteria		Bacteroidetes		Cand. Division TM7		Firmicutes		Proteobacteria		Tenericutes	
Mean Relative Abundance (%)	Stool	DNA	Stool	DNA	Stool	DNA	Stool	DNA	Stool	DNA	Stool	DNA
P1	5.28	5.57	29.56	26.36	0.06	0.04	63.39	66.69	1.32	1.02	0.38	0.33
P2	6.44	9.95	44.36	37.95	0.01	0.01	45.98	45.42	0.72	0.86	2.5	5.83
P3	13.02	24.55	14.77	6.76	0.05	0.06	70.33	66.21	1.74	2.2	0.1	0.22
P4	16.87	n.d.	2.43	n.d.	0.01	n.d.	78.14	n.d.	0.82	n.d.	1.74	n.d.
P5	0.43	0.51	22.46	15.29	0.01	0.02	75.79	82.31	0.74	0.56	0.55	1.3
P6	19.42	12.12	6.92	6.31	0.01	0.04	70.72	78.14	2.5	1.78	0.4	1.6
P7	6.33	6.67	30.84	27.5	0.03	0.05	60.75	64.23	1.67	0.95	0.37	0.59
P8	24.62	12.43	0.96	8.74	0.01	0.02	73.86	77.32	0.37	1.22	0.05	0.22
P9	2.41	2.5	32.5	13.27	0.01	0.02	60.97	74.44	0.27	1.64	3.78	8.09

the reported relative abundance from the phylum to the genus level exhibited an unexpectedly high degree of compositional differences between the participating centers, while only slight intra-center deviations were observed. Procedures for data analysis led to substantial differences as shown by comparison of reported and reanalyzed data. In particular, this was impressively demonstrated in our study when OTUs that were classified to the genus *Enterococcus* by one participant reached up to nine percent relative abundance, while reanalyzed and fellow participants showed no presence. Various in-depth studies have already comprehensively identified sources of errors introduced by the use of data analysis pipelines (D'Argenio et al., 2014), which may be introduced during initial quality filtering of raw sequencing reads (Schloss, 2010), OTU clustering (Chen et al., 2013; Schmidt et al., 2014), quality and composition of 16S ribosomal databases, or in general by improper application of implemented algorithms and metrics.

Eliminating potential bias introduced by the wide variety of downstream bioinformatic procedures implemented, we aimed to increase comparability by uniformly reanalyzing the entire EQA data sets from next-generation-sequencing raw data provided by the participants. Potential methodological distortion of community profiles may still arise from DNA extraction (Yuan et al., 2012), unequal amplification of individual 16S rRNA genes during PCR amplification (Pinto and Raskin, 2012; Suzuki and Giovannoni, 1996), generation of sequencing libraries (van Dijk et al., 2014) or technologically inherent error characteristics of next-generation-sequencing platforms (Liu et al., 2012; Luo et al., 2012; Salipante et al., 2014). However, large inter-center variability was still present in our study after reanalysis of EQA raw data. By sending DNA samples, readily extracted from the EQA sample set, we could further show that different DNA extraction methods had, as expected, a specific influence on the observed microbiota profiles. As analyzed by NMDS ordination analyses, the selection of amplification primers and the covered hypervariable 16S rDNA regions deduced from them were determined to be the probable main cause of difference between centers within the EQA dataset. Individual coverage of certain taxa by selected DNA sequences of universal 16S rDNA primers is probably one main decisive factor, since hierarchal clusters calculated from SILVA release 123 database coverage of primers perfectly matched ordination groups based on covered variable 16S rDNA regions. Tremblay et al. (2015) analyzed mock communities using different primer sets targeting the V3, V6–V8 and V7–V8 hypervariable regions using the Illumina MiSeq and the 454 Titanium FLX instrument. In accordance with our findings, observed relative abundance varied significantly between mock communities amplified with different primer sets, while sequencing platforms had only a minor impact. In the same manner, Lozupone et al. (2013) found that samples from the human microbiome project (HMP) where storage and DNA extraction were performed according to the same protocols, but which

were amplified using different primers covering V1–V3 or V3–V5, hypervariable 16S rDNA regions clustered separately in principal coordinate analyses. For this reason, direct comparison of microbiome studies carried out with different primer sets are hardly possible. As shown in our study, intra-laboratory variance was relatively small and reliable comparisons could at least be made for datasets amplified with primer sets targeting the same hypervariable 16S rDNA regions. Additionally, discriminative effects of read length (Kuleshov et al., 2015; Hiergeist et al., 2015), as well as amplified discriminative variable and conserved sites of the 16S rRNA gene (Vinje et al., 2014), have to be considered. Comprehensive validation of universal 16S rDNA amplification primers and effective comparisons of covered relevant bacterial taxa are necessary (Klindworth et al., 2013) to ensure a consistent representation of microbiota profiles in microbiome analyses.

The addition of verifiable quantitative amounts of exogenic bacterial marker species allowed us to further investigate the recovery of defined ratios between bacterial species. Our analyses revealed major differences in the correct retrieval of expected proportions, at least for known species. Spike-in controls are widely used in other research fields like RNA-Seq (Li et al., 2014), and should be considered as controls for validation of protocols and as process controls in microbiome research. Since marker gene-based studies are highly dependent on high quality reads and the accuracy of next-generation-sequencing, and downstream data analyses algorithms will continue to improve, close attention should be particularly paid to the further development and standardization of wet-lab protocols, as well as their specific adjustment to microbiome analyses. This would allow for uniform DNA extraction and PCR amplification in order to minimize the methodological variance that outweighs biological differences and, moreover, allow for appropriate interpretation of cross-study comparisons in the human microbiome field. In addition, researches should be encouraged to accurately deposit methodological details (their standard operation procedures (SOPs)) to public sequence databases such as the European Nucleotide Archive (ENA), to enable traceability of methodological bias and to further improve the inter-center comparability of microbiome analysis data.

5. Conclusion

High inter-laboratory deviations were observed from our first external quality assessment of 16S rDNA next-generation-sequencing-based microbiota profiling. Our findings primarily indicate that the standardization and development of methods to increase cross-study comparability is urgently needed. This is particularly important in light of the accelerating transition of sophisticated methods in microbiome analysis from research to clinical routine diagnostics, which places very complex qualitative demands on laboratories. With the International Human Microbiome Standards (IHMS) project, initial efforts have already been undertaken to encourage methodological standardization of microbiome analyses, and the development of uniform standard operating procedures (Sinha et al., 2015). Here, we reported on the first inter-laboratory quality assessment scheme for microbiome analyses in a comparative study.

Funding

This study was supported by INSTAND e.V. and the German Research Foundation (DFG GE 671/14-1).

Conflict of interest

All authors declare no competing financial interests.

Acknowledgements

We would like to thank all participating centers and their coworkers for their invaluable support of this study by performing microbiome analyses, providing sequencing data and completing important questionnaires. Furthermore, the authors acknowledge the assistance of Claudia Deinzer in sample preparation, performing DNA extraction and qPCR quantification, Nadja Reul for 16S rDNA amplification and Christine Irtenkauf and Holger Melzl for performing next-generation-sequencing as well as Sabrina Woltemate for excellent technical assistance. We would additionally like to thank INSTAND e.V., Society for Promoting Quality Assurance in Medical Laboratories e.V. in Düsseldorf, Germany for their help in organizing quality assessment schemes for microbiome analyses and the members of the DFG Priority Programme “Intestinal Microbiota” (SPP1656) for fruitful discussions.

Appendix A.

Priority Program 1656 Intestinal Microbiota Consortium/ quality assessment participants: Debora Garzetti [b,c] Bärbel Stecher [b,c] Eric JC Gálvez [d] Till Strowig [d] Ines Yang [e] Sebastian Suerbaum [e] Nicole Fischer [f] Adam Grundhoff [f] Karl Kashofer [g] Andrea Thüringer [g] John F. Baines [h,i] Marie Vallier [h,i] Thomas Clavel [j] Dirk Haller [j,k]

[b]Ludwig-Maximilians University (LMU) Munich, Max von Pettenkofer Institute, Munich, Germany

[c]Center for Gut Microbiome Reseaerch (CEGIIMIR), German Center of Infection Research (DZIF), Partner site Munich, Germany

[d]Helmholtz Centre for Infection Research, Braunschweig, Germany

[e]Institute for Medical Microbiology and Hospital Epidemiology. Hannover Medical School, Hannover, Germany

[f]Heinrich-Pette Institute, Leibniz Institute for Experimental Virology, Department Virus Genomic, Hamburg, Germany

[g]Institute of Pathology, Medical University Graz, Graz, Austria

[h]Max Planck Institute for Evolutionary Biology, Plön, Germany

[i]Institute for Experimental Medicine, Christian-Albrechts-University of Kiel, Kiel, Germany

[j]ZIEL Institute for Food and Health, Technische Universität München, Germany

[k]Chair of Nutrition and Immunology, Technische Universität München, Germany

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijmm.2016.03.005.

References

Aronesty, E., Ea-utils: command-line tools for processing biological sequencing data. 2011 https://code.google.com/archive/p/ea-utils/.

Bahl, M.I., Bergström, A., Licht, T.R., 2012. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. FEMS Microbiol. Lett. 329, 193–197, http://dx.doi.org/10.1111/j.1574-6968.2012.02523.x.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336, http://dx.doi.org/10.1038/nmeth.f.303.

Chen, W., Zhang, C.K., Cheng, Y., Zhang, S., Zhao, H., 2013. A comparison of methods for clustering 16S rRNA sequences into OTUs. PLoS One 8, e70837, http://dx.doi.org/10.1371/journal.pone.0070837.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal database project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633–D642, http://dx.doi.org/10.1093/nar/gkt1244.

D'Argenio, V., Casaburi, G., Precone, V., Salvatore, F., D'Argenio, V., Casaburi, G., Precone, V., Salvatore, F., 2014. Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines. Biomed. Res. Int. 2014, 325340, http://dx.doi.org/10.1155/2014/325340.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072, http://dx.doi.org/10.1128/AEM.03006-05.

Dominianni, C., Wu, J., Hayes, R.B., Ahn, J., 2014. Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiol. 14, 1–6, http://dx.doi.org/10.1186/1471-2180-14-103.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461, http://dx.doi.org/10.1093/bioinformatics/btq461.

Gaspar, J.M., Thomas, W.K., 2015. FlowClus: efficiently filtering and denoising pyrosequenced amplicons. BMC Bioinf. 16, 105, http://dx.doi.org/10.1186/s12859-015-0532-1.

Hiergeist, A., Gläsner, J., Reischl, U., Gessner, A., 2015. Analyses of intestinal microbiota: culture versus sequencing: fig. 1. ILAR J. 56, 228–240, http://dx.doi.org/10.1093/ilar/ilv017.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., Glockner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41, http://dx.doi.org/10.1093/nar/gks808, e1–e1.

Kuleshov, V., Jiang, C., Zhou, W., Jahanbani, F., Batzoglou, S., Snyder, M., 2015. Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome. Nat. Biotechnol. 34, 64–69, http://dx.doi.org/10.1038/nbt.3416.

Lagier, J.-C., Million, M., Hugon, P., Armougom, F., Raoult, D., 2012. Human gut microbiota: repertoire and variations. Front. Cell. Infect. Microbiol. 2, 136, http://dx.doi.org/10.3389/fcimb.2012.00136.

Lee, Z.M.P., Bussema, C., Schmidt, T.M., 2009. rrn DB: Documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res. 37, http://dx.doi.org/10.1093/nar/gkn689.

Li, S., Labaj, P.P., Zumbo, P., Sykacek, P., Shi, W., Shi, L., Phan, J., Wu, P.-Y., Wang, M., Wang, C., Thierry-Mieg, D., Thierry-Mieg, J., Kreil, D.P., Mason, C.E., 2014. Detecting and correcting systematic variation in large-scale RNA sequencing data. Nat. Biotechnol. 32, 888–895, http://dx.doi.org/10.1038/nbt.3000.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol., 251364, http://dx.doi.org/10.1155/2012/251364.

Lozupone, C.A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vazquez-Baeza, Y., Jansson, J.K., Gordon, J.I., Knight, R., 2013. Meta-analyses of studies of the human microbiota. Genome Res. 23, 1704–1714.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One 7, http://dx.doi.org/10.1371/journal.pone.0030087.

Marsland, B.J., Gollwitzer, E.S., 2014. Host–microorganism interactions in lung diseases. Nat. Rev. Immunol. 14, 827–835, http://dx.doi.org/10.1038/nri3769.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., OHara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2015. Vegan: Community Ecology Package. R Packag. version 2.3-1, http://dx.doi.org/10.4135/9781412971874.n145.

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., Stahl, D.A., 1986. Microbial ecology and evolution: a ribosomal RNA approach. Annu. Rev. Microbiol. 40, 337–365, http://dx.doi.org/10.1146/annurev.mi.40.100186.002005.

Pinto, A.J., Raskin, L., 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. PLoS One 7, http://dx.doi.org/10.1371/journal.pone.0043093.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glockner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596, http://dx.doi.org/10.1093/nar/gks1219.

R Developement Core Team, 2008. Computational Many-Particle Physics, R Foundation for Statistical Computing, Lecture Notes in Physics. Springer, Berlin, Heidelberg, http://dx.doi.org/10.1007/978-3-540-74686-7.

Salipante, S.J., Kawashima, T., Rosenthal, C., Hoogestraat, D.R., Cummings, L.A., Sengupta, D.J., Harkins, T.T., Cookson, B.T., Hoffman, N.G., 2014. Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling. Appl. Environ. Microbiol. 80, 7583–7591, http://dx.doi.org/10.1128/AEM.;1;02206-14.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541, http://dx.doi.org/10.1128/AEM.01541-09.

Schloss, P.D., 2010. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput. Biol. 6, 19, http://dx.doi.org/10.1371/journal.pcbi.1000844.

Schmidt, T.S.B., Matias Rodrigues, J.F., von Mering, C., 2014. Ecological consistency of SSU rRNA-Based operational taxonomic units at a global scale. PLoS Comput. Biol. 10, http://dx.doi.org/10.1371/journal.pcbi.1003594.

Sekirov, I., Russell, S.L., Antunes, L.C.M., Finlay, B.B., 2010. Gut microbiota in health and disease. Physiol. Rev. 90, 859–904, http://dx.doi.org/10.1152/physrev.00045.2009.

Shreiner, A.B., Kao, J.Y., Young, V.B., 2015. The gut microbiome in health and in disease. Curr. Opin. Gastroenterol. 31, 69–75, http://dx.doi.org/10.1097/MOG.0000000000000139.

Sinha, R., Abnet, C.C., White, O., Knight, R., Huttenhower, C., 2015. The microbiome quality control project: baseline study design and future directions. Genome Biol. 16, 276, http://dx.doi.org/10.1186/s13059-015-0841-8.

Suzuki, M.T., Giovannoni, S.J., 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630.

Tremblay, J., Singh, K., Fern, A., Kirton, E.S., He, S., Woyke, T., Lee, J., Chen, F., Dangl, J.L., Tringe, S.G., 2015. Primer and platform effects on 16S rRNA tag sequencing. Front. Microbiol. 6, http://dx.doi.org/10.3389/fmicb.2015.00771.

Vinje, H., Almøy, T., Liland, K.H., Snipen, L., 2014. A systematic search for discriminating sites in the 16S ribosomal RNA gene. Microb. Inform. Exp. 4, 2, http://dx.doi.org/10.1186/2042-5783-4-2.

van Dijk, E.L., Jaszczyszyn, Y., Thermes, C., 2014. Library preparation methods for next-generation sequencing: tone down the bias. Exp. Cell Res. 322, 12–20, http://dx.doi.org/10.1016/j.yexcr.2014.01.008.

Yuan, S., Cohen, D.B., Ravel, J., Abdo, Z., Forney, L.J., 2012. Evaluation of methods for the extraction and purification of DNA from the human microbiome. PLoS One 7, http://dx.doi.org/10.1371/journal.pone.0033865.