

RESEARCH ARTICLE

# Usability of rectal swabs for microbiome sampling in a cohort study of hematological and oncological patients

Lena M. Biehl<sup>1,2,3</sup>, Debora Garzetti<sup>4,5</sup>, Fedja Farowski<sup>1,3</sup>, Diana Ring<sup>4</sup>, Martin B. Koepfel<sup>4,5</sup>, Holger Rohde<sup>2,6</sup>, Philippe Schafhausen<sup>7</sup>, Bärbel Stecher<sup>4,5</sup>, Maria J. G. T. Vehreschild<sup>1,3,8\*</sup>

**1** University of Cologne, Department I of Internal Medicine, University Hospital of Cologne, Cologne, Germany, **2** Institute for Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, **3** German Centre for Infection Research (DZIF), Partner site Bonn-Cologne, Cologne, Germany, **4** Max von Pettenkofer Institute, Faculty of Medicine, LMU Munich, Munich, Germany, **5** German Centre for Infection Research (DZIF), Partner site Munich, Munich, Germany, **6** German Centre for Infection Research (DZIF), Partner site Hamburg-Borstel, Hamburg, Germany, **7** Department of Oncology and Hematology, Hubertus Wald Tumorzentrum/University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, **8** Infectious Diseases Unit, Medical Clinic II, University Hospital Frankfurt, Frankfurt am Main, Germany

\* These authors contributed equally to this work.

\* [maria.vehreschild@uk-koeln.de](mailto:maria.vehreschild@uk-koeln.de)



## OPEN ACCESS

**Citation:** Biehl LM, Garzetti D, Farowski F, Ring D, Koepfel MB, Rohde H, et al. (2019) Usability of rectal swabs for microbiome sampling in a cohort study of hematological and oncological patients. PLoS ONE 14(4): e0215428. <https://doi.org/10.1371/journal.pone.0215428>

**Editor:** Alexander V. Alekseyenko, Medical University of South Carolina, UNITED STATES

**Received:** April 11, 2018

**Accepted:** April 3, 2019

**Published:** April 15, 2019

**Copyright:** © 2019 Biehl et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The datasets generated and analyzed during the current study are available in the NCBI Sequence Read Archive under the BioProject accession number PRJNA376506 (validation and patient samples). Sequences from the negative controls are available as fasta [S2 File](#).

**Funding:** This work was supported by the German Center for Infection Research (DZIF) under grant numbers TI 07.001 to LMB and TTU 08.904 and the DZIF Center for Gastrointestinal Microbiome

## Abstract

### Objectives

Large-scale clinical studies investigating associations between intestinal microbiota signatures and human diseases usually rely on stool samples. However, the timing of repeated stool sample collection cannot be predefined in longitudinal settings. Rectal swabs, being straightforward to obtain, have the potential to overcome this drawback. Therefore, we assessed the usability of rectal swabs for microbiome sampling in a cohort of hematological and oncological patients.

### Study design

We used a pipeline for intestinal microbiota analysis from deep rectal swabs which was established and validated with test samples and negative controls. Consecutively, a cohort of patients from hematology and oncology wards was established and weekly deep rectal swabs taken during their admissions and re-admissions.

### Results

Validation of our newly developed pipeline for intestinal microbiota analysis from rectal swabs revealed consistent and reproducible results. Over a period of nine months, 418 rectal swabs were collected longitudinally from 41 patients. Adherence to the intended sampling protocol was 97%. After DNA extraction, sequencing, read pre-processing and filtering of chimeric sequences, 405 of 418 samples (96.9%) were eligible for further analyses. Follow-up samples and those taken under current antibiotic exposure showed a significant

Research (CEGIMIR) to BS. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** LMB has received lecture honoraria from Astellas Pharma and MSD and travel grants from 3M and Gilead. MJGTV has served at the speakers' bureau of Pfizer, Merck, Gilead Sciences and Astellas Pharma, received research funding from 3M, DaVolterra, Astellas Pharma, Gilead Sciences and Merck and is a consultant to Berlin Chemie, Astellas Pharma and Merck. All other authors: none to declare. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

decrease in alpha diversity as compared to baseline samples. Microbial domination occurred most frequently by Enterococcaceae (99 samples, 24.4%) on family level and *Enterococcus* (90 samples, 22.2%) on genus level. Furthermore, we noticed a high abundance of potential skin commensals in 99 samples (24.4%).

## Summary

Deep rectal swabs were shown to be reliable for microbiome sampling and analysis, with practical advantages related to high sampling adherence, easy timing, transport and storage. The relatively high abundance of putative skin commensals in this patient cohort may be of potential interest and should be further investigated. Generally, previous findings on alpha diversity dynamics obtained from stool samples were confirmed.

## Introduction

The intestinal tract harbors a complex microbial community which plays a central role in human health. Disruption of the gut microbiota (or dysbiosis) is associated with pathological intestinal conditions such as obesity and malnutrition, metabolic diseases such as diabetes, and chronic inflammatory diseases such as inflammatory bowel diseases [1–4]. The recent increase in clinical microbiome studies has been facilitated by advances in high throughput sequencing technologies, which offer rapid and comprehensive culture-independent techniques, mostly relying on DNA sequencing of the 16S ribosomal RNA gene [5].

When conducting clinical studies addressing microbiome dynamics in humans, longitudinal sampling at precisely determined intervals is often needed. Except for settings where mucosal biopsies are available, such as during colonoscopies, stool samples are the “gold standard” starting material for this purpose [6]. However, their acquisition can be challenging. Most importantly, timing can be jeopardized if a study participant does not produce a stool sample within the desired time window. Furthermore, immediate transport and adequate storage is difficult during evenings and on weekend days. These organizational drawbacks may be overcome if rectal swabs are used instead. In fact, they can be obtained at specific time points, thus preventing prolonged periods between sampling and storage. As shown by previous studies, microbiota profiles of rectal swab samples are, to some extent, similar to the ones of fecal samples, even though contamination with skin bacteria can occur and may be caused by incorrect sample collection [6–9]. To our knowledge, so far only few studies have used rectal swabs for microbiota sampling in a clinical setting. Published work mainly focused on healthy subjects and on patients with inflammatory bowel disease [6, 10]. Therefore, little is known about the applicability of rectal swabs during chemotherapy and antibiotic treatment, interventions which are both known to affect the intestinal microbiome [11, 12]. In this study we report on the usability of rectal swabs for microbiota sampling in a longitudinal cohort of high-risk hematological and oncological patients undergoing chemotherapy and antibiotic treatment.

## Materials and methods

### Validation and control samples

In order to validate our pipeline for microbiome analysis and its application on rectal swabs, we tested its performance and reproducibility on a small number of subjects (S1 Table). Stool and deep rectal swab samples (FecalSwab<sup>TM</sup>, COPAN Italia, Brescia, Italy) were obtained from

one patient and seven healthy volunteers. For each individual, specimens were collected simultaneously. To account for different sampling times, storage conditions, primer pairs and DNA extraction batches, we included technical replicates for volunteer H1 (S1 Table). Moreover, as the transport medium may impact microbiome profiling at several steps and influence bacterial survival/growth, DNA extraction and PCR efficiency, we obtained a second deep rectal swab from four of the volunteers immediately after the first swab. While a Cary-Blair medium (composition:  $\text{Na}_2\text{HPO}_4$  1.1 g/l,  $\text{C}_2\text{H}_3\text{NaO}_2\text{S}$  1.5 g/l, NaCl 5.0 g/l,  $\text{CaCl}_2$  0.09 g/l, agar 5.6 g/l) is used as a standard solution in our swabs, the second swabs were transported and frozen in saline solution (0.9% NaCl). In order to exclude errors which may arise during any step of our microbiome analysis pipeline, we included a mock community composed of twelve bacterial strains [13]. In detail, 200  $\mu\text{l}$  of the mock community were inoculated either in 1 ml Cary Blair medium or 1 ml saline solution, and processed in the same way as the other validation samples.

PCR, sequencing and Bioinformatic analyses were performed as described in the paragraph “Bioinformatic analyses and taxonomic profiling”, except that a rarefaction level of 10,000 sequences per sample was applied before alpha and beta diversity calculation. Correlation analysis of taxonomy profiles was done calculating  $R^2$  correlation coefficients for all pairs of samples, while statistics significance was calculated with a 2-way ANOVA test with a Bonferroni post hoc test.

To monitor possible contaminations, the pipeline was also applied to experimental controls from sterile swabs, DNA extraction buffers and reagents used for amplicon library preparation, which were processed together with the patient samples. For these negative control samples, sequencing was performed with specific primers (see below and S2 Table) and taxonomic classification was performed aligning the pre-processed high-quality and chimera-free sequences to the SILVA databases with the online tool SILVA Incremental Aligner, version 1.2.11 [14].

## Patients and study settings

Patients were eligible for inclusion into the longitudinal cohort if they were admitted for intensive cytotoxic treatment with subsequent neutropenia requiring an expected hospitalization of at least 3 weeks. The cohort study was approved by the local institutional review board and ethics committee of the Medical Chamber in Hamburg (Ethik-Kommission der Ärztekammer Hamburg), Germany (Study ID: PV4722). Samples from healthy volunteers were collected within a volunteer sample collection protocol approved by the Ethics Commission of the Faculty of Medicine of Cologne University (Ethikkommission der Medizinischen Fakultät der Universität zu Köln), Germany (Study ID: 16–234). Written informed consent was obtained from all patients and healthy volunteers prior to inclusion into the study.

All patients were treated according to local standards of care at the Medical Clinic and Polyclinic of the University Medical Center Hamburg-Eppendorf, Germany, which includes antimicrobial prophylaxis with low-dose trimethoprim/sulfamethoxazole twice weekly and ciprofloxacin or sulfamycin daily during neutropenia.

## Clinical data and definitions

Clinical data collected from all patients included underlying disease, chemotherapy, antibiotic prophylaxis and treatment, duration of neutropenia (defined as a neutrophil count below 500/ $\mu\text{l}$ ), abdominal symptoms, temperature, as well as results of microbiological tests. Samples under current antibiotic exposure were defined as samples taken during ongoing antibiotic

administration of at least 3 days prior sampling or samples taken within 2 days after termination of antibiotic treatment.

## Sample collection and procession

Deep rectal swabs (FecalSwab, COPAN Italia, Brescia, Italy, lot number N02T01) were obtained at baseline and weekly during inpatient treatment. Sampling was resumed in case of re-admission until no further oncological treatment was planned or until the maximum study duration of 9 months. Deep rectal swabs were performed by trained medical staff (nurses, medical students), with swabs being inserted at least 2 cm intrarectally and rotated while avoiding contact with the perianal skin of the patient.

After sample collection, storage and transport of the fecal swabs were allowed at room temperature for a maximum of 4 hours or at 4–8°C for a maximum of 5 hours. Swab Cary-Blair media (2,000 µl) were vortexed, and aliquots of 1,200 µl were stored at -80°C until microbiome analysis, while the remaining 800 µl were kept at -80°C for any further necessary experiment.

## Microbiota analysis

Profiling of the gut microbiota was performed as recently published [15] with some modifications, as described below.

**DNA isolation.** Genomic DNA (gDNA) was isolated from rectal swabs according to the procedures previously described [16, 17]. Briefly, each frozen sample was thawed, pelleted and suspended in 500 µl of extraction buffer (200 mM Tris pH 8.0, 200 mM NaCl, 20 mM EDTA), 210 µl of 20% SDS, 500 µl of a mix of phenol:chloroform:isoamyl alcohol (24:24:1), and 500 µl of zirconia/silica beads (0.1-mm diameter). Bacterial cells were lysed by mechanical disruption with a bead beater (50 s<sup>-1</sup>) for 4 min, followed by DNA extraction in phenol:chloroform:isoamyl alcohol and precipitation with ethanol. The isolated gDNA was resuspended in 10 mM Tris buffer and purified with NucleoSpin gDNA clean-up columns (Macherey-Nagel, Düren, Germany) in a final elution volume of 50 µl. gDNA was finally quantified by a fluorometric method (Quant-iT PicoGreen dsDNA Reagent, Thermo Fisher Scientific, Darmstadt, Germany), and stored at 4°C before further processing.

**16S rRNA amplicon sequencing.** For microbiota profiling, a dual-index strategy multiplexing 8 forward and 12 reverse primers for 16S rRNA gene sequencing on the MiSeq Illumina platform was developed, as previously described [18]. PCR amplification primers targeting the V3-V4 variable regions of the 16S rRNA gene were designed by Eurofins Genomics (Ebersberg, Germany) using a proprietary layout based on established methods (S2 Table) [19, 20]. To monitor possible errors due to contaminations or read-to-sample misassignment, primers for 8 control samples with specific barcodes were also included (S2 Table). PCR reactions contained 1X Q5 Hot Start High-Fidelity Master Mix (New England BioLabs, GmbH, Frankfurt Am Main, Germany), 0.5 µM each forward and reverse primers, and 15–20 ng of genomic DNA template. Reactions consisted of 30 s at 94°C; 30 cycles of 10 sec at 95°C, 20 sec at 55°C and 10 sec at 72°C; and a final extension at 72°C for 2 min. Each sample was amplified in duplicate and purified using the Agencourt AMPure XP PCR Purification system (Beckman Coulter, Krefeld, Germany). Purified amplicons were quantified using the fluorometric PicoGreen dsDNA reagent. A pooled sample was created by combining equimolar amounts of the 96 individual amplicons and sent to Eurofins Genomics for library quality control and sequencing. Libraries were mixed with 1% PhiX genomic control library and sequenced on the Illumina MiSeq v3 as 300-bp paired-end runs.

**Bioinformatic analyses and taxonomic profiling.** Prior to taxonomic analysis, reads were pre-processed to retain only high-quality sequences, using a combination of scripts from

the BBTools package [<http://jgi.doe.gov/data-and-tools/bbtools/>] and Trimmomatic [21] (see S1 File for a full list of used commands). Reads had to possess the full barcode free of errors, be longer than 200 bp after removal of the variability region and quality trimming, and have a minimum of 30 bp overlap with no mismatches during merging of paired-end reads. Joined reads, which had to be longer than 400 bp, were then quality filtered during demultiplexing with QIIME v. 1.9 [22], allowing a minimum Phred quality score of 20 and no ambiguous bases. Failed sequencing was considered for samples having a number of reads lower than the negative sample(s) included in the run. Chimeric sequences were then removed with USEARCH v. 6.1 [23]. Samples containing less than 2,000 sequences were heuristically considered not adequate for accurate profiling of the bacterial community composition and omitted for subsequent analyses. OTU (Operational Taxonomic Unit) picking and taxonomic classification were performed with QIIME. Open-reference OTU clustering and taxonomy assignment of sequences were performed with UCLUST [23] against the Silva database Release 119 [24] at the 97% similarity level. OTUs with a number of sequences < 0.01% of the total number of sequences were discarded as a second level of quality-filtering. After conducting a rarefaction test to determine the sufficient number of sequences for an accurate estimation of species diversity, the OTU table was finalized using a rarefaction level of 2,000 sequences per sample.

**Statistical analysis.** All further analyses were carried out using R for Statistical Computing version 3.2.5 (R Foundation for Statistical Computing, Vienna, Austria) [25]. The alpha and beta diversity metrics were calculated using functions provided in the phyloseq R package [26]. Continuous data was presented as mean ( $\pm$  standard deviation) and/or median (range) and tested with appropriate statistical analyses (two-tailed t-tests). Absolute and relative frequencies were given for categorical data (i.e. dominations) and tested using the chi-square test. Domination was defined as a single bacterial taxon comprising at least 30% of sequences and being the most abundant taxon [27].

## Presence of skin commensals

In order to evaluate whether detected bacterial taxa were originating from the anal skin rather than the rectal mucosa as a possible contamination introduced by inadequate sampling we investigated the presence of skin commensals [28]. From all detected taxa, the genera *Corynebacterium*, *Staphylococcus* and *Streptococcus* were chosen as potential contaminants and a relative abundance of more than 10% was defined as increased. Furthermore, frequencies of domination by these genera and a shift towards the dominating genus in the previous and subsequent samples were assessed. A shift was defined as an increased abundance of the respective genus of more than 10% in the sample directly prior or after the sample showing the domination. Of note, both terms, domination and shift only refer to results from compositional data and not to the total amount of bacteria in a sample.

## Availability of data and material

The datasets generated and analyzed during the current study are available in the NCBI Sequence Read Archive under the BioProject accession number PRJNA376506 (validation and patient samples). Sequences from the negative controls are available as fasta S2 File.

## Results

### Pipeline validation

We used a set of 35 samples for validation of our microbiota analysis pipeline and for evaluating reproducibility of the results. In particular, we designed experiments to control for effects



of sample collection, transport media, gDNA extraction and 16S rRNA amplicon sequencing. The analysis included samples from a patient and seven healthy volunteers as well as a mock bacterial community with known composition (S1 Table). Overall, highly consistent and reproducible results were obtained, proving that our new pipeline allows microbiota analysis of complex human intestinal samples. Specifically, we obtained a high number of reads from both stools and rectal swabs (range 12,490–415,602, median 93,112). Although the gDNA concentration from stools was higher than that from rectal swabs, this difference was not maintained during PCR amplification nor reflected in the sequencing output (S1 Table).

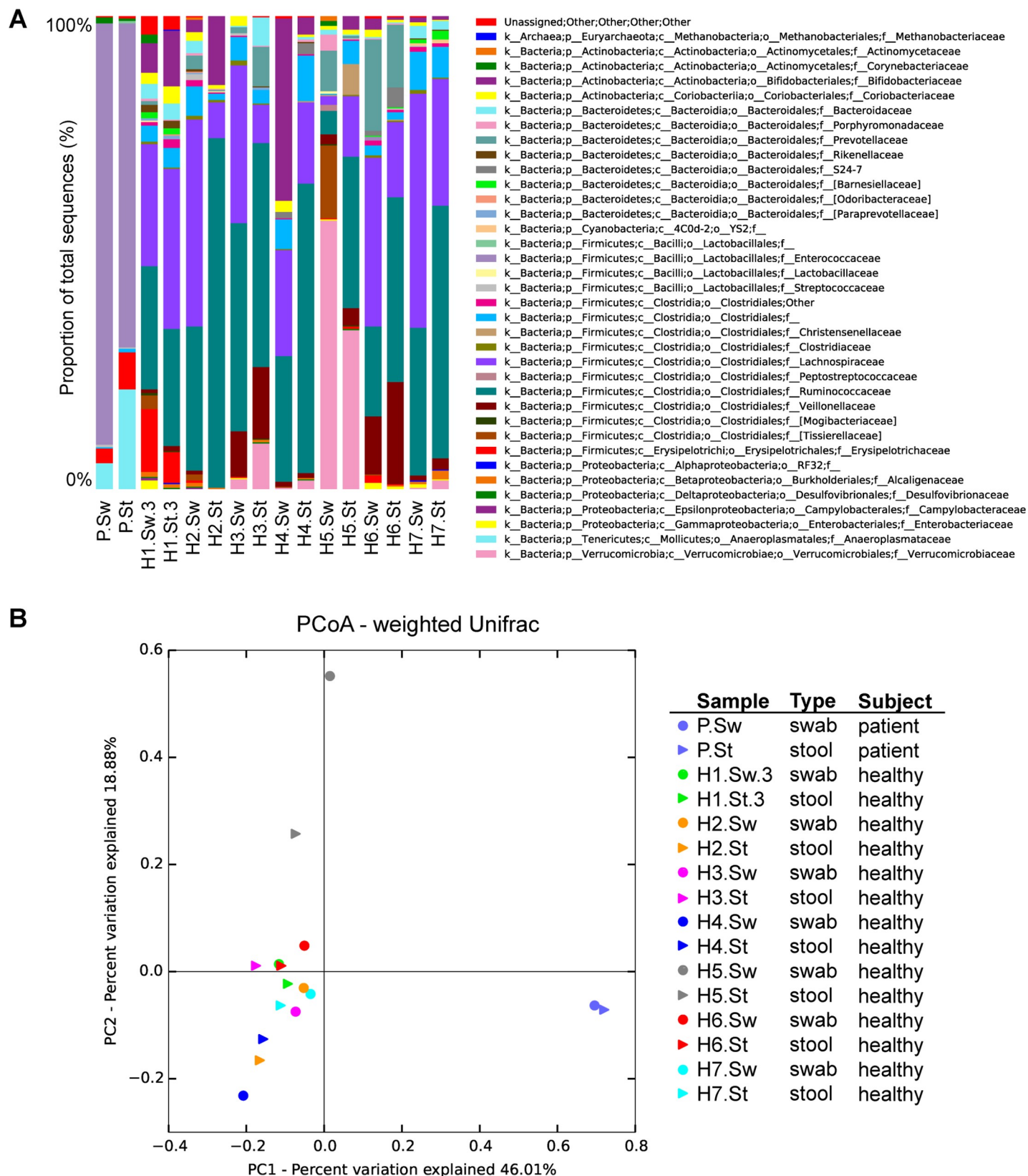
Rectal swabs and stools sampled at the same time point from the same subject yielded similar results in terms of microbiota composition. This applied both, for the patient ( $R^2 = 0.97$ , Shannon index: swab = 1.23, stool = 2.28) and for the healthy volunteers ( $R^2 = 0.86$ , Shannon index mean: swabs = 5.53, stools = 5.29) (Fig 1). Besides overall similarities of the taxonomic profiles, few differences in the individual identified taxa were detected (Fig 1A). In particular, the relative abundance of Enterococcaceae was higher in the patient swab sample ( $p < 0.01$ ), and Anaeroplasmataceae were more abundant in the patient stool ( $p < 0.05$ ), while stool samples from the healthy volunteers showed a significantly lower abundance of Lachnospiraceae and a higher proportion of Ruminococcaceae, compared to the swab samples ( $p < 0.001$ ). As expected, the patient showed a completely different microbiota profile compared to the healthy subjects (Fig 1B). Interestingly, our validation samples did not reveal higher abundances of skin commensals in swabs compared to stools, as seen in the patient samples (see below). In fact, Corynebacteriaceae were only detected in the patient samples and in two healthy subjects (H1 and H5), while the Streptococcaceae family was present in very low abundance in all rectal swabs from healthy individuals (mean of all samples: 0.2%).

The reproducibility and consistency of the results were confirmed by analyzing microbiota profiles from technical replicates of the swabs and stools taken from volunteer H1 (S1 Table). Indeed, major variations due to extraction of the gDNA in three batches ( $R^2 = 0.98$ ), or the use of alternative primer pairs ( $R^2 = 1$ ), or different storage conditions of stools ( $R^2 = 0.99$ ), or sampling of swabs at two time points ( $R^2 = 0.89$ ) could be excluded (S1 Fig).

Likewise, any impact of the Cary-Blair medium on microbiome profiles was ruled out by comparison with saline solution as a transport medium in swabs from four volunteers and a mock bacterial suspension (S1 Table and S2 Fig). Importantly, both mock samples revealed the expected bacterial composition, with 12 bacterial genera and 10 families detected [13].

## Control samples

To control for possible sources of contaminating bacterial DNA, we included experimental negative controls from DNA extraction buffers, PCR reagents, and sterile swabs. Overall, 14 control samples were sequenced, generating between 0 and 298 reads per sample. After pre-processing and chimera filtering, a total of 860 sequences remained (S3 Table and S2 File). All sequences could be classified as belonging to the Bacteria kingdom, and for 492 (57.2%) and 66 (7.7%) sequences classification was possible at the genus and species level, respectively. The predominant contaminants detected in our control samples are summarized in Table 1. Mostly, they have been already identified as contaminants of DNA extraction kits, PCR reagents, laboratory consumables, water and personnel [29–32]. Of the still unreported contaminants, some are common environmental organisms (e.g. *Sediminibacterium*, *Legionellales*, *Cyanobacteria*, *Hydrogenophilus*), while others are skin/mucosa-associated bacteria (e.g. *Staphylococcus*, *Dermabacteraceae*). Sequences assigned to intestinal bacteria (e.g. *Enterococcus*, *Roseburia*, *Ruminococcus*) may have been introduced during sample processing. However,



**Fig 1. Comparison of stool- versus swab-derived microbiota profiles.** (A) Taxonomic profiles at the family level. The legend has the same color direction as in the bar plots. (B) Beta-diversity calculated as weighted Unifrac and visualized by principal coordinate analysis (PCoA). P-patient; H-healthy volunteer; Sw-swab; St-stool.

<https://doi.org/10.1371/journal.pone.0215428.g001>

Table 1. List of the main contaminants detected in negative control samples.

Negative control	Phylum	Family	Genus	Proportion (%) <sup>b</sup>
Sterile swabs	Bacteroidetes	Flavobacteriaceae	<i>Flavobacterium</i> <sup>a</sup>	17.9
	Firmicutes	Enterococcaceae	<i>Enterococcus</i>	9.2
		Lachnospiraceae	Unclassified	5.8
	Proteobacteria	Rhizobiaceae	<i>Agrobacterium</i> <sup>a</sup>	5.8
		Pseudomonadaceae	<i>Pseudomonas</i> <sup>a</sup>	4.8
DNA extraction buffers	Actinobacteria	Corynebacteriaceae	<i>Corynebacterium</i> <sup>a</sup>	5.8
		Mycobacteriaceae	<i>Mycobacterium</i> <sup>a</sup>	4.6
	Armatimonadetes	[Fimbriimonadaceae]	Unclassified	5.8
	Proteobacteria	Comamonadaceae	<i>Schlegelella</i> <sup>a</sup>	5.2
			Unclassified <sup>a</sup>	5.8
		Enterobacteriaceae	Unclassified	6.4
		Legionellales, Unclassified	Unclassified	4.6
		Pseudomonadaceae	<i>Pseudomonas</i> <sup>a</sup>	4.6
PCR reagents	Armatimonadetes	[Fimbriimonadaceae]	Unclassified	5.2
	Bacteroidetes	Chitinophagaceae	<i>Sediminibacterium</i>	4.8
	Proteobacteria	Caulobacteraceae	Unclassified	5.6
		Brucellaceae	<i>Ochrobactrum</i> <sup>a</sup>	5.4
		Methylobacteriaceae	<i>Methylobacterium</i> <sup>a</sup>	9.2
		Comamonadaceae	Unclassified <sup>a</sup>	23.9
		Legionellales, Unclassified	Unclassified	6.3
		Pseudomonadaceae	<i>Pseudomonas</i> <sup>a</sup>	5.4

<sup>a</sup>Previously reported as contaminants [29–32]

<sup>b</sup>Proportion of sequences relative to each negative control type; only contaminants with proportions >4.5% are reported. For full list of contaminants, see S3 Table.

<https://doi.org/10.1371/journal.pone.0215428.t001>

due to the very low number of sequences, they are unlikely to affect the analysis of patient microbiome profiles.

## Patient characteristics and sample collection

From July 2014 to March 2015, 41 patients were included into the study. Median age was 60 years (range 19–76) with 22 patients (53.7%) being female. Acute leukemia was the most frequent underlying disease (29 patients; 70.7%) followed by lymphoma (7; 17.1%) and solid tumor (3; 7.3%). One patient had suspected rectal carcinoma, which was not confirmed, and one patient was admitted for hematopoietic stem cell transplantation for treatment of Multiple Sclerosis. The median duration of longitudinal sampling was 55 days (range 1–246). All except one patient received antibiotic treatment during their study participation, while all except three patients received chemotherapy.

According to the sample collection protocol, 431 weekly rectal swabs were planned to be collected during all admissions and re-admissions of study patients. We were able to collect 418 samples (range 1–30 per patient), resulting in a completeness rate of 97.0%. Reasons for missed weekly samples were short inpatient stay of less than 3 days and rejection of sampling by the patient, in one case due to general weakness and tiredness, and in one case due to an anal fissure and anticipated pain during sampling.

No adverse events occurred during rectal swab sampling, especially no bleeding, despite the low platelet counts of as low as  $10 \times 10^9/l$  and no consecutive infections, despite patients being neutropenic during some of the sampling. Two patients decided to quit their study



participation due to their wish to have fewer interventions and fewer disturbances during their oncological treatment.

### DNA extraction from rectal swabs

Isolation of genomic DNA, as assessed by fluorometric quantification, generated an average DNA concentration of 7.4 ng/μl (range: -0.228–110 ng/μl) (S4 Table). Importantly, no relation was found between the genomic DNA amount and the concentration of the corresponding PCR amplicon or the read output.

### Microbiome analysis from rectal swabs

PCR amplicons obtained from all the 418 rectal swabs were achieved in five different runs (96 samples each, including negative controls). After read pre-processing and filtering of chimeric sequences, on average  $85,106 \pm 60,725$  sequences per sample were obtained (S4 Table). Samples that failed sequencing ( $n = 4$ ), contained less than 2,000 reads after raw read pre-processing (6), or had less than 2,000 sequences represented in the OTU table used for calculation of diversity metrics (3) were omitted (total 13 out of 418 samples, 3.1%), resulting in a final collection of 405 samples with on average  $87,811 \pm 59,752$  sequences per sample (Table 2). Importantly, we observed a significantly higher amount of sequences ( $93,897 \pm 60,642$  vs  $67,950 \pm 52,297$ ;  $p < 0.001$ ) in samples from patients without current antibiotic exposure ( $n = 310/405$ ) (Table 3).

### Bacterial diversity analysis of rectal swab samples

The alpha diversity metrics of the analyzed swab samples are shown in Table 2. The Shannon index, the inverse Simpson index and the phylogenetic diversity of all samples were  $2.28 \pm 1.11$ ,  $8.02 \pm 8.08$  and  $12.21 \pm 6.70$ , respectively. The Shannon index, as well as the phylogenetic diversity, were significantly increased in the baseline samples compared to samples obtained during the course of chemotherapy ( $p = 0.013$  and  $p < 0.001$ , respectively). Additionally, in samples from patients with current antibiotic exposure ( $n = 95$ ) all three alpha diversity scores were significantly decreased (Table 3 and Fig 2).

### Bacterial domination in samples

Overall, bacterial domination, i.e. at least 30% of the sequences in a sample being assigned to a single taxon, was frequently observed both on family (338/405 samples; 83.5%) and genus level (231/405 samples; 57.0%). These dominations were caused by a variety of families and genera (Table 4), the most frequent one being Enterococcaceae (99 swabs, 24.4%; 31 patients, 75.6%) and *Enterococcus* (90 swabs, 22.2%; 28 patients, 68.3%), respectively. While the overall proportion of samples dominated by any genus or family did not differ between baseline and post-baseline samples, there was a significant shift towards domination by the genus *Enterococcus* in samples taken after day 0 ( $p = 0.01$ ; Table 4).

### Detection of skin microbiota in rectal swab samples

An increased abundance of more than 10% of one of the genera *Corynebacterium*, *Staphylococcus* and *Streptococcus* was present in 99 samples (24.4%) from 34 different patients (82.9%). These genera are abundant taxa of the human skin microbiota [28]. Domination by one of these genera was present in 44 samples (10.9%) from 23 patients (56.1%). Of note, in half of the samples with skin bacterial dominations there was a shift towards the dominating genus seen in the previous or subsequent sample (Table 5).

**Table 2. Sequences per sample, alpha diversity scores and frequency of dominations.**

	Total (n = 405)	Comparison		
		Baseline samples (day = 0; n = 39)	Post baseline samples (day > 0; n = 366)	p-value
<b>Sequences per sample mean ± SD</b>	<b>87811 ± 59752</b>	<b>57513 ± 37036</b>	<b>91039 ± 60831</b>	<b>&lt;0.001</b>
(range)	(2787-279263)	(16620-170624)	(2787-279263)	
<b>Shannon index mean ± SD</b>	<b>2.28 ± 1.11</b>	<b>2.68 ± 1.01</b>	<b>2.24 ± 1.11</b>	<b>0.013</b>
(range)	(0.01-4.46)	(0.10-4.23)	(0.01-4.46)	
<b>Inverse Simpson mean ± SD</b>	<b>8.02 ± 8.08</b>	<b>9.55 ± 7.30</b>	<b>7.86 ± 8.14</b>	<b>0.180</b>
(range)	(1.00-48.74)	(1.04-26.52)	(1.00-48.74)	
<b>PD whole tree mean ± SD</b>	<b>12.21 ± 6.70</b>	<b>16.72 ± 7.63</b>	<b>11.73 ± 6.42</b>	<b>&lt;0.001</b>
(range)	(1.07-34.25)	(1.07-32.99)	(1.11-34.25)	
<b>Domination</b>				
any domination by family % (n)	83.5 (338)	71.8 (28)	84.7 (310)	0.066
any domination by genus % (n)	57.0 (231)	48.7 (19)	57.9 (212)	0.350
domination by <i>Enterococcus</i> % (n)	22.2 (90)	5.1 (2)	24.0 (88)	0.012

SD: standard deviation; PD: phylogenetic diversity

<https://doi.org/10.1371/journal.pone.0215428.t002>

## Discussion

In this analysis, we present an intestinal microbiome analysis from a longitudinal cohort study utilizing deep rectal swabs for intestinal microbiota profiling. Previous studies have already compared rectal swabs with samples from different origins [6–9]. We validated the pipeline for DNA extraction and 16S rRNA gene sequencing on rectal swabs and stools obtained from seven healthy volunteers and one patient, confirming reproducibility of the results. Between swab- and stool-derived microbiota profiles we found some small differences, which may partly be due to dissimilarities in the sampled body sites (rectal mucosal wall versus stool), as previously reported [33]. Since it was not our aim to prove equality of these two sampling techniques, only this small number of comparative specimen from volunteers or patients were collected. Furthermore, the relatively low number of contaminant reads found in sterile swabs and negative controls is within the range of previously reported contamination of laboratory consumables [29], and is therefore expected not to confound the taxonomic composition of the patient samples.

Studies using stool samples are often confronted with low patient compliance and difficulties to obtain and process the required samples in a predefined time window. Hence, these studies usually face missing data as an issue. Over a period of 9 months, we were able to collect 418 of 431 (97%) intended samples as per collection protocol. Unfortunately, few studies have

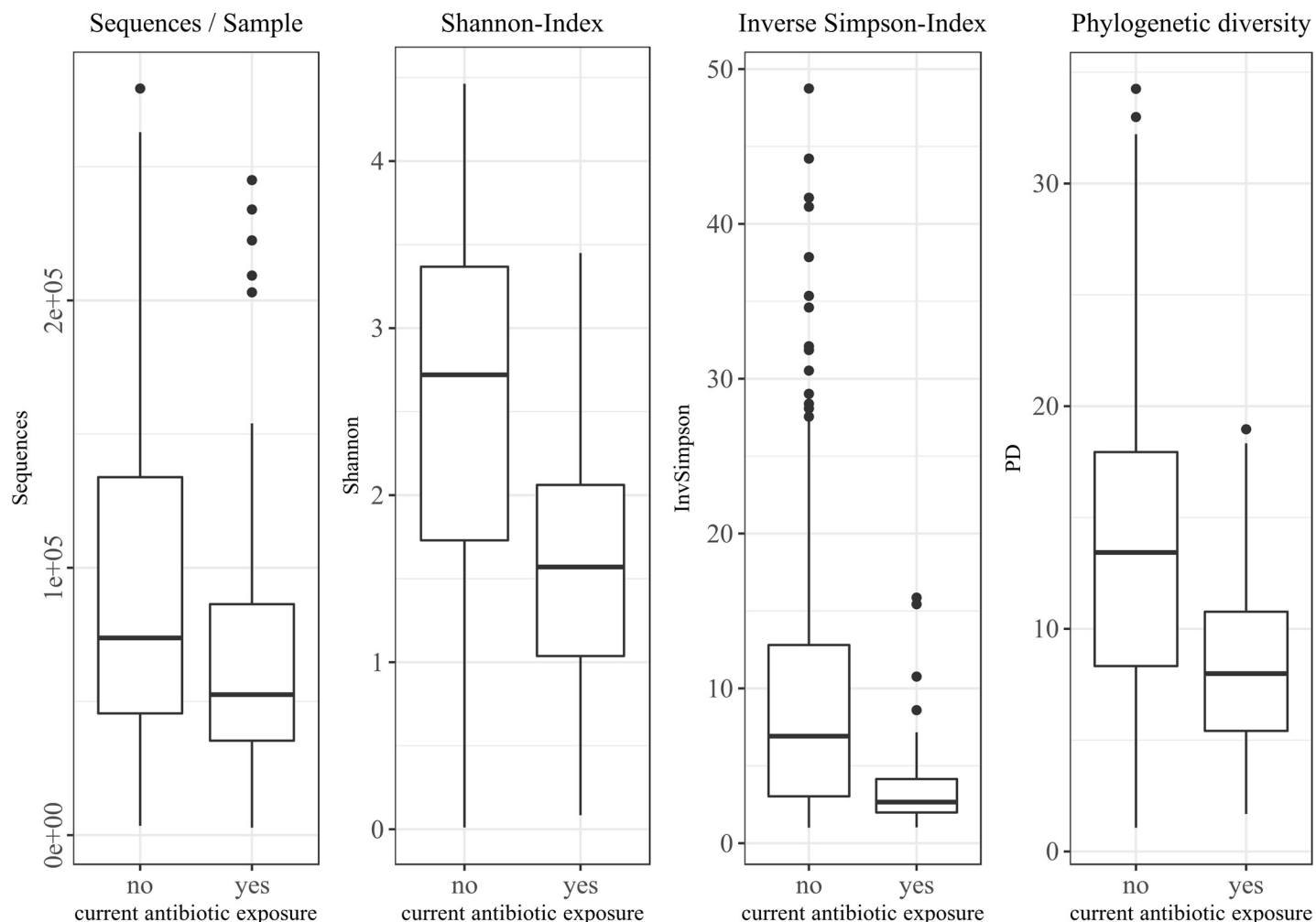
**Table 3. Effect of antibiotic exposure on sequences per sample and alpha diversity metrics.**

	No current antibiotic exposure <sup>a</sup> (n = 310)	Current antibiotic exposure <sup>a</sup> (n = 95)	p-value
Sequences per sample mean ± SD	93897 ± 60642	67950 ± 52297	<0.001
Shannon index mean ± SD	2.50 ± 1.11	1.59 ± 0.76	<0.001
Inverse Simpson mean ± SD	9.43 ± 8.66	3.43 ± 2.57	<0.001
PD whole tree mean ± SD	13.34 ± 6.95	8.53 ± 4.03	<0.001

SD: standard deviation; PD: phylogenetic diversity

<sup>a</sup>Samples under current antibiotic exposure were defined as i) samples taken during an ongoing antibiotic exposure that started at least 3 days prior to sampling or ii) samples taken within 2 days after termination of antibiotic treatment.

<https://doi.org/10.1371/journal.pone.0215428.t003>



**Fig 2. Effect of current antibiotic exposure on sequences per sample, Shannon-Index, inverse Simpson-Index and phylogenetic diversity (PD).** No current antibiotic exposure: 310 samples; current antibiotic exposure: 95 samples.

<https://doi.org/10.1371/journal.pone.0215428.g002>

so far reported the actual adherence to their sample collection protocol, but, as an example, the exclusion of up to 16% of patients due to insufficient samples possibly indicates adherence problems [27]. From our experience, the sample yield in the present study is very high, demonstrating a practical advantage of rectal swabs over stool samples. There were no adverse events related to obtaining deep rectal swabs in this patient cohort. However, the current study was not designed to assess safety and we are not aware of other studies systematically assessing adverse events of this sampling technique in a similar patient cohort. Nevertheless, rectal swabs are broadly implemented in hematological and oncological departments in Germany for several years now, further supporting our safety observations.

A major concern associated with the use of rectal swabs relates to the recovered amount of bacterial gDNA for successful taxonomic profiling. The DNA quantity extracted from swab samples showed very divergent values including negative ones. However, we found no relation between the amount of DNA and sequencing output. This suggests that differences in DNA yields were caused by technical issues producing inaccurate negative values, i.e. the phenol and ethanol used during DNA isolation might have interfered with the signal intensity of the DNA quantitation assay, producing inaccurate negative values. After omitting samples with less

Table 4. Most frequent dominating taxa on family and genus level.

	Dominating taxon	Percentage of samples (n); n = 405	Percentage of patients (n); n = 41
Family level	Enterococcaceae	24.4 (99)	75.6 (31)
	Lachnospiraceae	15.3 (62)	53.7 (22)
	Ruminococcaceae	8.1 (33)	46.3 (19)
	Prevotellaceae	7.4 (30)	36.6 (15)
	Enterobacteriaceae	6.7 (27)	43.9 (18)
	Other families <sup>a</sup>	21.5 (87)	75.6 (31)
	Total	83.5 (338)	100 (41)
Genus level	<i>Enterococcus</i>	22.2 (90)	68.3 (28)
	<i>Prevotella</i>	7.4 (30)	34.1 (14)
	<i>Corynebacterium</i>	4.2 (17)	29.3 (12)
	<i>Lactobacillus</i>	3.7 (15)	24.4 (10)
	<i>Bacteroides</i>	3.7 (15)	17.1 (7)
	Other genera <sup>b</sup>	15.8 (64)	63.4 (26)
	Total	57.0 (231)	95.1 (39)

<sup>a</sup>Other families included: Anaeroplasmataceae, Bacteroidaceae, Bifidobacteriaceae, Campylobacteraceae, Corynebacteriaceae, Dethiosulfovibrionaceae, Erysipelotrichaceae, Lactobacillaceae, Mycoplasmataceae, Porphyromonadaceae, Pseudomonadaceae, Staphylococcaceae, Veillonellaceae, Verrucomicrobiaceae

<sup>b</sup>Other genera included: *Akkermansia*, *Bifidobacterium*, *Blautia*, *Campylobacter*, *Dorea*, *Faecalibacterium*, *Fusobacterium*, *Mycoplasma*, *Porphyromonas*, *Pseudomonas*, *Pyramidobacter*, *Ruminococcus*, *Staphylococcus*, *Streptococcus*, *Ureaplasma*, *Veilonella*

<https://doi.org/10.1371/journal.pone.0215428.t004>

than 2,000 reads, 405 of 418 samples (96.9%) could be included for further analyses. In comparison to previous studies performed in the setting of hematological and oncological patients,

Table 5. Increased relative abundance of and domination by common skin microbiota.

	Percentage of samples (n); n = 405	Percentage of patients (n); n = 41
<i>Corynebacterium</i>		
Relative abundance in sample >10%	12.8 (52)	51.1 (21)
Domination by this genus	4.2 (17)	29.3 (12)
Shift towards domination seen in previous or subsequent sample <sup>a</sup>	3.7 (15)	26.8 (11)
<i>Staphylococcus</i>		
Relative abundance in sample >10%	6.4 (26)	34.1 (14)
Domination by this genus	3.5 (14)	14.6 (6)
Shift towards domination seen in previous or subsequent sample <sup>a</sup>	1.7 (7)	2.4 (1)
<i>Streptococcus</i>		
Relative abundance in sample >10%	10.4 (42)	58.5 (24)
Domination by this genus	3.2 (13)	29.3 (12)
Shift towards domination seen in previous or subsequent sample <sup>a</sup>	1.2 (5)	12.2 (5)
Any of the three genera		
Relative abundance in sample >10%	24.4 (99)	82.9 (34)
Domination by this genus	10.9 (44)	56.1 (23)
Shift towards domination seen in previous or subsequent sample <sup>a</sup>	6.7 (27)	39.0 (16)

<sup>a</sup>Abundance of respective genus >10% in sample from same patient directly before or after sample with domination.

<https://doi.org/10.1371/journal.pone.0215428.t005>

our rectal swab processing and sequencing yielded even a higher number of reads per sample than in stool samples [27, 34]. Importantly, the amount of DNA extracted from most of the rectal swabs (range 600 pg–5.5 µg), after depletion of host DNA, might allow metagenome sequencing (e.g. on an Illumina platform), which nowadays produces high quality data with as little as 10 pg of starting material [35, 36].

In terms of microbiota composition, our results can be compared with published studies, with the limitation that those were conducted using stool samples and analyzing different 16S rRNA variable regions. In a recent stool sample-based study comprising 34 patients receiving induction chemotherapy for the treatment of acute myelogenous leukemia [34], a wide range of alpha diversity among baseline samples was observed, with a mean Shannon index of 2.0 (95% confidence interval (CI), 1.7–2.4; range, 0.1–3.5). Similarly, in the present cohort, the mean Shannon index at baseline was 2.7 (95% CI, 2.4–3.0; range 0.01–4.5). The observed rates of domination on family (83.5% of samples) and genus level (57% of samples) are also comparable to their reported rate, with domination found in 79.2% of the samples [34]. In our cohort, Enterococcal domination was the most frequently observed one, as previously reported [27, 34]. In line with previous studies, we observed a statistically significant decrease in the microbial diversity over the course of chemotherapy [11, 27, 34] and after antibiotic exposure [12, 27, 34].

Our analysis, however, also revealed a possible difference in the microbiome composition from different intestinal locations captured by the use of a swab as compared to stool samples. We identified a relatively high number of patient samples in which typical skin commensals (*Corynebacterium*, *Staphylococcus*, *Streptococcus*) were present or even dominated the microbiota profile of the samples. This difference was not seen in the validation samples from a patient and various healthy volunteers. While we cannot rule out that these taxa represent actual skin microbiota of the patients resulting from flawed sampling technique, one could hypothesize that the high antibiotic burden and ongoing chemotherapy led to a shift towards these taxa in the distal intestinal microbiome. In half of the cases with observed domination by skin commensals, we found an increased abundance of the respective genus in the previous or subsequent sample, suggesting a true and possibly treatment-related change in the microbiota composition of the upper rectal mucosa. Interestingly, high relative abundances of skin bacteria, most frequently *Staphylococcus* spp. and *Streptococcus* spp., were also observed in studies using stool samples [27, 34]. *Corynebacterium* was reported to be more frequently found in rectal mucosal biopsies than stool samples in one study performing 16S rRNA gene sequencing on samples from healthy volunteers [33], suggesting that they may not only be part of the anal skin flora but also that of the rectal mucosa. Further studies are needed to address this issue including a larger number of comparative stool and rectal swab samples from patients.

Our study shows that the utilization of rectal swabs is adequate and feasible for the analysis of intestinal microbiota compositions in patients with exposure to chemotherapy and antibiotics. The characteristics of the rectal swab sample collection, including diversity metrics and domination frequencies, are comparable with those of similar clinical cohorts using stool samples. Our results confirm previous findings about the impact of chemotherapy and antibiotic exposure on microbial diversity. The increased abundance of skin commensals may possibly be related to chemotherapy and antibiotic exposure in this specific cohort, rather than contamination with skin bacteria during sampling. This observation needs to be investigated in future studies in more detail. In conclusion, rectal swabs have a number of practical advantages possibly leading to higher sample adherence and patient compliance. The total amount of DNA recovered from rectal swabs should also be sufficient for metagenome sequencing, which could provide more comprehensive insights into the functional and metabolic potential of the microbial community.



## Supporting information

**S1 Fig. Reproducibility and consistency of microbiota profiles.** (A): Taxonomic profiles and respective legend at the family level of different technical replicates from swabs and stools of volunteer H1. (B): Beta-diversity calculated as weighted Unifrac and visualized by principal coordinate analysis (PCoA).

(PDF)

**S2 Fig. Impact of Cary-Blair medium on microbiota profiles.** (A): Taxonomic profiles and respective legend at the family level of mock and volunteer samples transported in both Cary-Blair medium and saline solution (NaCl). (B): Beta-diversity calculated as weighted Unifrac and visualized by principal coordinate analysis (PCoA). The mock community is composed of 12 bacterial species belonging to 10 families.

(PDF)

**S1 File. List of commands used for bioinformatics processing of raw data and microbiome analyses.**

(TXT)

**S2 File. Fasta control sequences.**

(FNA)

**S1 Table. Characteristics of the samples used for pipeline validation.**

(XLSX)

**S2 Table. Dual-indexed PCR amplification primer sequences.** Red: Illumina adapter sequence. Black: variability region. Green: specific sample barcode. Blue: forward primer targeting the specific 16S rRNA regions V3-V4. Purple: reverse primer targeting the specific 16S rRNA region V3-V4.

(XLSX)

**S3 Table. Information and taxonomy classification of the control samples.**

(XLSX)

**S4 Table. List of sequenced swab samples and relative information.**

(XLSX)

## Acknowledgments

We thank all patients for participation in this study. Furthermore, we thank Walter Fiedler and Maximilian Christopeit for their support in patient enrolment, Cristina Belmar Campos for technical laboratory support and Nadine Mirwald for clinical documentation.

## Author Contributions

**Conceptualization:** Lena M. Biehl, Martin B. Koeppel, Holger Rohde, Bärbel Stecher, Maria J. G. T. Vehreschild.

**Data curation:** Lena M. Biehl, Debora Garzetti.

**Formal analysis:** Lena M. Biehl, Debora Garzetti, Fedja Farowski.

**Funding acquisition:** Lena M. Biehl, Bärbel Stecher.

**Investigation:** Lena M. Biehl, Debora Garzetti, Holger Rohde, Philippe Schafhausen.

**Methodology:** Lena M. Biehl, Debora Garzetti, Diana Ring, Holger Rohde.

**Project administration:** Lena M. Biehl.

**Resources:** Philippe Schafhausen.

**Software:** Debora Garzetti, Fedja Farowski.

**Supervision:** Lena M. Biehl, Martin B. Koepfel, Holger Rohde, Bärbel Stecher, Maria J. G. T. Vehreschild.

**Validation:** Lena M. Biehl, Debora Garzetti, Fedja Farowski, Diana Ring.

**Visualization:** Debora Garzetti, Fedja Farowski.

**Writing – original draft:** Lena M. Biehl, Debora Garzetti, Fedja Farowski, Martin B. Koepfel.

**Writing – review & editing:** Lena M. Biehl, Debora Garzetti, Fedja Farowski, Diana Ring, Martin B. Koepfel, Holger Rohde, Philippe Schafhausen, Bärbel Stecher, Maria J. G. T. Vehreschild.

## References

1. Zhao L. The gut microbiota and obesity: from correlation to causality. *Nature reviews Microbiology*. 2013; 11(9):639–47. Epub 2013/08/06. <https://doi.org/10.1038/nrmicro3089> PMID: 23912213.
2. Jonkers DM. Microbial perturbations and modulation in conditions associated with malnutrition and malabsorption. *Best practice & research Clinical gastroenterology*. 2016; 30(2):161–72. Epub 2016/04/19. <https://doi.org/10.1016/j.bpg.2016.02.006> PMID: 27086883.
3. Blandino G, Inturri R, Lazzara F, Di Rosa M, Malaguarnera L. Impact of gut microbiota on diabetes mellitus. *Diabetes & metabolism*. 2016; 42(5):303–15. Epub 2016/05/18. <https://doi.org/10.1016/j.diabet.2016.04.004> PMID: 27179626.
4. Chiriac MT, Mahapatro M, Neurath MF, Becker C. The Microbiome in Visceral Medicine: Inflammatory Bowel Disease, Obesity and Beyond. *Visceral medicine*. 2017; 33(2):153–62. Epub 2017/06/01. <https://doi.org/10.1159/000470892> PMID: 28560232; PubMed Central PMCID: PMC5447169.
5. Morgan XC, Huttenhower C. Chapter 12: Human microbiome analysis. *PLoS computational biology*. 2012; 8(12):e1002808. Epub 2013/01/10. <https://doi.org/10.1371/journal.pcbi.1002808> PMID: 23300406; PubMed Central PMCID: PMC3531975.
6. Budding AE, Grasman ME, Eck A, Bogaards JA, Vandenbroucke-Grauls CM, van Bodegraven AA, et al. Rectal swabs for analysis of the intestinal microbiota. *PloS one*. 2014; 9(7):e101344. Epub 2014/07/16. <https://doi.org/10.1371/journal.pone.0101344> PubMed Central PMCID: PMC4096398. PMID: 25020051
7. Araujo-Perez F, McCoy AN, Okechukwu C, Carroll IM, Smith KM, Jeremiah K, et al. Differences in microbial signatures between rectal mucosal biopsies and rectal swabs. *Gut microbes*. 2012; 3(6):530–5. Epub 2012/10/13. <https://doi.org/10.4161/gmic.22157> PMID: 23060016; PubMed Central PMCID: PMC3495790.
8. Shaw AG, Black N, Rushd A, Sim K, Randell P, Kroll JS, et al. Assessing the Colonic Microbiota in Children: Effects of Sample Site and Bowel Preparation. *Journal of pediatric gastroenterology and nutrition*. 2016. Epub 2016/04/14. <https://doi.org/10.1097/mpg.0000000000001233> PMID: 27070657.
9. Bassis CM, Moore NM, Lolans K, Seekatz AM, Weinstein RA, Young VB, et al. Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles. *BMC microbiology*. 2017; 17(1):78. Epub 2017/04/01. <https://doi.org/10.1186/s12866-017-0983-9> PMID: 28359329; PubMed Central PMCID: PMC5374586.
10. Patel AL, Mutlu EA, Sun Y, Koenig L, Green S, Jakubowicz A, et al. Longitudinal Survey of Microbiota in Hospitalized Preterm Very-Low-Birth-Weight Infants. *Journal of pediatric gastroenterology and nutrition*. 2016; 62(2):292–303. Epub 2015/08/01. <https://doi.org/10.1097/MPG.0000000000000913> PMID: 26230901; PubMed Central PMCID: PMC4724288.
11. Montassier E, Batard E, Massart S, Gastinne T, Carton T, Caillon J, et al. 16S rRNA gene pyrosequencing reveals shift in patient faecal microbiota during high-dose chemotherapy as conditioning regimen for bone marrow transplantation. *Microbial ecology*. 2014; 67(3):690–9. Epub 2014/01/10. <https://doi.org/10.1007/s00248-013-0355-4> PMID: 24402367.
12. Holler E, Butzhammer P, Schmid K, Hundsruker C, Koestler J, Peter K, et al. Metagenomic analysis of the stool microbiome in patients receiving allogeneic stem cell transplantation: loss of diversity is associated with use of systemic antibiotics and more pronounced in gastrointestinal graft-versus-host disease.

- Biology of blood and marrow transplantation: journal of the American Society for Blood and Marrow Transplantation. 2014; 20(5):640–5. Epub 2014/02/05. <https://doi.org/10.1016/j.bbmt.2014.01.030> PMID: 24492144.
13. Brugiroux S, Beutler M, Pfann C, Garzetti D, Ruscheweyh HJ, Ring D, et al. Genome-guided design of a defined mouse microbiota that confers colonization resistance against *Salmonella enterica* serovar Typhimurium. *Nature microbiology*. 2016; 2:16215. Epub 2016/11/22. <https://doi.org/10.1038/nmicrobiol.2016.215> PMID: 27869789.
14. Priesse E, Peplies J, Glockner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics (Oxford, England)*. 2012; 28(14):1823–9. Epub 2012/05/05. <https://doi.org/10.1093/bioinformatics/bts252> PMID: 22556368; PubMed Central PMCID: PMC3389763.
15. Barthels C, Ogrinc A, Steyer V, Meier S, Simon F, Wimmer M, et al. CD40-signalling abrogates induction of RORgammat+ Treg cells by intestinal CD103+ DCs and causes fatal colitis. *Nature communications*. 2017; 8:14715. Epub 2017/03/10. <https://doi.org/10.1038/ncomms14715> PMID: 28276457; PubMed Central PMCID: PMC5347138.
16. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009; 457(7228):480–4. Epub 2008/12/02. <https://doi.org/10.1038/nature07540> PMID: 19043404; PubMed Central PMCID: PMC267729.
17. Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, et al. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *The Journal of experimental medicine*. 2012; 209(8):1445–56. Epub 2012/07/25. <https://doi.org/10.1084/jem.20120504> PMID: 22826298; PubMed Central PMCID: PMC3409501.
18. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*. 2013; 79(17):5112–20. Epub 2013/06/25. <https://doi.org/10.1128/AEM.01043-13> PMID: 23793624; PubMed Central PMCID: PMC3753973.
19. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. *PloS one*. 2014; 9(8):e105592. Epub 2014/08/22. <https://doi.org/10.1371/journal.pone.0105592> PMID: 25144201; PubMed Central PMCID: PMC4140814.
20. Fadrosch DW, Ma B, Gajer P, Sengamalai N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. 2014; 2(1):6. Epub 2014/02/25. <https://doi.org/10.1186/2049-2618-2-6> PMID: 24558975; PubMed Central PMCID: PMC3940169.
21. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*. 2014; 30(15):2114–20. Epub 2014/04/04. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404; PubMed Central PMCID: PMC4103590.
22. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*. 2010; 7(5):335–6. Epub 2010/04/13. <https://doi.org/10.1038/nmeth.f.303> PMID: 20383131; PubMed Central PMCID: PMC3156573.
23. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxford, England)*. 2010; 26(19):2460–1. Epub 2010/08/17. <https://doi.org/10.1093/bioinformatics/btq461> PMID: 20709691.
24. Quast C, Priesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*. 2013; 41 (Database issue):D590–6. Epub 2012/11/30. <https://doi.org/10.1093/nar/gks1219> PMID: 23193283; PubMed Central PMCID: PMC3531112.
25. R Core Team. R: A language and environment for statistical computing. 3.2.5 ed. Vienna, Austria: R Foundation for Statistical Computing; 2016.
26. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013; 8(4):e61217. Epub 2013/05/01. <https://doi.org/10.1371/journal.pone.0061217> PONE-D-12-31789 [pii]. PMID: 23630581; PubMed Central PMCID: PMC3632530.
27. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, et al. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*. 2012; 55(7):905–14. Epub 2012/06/22. <https://doi.org/10.1093/cid/cis580> PMID: 22718773; PubMed Central PMCID: PMC3657523.

28. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(26):11971–5. Epub 2010/06/23. <https://doi.org/10.1073/pnas.1002601107> PMID: 20566857; PubMed Central PMCID: PMC2900693.
29. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC biology*. 2014; 12:87. Epub 2014/11/13. <https://doi.org/10.1186/s12915-014-0087-z> PMID: 25387460; PubMed Central PMCID: PMC4228153.
30. Keki Z, Grebner K, Bohus V, Marialigeti K, Toth EM. Application of special oligotrophic media for cultivation of bacterial communities originated from ultrapure water. *Acta microbiologica et immunologica Hungarica*. 2013; 60(3):345–57. Epub 2013/09/26. <https://doi.org/10.1556/AMicr.60.2013.3.9> PMID: 24060557.
31. Shen H, Rogelj S, Kieft TL. Sensitive, real-time PCR detects low-levels of contamination by *Legionella pneumophila* in commercial reagents. *Molecular and cellular probes*. 2006; 20(3–4):147–53. Epub 2006/04/25. <https://doi.org/10.1016/j.mcp.2005.09.007> PMID: 16632318.
32. Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PloS one*. 2014; 9(5):e97876. Epub 2014/05/20. <https://doi.org/10.1371/journal.pone.0097876> PMID: 24837716; PubMed Central PMCID: PMC4023998.
33. Ringel Y, Maharshak N, Ringel-Kulka T, Wolber EA, Sartor RB, Carroll IM. High throughput sequencing reveals distinct microbial populations within the mucosal and luminal niches in healthy individuals. *Gut microbes*. 2015; 6(3):173–81. Epub 2015/04/29. <https://doi.org/10.1080/19490976.2015.1044711> PMID: 25915459; PubMed Central PMCID: PMC4615648.
34. Galloway-Pena JR, Smith DP, Sahasrabhojane P, Ajami NJ, Wadsworth WD, Daver NG, et al. The role of the gastrointestinal microbiome in infectious complications during induction chemotherapy for acute myeloid leukemia. *Cancer*. 2016; 122(14):2186–96. Epub 2016/05/05. <https://doi.org/10.1002/cncr.30039> PMID: 27142181.
35. Bowers RM, Clum A, Tice H, Lim J, Singh K, Ciobanu D, et al. Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. *BMC Genomics*. 2015; 16:856. <https://doi.org/10.1186/s12864-015-2063-6> PMID: 26496746; PubMed Central PMCID: PMC4619416.
36. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*. 2018; 6(1):42. <https://doi.org/10.1186/s40168-018-0426-3> PMID: 29482639