



OPEN ACCESS

EDITED BY

Zheng Kuang,
Carnegie Mellon University, United States

REVIEWED BY

Honghua Hu,
Jinhua Institute of Zhejiang University,
China

Bruno Fosso,
University of Bari Aldo Moro, Italy
Justin P. Shaffer,
University of California, San Diego,
United States

*CORRESPONDENCE

Bahtiyar Yilmaz,
✉ bahtiyar.yilmaz@unibe.ch

†These authors have contributed equally to this work and share first authorship

RECEIVED 11 March 2023

ACCEPTED 10 April 2023

PUBLISHED 26 April 2023

CITATION

Marchukov D, Li J, Juillerat P, Misselwitz B and Yilmaz B (2023), Benchmarking microbial DNA enrichment protocols from human intestinal biopsies. *Front. Genet.* 14:1184473. doi: 10.3389/fgene.2023.1184473

COPYRIGHT

© 2023 Marchukov, Li, Juillerat, Misselwitz and Yilmaz. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Benchmarking microbial DNA enrichment protocols from human intestinal biopsies

Dmitrij Marchukov^{1†}, Jiaqi Li^{2,3†}, Pascal Juillerat^{2,3,4}, Benjamin Misselwitz^{2,3} and Bahtiyar Yilmaz^{2,3*}

¹University Hospital Zürich, University of Zürich, Zürich, Switzerland, ²Department of Visceral Surgery and Medicine, Bern University Hospital, University of Bern, Bern, Switzerland, ³Maurice Müller Laboratories, Department for Biomedical Research, University of Bern, Bern, Switzerland, ⁴Crohn's and Colitis Center, Gastroenterologie Beaulieu, Lausanne, Switzerland

Shotgun metagenomic sequencing is a powerful tool for studying bacterial communities in their natural habitats or sites of infection, without the need for cultivation. However, low microbial signals in metagenomic sequencing can be overwhelmed by host DNA contamination, resulting in decreased sensitivity for microbial read detection. Several commercial kits and other methods have been developed to enrich bacterial sequences; however, these assays have not been tested extensively for human intestinal tissues yet. Therefore, the objective of this study was to assess the effectiveness of various wet-lab and software-based approaches for depleting host DNA from microbiome samples. Four different microbiome DNA enrichment methods, namely the NEBNext Microbiome DNA Enrichment kit, Molzym Ultra-Deep Microbiome Prep, QIAamp DNA Microbiome kit, and Zymo HostZERO microbial DNA kit, were evaluated, along with a software-controlled adaptive sampling (AS) approach by Oxford Nanopore Technologies (ONT) providing microbial signal enrichment by aborting unwanted host DNA sequencing. The NEBNext and QIAamp kits proved to be effective in shotgun metagenomic sequencing studies, as they efficiently reduced host DNA contamination, resulting in 24% and 28% bacterial DNA sequences, respectively, compared to <1% in the AllPrep controls. Additional optimization steps using further detergents and bead-beating steps improved the efficacy of less efficient protocols but not of the QIAamp kit. In contrast, ONT AS increased the overall number of bacterial reads resulting in a better bacterial metagenomic assembly with more bacterial contigs with greater completeness compared to non-AS approaches. Additionally, AS also allowed for the recovery of antimicrobial resistance markers and the identification of plasmids, demonstrating the potential utility of AS for targeted sequencing of microbial signals in complex samples with high amounts of host DNA. However, ONT AS resulted in relevant shifts in the observed bacterial abundance, including 2 to 5 times more *Escherichia coli* reads. Furthermore, a modest enrichment of *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* was also observed with AS. Overall, this study provides insight into the efficacy and limitations of various methods for reducing host DNA contamination in human intestinal samples to improve the utility of metagenomic sequencing.

KEYWORDS

gut microbiome, host DNA depletion, metagenomic, phyloseq, human small intestine, microbial enrichment

1 Introduction

The gut microbiota is a complex community of microorganisms living in the mammalian digestive tract (Martin et al., 2007; Hilt et al., 2014; Sender et al., 2016). The highly co-evolved mutualism between inhabitants on our body surfaces and the host immune system has promoted beneficial co-existence and interdependency over millions of years (Young, 2017). The role of the bacterial microbiota in maintaining homeostasis starts at birth and continues throughout life (Dominguez-Bello et al., 2010; Mueller et al., 2015). It is notably evident that viable gut microbiota is crucial for maintaining the host health status (Lloyd-Price et al., 2019), and this is in nearly everyone's interests to keep the habitat and its distinctive niches healthy (Stecher et al., 2005; Smith et al., 2007; Hooper and Macpherson, 2010; Yilmaz et al., 2014; Macpherson et al., 2018; Uchimura et al., 2018; Lynch and Hsiao, 2019). The composition of the gut microbiota remains relatively stable over the years within individuals in the absence of major events such as medications or surgery. Over time, gut microbial strains undergo genetic changes *via* various mechanisms (e.g., mutations, horizontal and vertical gene transfer), and selection resulting in rapid adaptation and/or long-term evolution of sub-strains. These processes can lead to positive and negative dynamic structural and functional changes in the gut, which in turn might also impact human health (Yilmaz et al., 2021).

Changes in the gut microbiota have been associated with a wide range of diseases, including inflammatory bowel diseases (IBD) (Lloyd-Price et al., 2019; Yilmaz et al., 2019), celiac diseases (Olivares et al., 2018), colorectal cancer (CRC) (Feng et al., 2015; Liang et al., 2017; Yu et al., 2017) (Cheng et al., 2020), chronic inflammation and metabolic diseases (Cox et al., 2014) such as obesity (Ley et al., 2005) (Backhed et al., 2004; Turnbaugh et al., 2006) and diabetes (Stewart et al., 2018; Zhou et al., 2022). However, studying the intestinal microbiome in the context of these diseases poses unique technical challenges. Identifying the diversity of gut microbiota using culture-based methods can be a laborious and time-consuming process that is often unable to capture the full range of microbial species present. However, some studies have attempted to address this limitation by using over 60 different culture conditions to isolate the most abundant taxa. In these studies, the researchers were able to successfully culture an average of 95% of the operational taxonomic units (OTUs) present at greater than 0.1% abundance in fecal samples. (Browne et al., 2016; Lau et al., 2016). In recent years, molecular-based approaches that do not rely on cultivation, such as 16 S rRNA gene sequencing and metagenomics, have brought a paradigm shift to our comprehension of the human microbiome's involvement in health and disease. These methods enable a thorough examination of the microbial community, including the detection of previously un-cultivable bacteria and the evaluation of their functional potential (Loman et al., 2012). Although 16 S rRNA amplicon sequencing is an expeditious and cost-effective approach for identifying the taxonomic composition of a sample (Schriefer et al., 2018), it is insufficient for characterizing the functional landscape of the gut microbiome to answer inquiries regarding microbial activities (Franzosa et al., 2018). Therefore, alternative strategies, such as shotgun metagenomic approaches, are needed to investigate the functional potential of the microbiome (Qin et al., 2010).

Shotgun metagenomic sequencing allows the simultaneous analysis of all genetic material present in a sample, regardless of the organisms. This approach enables the identification of numerous genes and their variants, along with the reconstruction of enzymatic pathways, thereby providing valuable insights into the functional capabilities of the microbial community (Ranjan et al., 2016; Robinson et al., 2021). In addition, recent advancements in microbiome research have expanded our ability to investigate the microbiota's functional and genetic profile in specific regions of the intestine, due to the development of bacterial profiling techniques that can be applied to biopsies rather than stool samples possible (Korem et al., 2015; Suez et al., 2018; Saffarian et al., 2019; Yilmaz et al., 2019; Zeevi et al., 2019; Yilmaz et al., 2022). This approach has been successfully utilized to unravel the molecular and cellular mechanisms underlying gut-associated diseases. For instance, a study conducted by Franzosa et al. utilized shotgun metagenomic sequencing of colonic biopsies to identify gene-level differences in the microbial community between patients with Crohn's disease and healthy individuals (Franzosa et al., 2019). This approach revealed that Crohn's disease was associated with significant alterations in bacterial metabolic pathways, including amino acid metabolism, energy production, and xenobiotic biodegradation. Furthermore, the use of biopsy-based bacterial profiling has allowed for a better understanding of the microbial communities' spatial organization in the intestine, with studies showing differences in microbial composition and diversity across various intestinal regions, including the duodenum, jejunum, ileum, and colon. For instance, a study by Leite et al. (2020) using 16 S rRNA gene sequencing of duodenal biopsies from healthy individuals revealed a distinct microbial community structure compared to that observed in fecal samples, highlighting the importance of analyzing specific regions of the intestine to gain a more comprehensive understanding of the microbiota's functional and genetic profile. Therefore, the use of biopsy-based bacterial profiling has provided a promising avenue for investigating the molecular and cellular mechanisms underlying gut-associated diseases and holds great potential for future microbiome research.

Bacterial metagenomic sequencing requires a sufficient abundance of microbial DNA without large-scale host DNA contaminations. However, biopsies or whole-tissue isolates contain large bulks of host DNA, vastly outnumbering microbial DNA (de Albuquerque et al., 2022). This phenomenon is not limited to intestinal biopsies. We have recently conducted a study with ileostomy patients, where we identified highly dynamic components of the microbiota present in the small intestine. These components were found to be highly responsive to dietary changes introduced after an overnight fast. (Yilmaz et al., 2022). The ratio of microbial/host DNA oscillates in accordance with fasting and feeding, and the appearance and disappearance of microbial sub-strains were also strongly associated with the provision of nutrition. This results in a higher ratio of host/microbe DNA in the fasting state, while the introduction of food leads to blooming in bacterial populations and a lower ratio of host/microbe DNA (Yilmaz et al., 2022). In this type of situation, characterizing the functional and genetic profile of low-abundance bacteria in microbiome samples can be a difficult task, particularly if the sample is contaminated with host DNA. To overcome this challenge, it is essential to perform host DNA depletion as the first step before conducting deep shotgun

metagenomic sequencing. By removing host DNA upstream of sequencing, it is possible to increase the detection of low-abundance bacteria, which would otherwise remain undetected. Nevertheless, the optimal approach to achieve host DNA depletion remains undetermined.

To reduce host DNA content and increase the yield of microbiota DNA prior to sequencing (Heravi et al., 2020), several commercial kits and general laboratory methods have been developed over the past few years. Some of them have already proven useful in enriching microbial DNA from liquid samples such as saliva (Marotz et al., 2018), blood (Feehery et al., 2013), sonicated fluid from prosthetic joint components (Thoendel et al., 2016), human milk (Yap et al., 2020), and cerebrospinal fluid (Hasan et al., 2016; Ji et al., 2020), but also from solid materials such as human breast tissue (Costantini et al., 2018), and tissue from an infected diabetic foot (Heravi et al., 2020). Surprisingly, to the best of our knowledge, commercially available kits for the depletion of host DNA from intestinal tissues have not been systematically compared or tested, except for a single study that developed a new technique to address this challenge. In this study, researchers optimized the sample lysis step by incorporating additional detergents and bead-beating protocols to achieve an efficient host DNA depletion (Bruggeling et al., 2021). The method was demonstrated to be relatively effective in reducing host DNA contamination in human fecal and mucosal samples. Notably, the approach resulted in higher bacterial read yield and a more accurate representation of the microbial community compared to the commercially available kits.

The present study aimed to evaluate the effectiveness of host DNA depletion methods, including commercially available kits, a laboratory-optimized protocol, and the software-controlled enrichment approach of Oxford Nanopore, in enhancing bacterial DNA yields from human intestinal biopsy samples. Our main objective was to enhance bacterial DNA yields from human intestinal biopsy samples, and the findings highlight the limitations of existing microbial DNA enrichment tools and the potential impact of different enrichment methods on the identification of bacterial groups. Despite the challenges of host DNA depletion from intestinal biopsies, we were able to increase the proportion of bacterial DNA to 30%–45% of total DNA in some cases. We also observed that different microbial enrichment methods could lead to shifts in the proportion of identified bacteria groups in each sample. Interestingly, we observed that the software-controlled enrichment approach of Oxford Nanopore increased the base pair numbers associated with bacterial DNA allowing to assemble the genome metagenomically but did not significantly increase the percentage of bacterial reads compared to commercially available kits. Our study provides valuable insights into the effectiveness of different microbial DNA enrichment methods and highlights the potential for the software-controlled enrichment approach of Oxford Nanopore to improve bacterial DNA yield and enable the identification of rare intestinal bacteria.

2 Materials and methods

2.1 Sample collection and ethics statement

To evaluate the performance of four host DNA depletion kits and one total DNA extraction control kit, we collected endoscopic

biopsies (~2 mm) from 10 subjects, with five biopsies per subject. Additionally, we included a second test group comprising 10 subjects to assess the effectiveness of a non-commercial method (Bruggeling et al., 2021) in combination with two commercial kits. In this group, three biopsies per subject were collected. The procedures used in this study included total DNA extraction, microbial DNA enrichment, and two variations of the published laboratory-optimized depletion method (Bruggeling et al., 2021).

The Bern Human Intestinal Community project was approved by the Bern Cantonal Ethics Commission (Ref: KEK-BE: 251/14) with signed informed consent obtained from all participants. Additionally, the bowel cleansing study was approved by ethical approval number 336/2014. Biopsy samples were collected from subjects registered for a screening ileo-colonoscopy without any gastrointestinal symptoms and without functional intestinal symptoms and negative results in all additional workups. The cohort comprised of 15 males and 5 females within the age range of 40–60 years old. It is noteworthy that none of the participants had taken antibiotics or any regular medications for the 6 months preceding the sampling. The licensed gastroenterologists collected biopsy samples and clinical data of all healthy subjects. Colonic biopsies were initially collected into 2 mL microfuge tubes containing 500 μ L RNeasy (Sigma-Aldrich) and stored at -20°C until DNA extraction.

Structured clinical metadata were prospectively gathered based on pre-determined standards, documented electronically using Research Electronic Data Capture (REDCap) (Harris et al., 2009) and handled in R (<http://www.r-project.org>) using the *xlsx* and *data.frame* packages. Assessment of the microbiota composition from intestinal biopsies was then analyzed according to numerous parameters such as age, sex, and sampling location. Statistical analyses were performed using Student's *t*-test, Wilcoxon's rank sum test, and Pearson's chi-squared test to assess differences between groups.

2.2 Microbial DNA extraction and host DNA depletion

Microbial DNA extraction is a critical step in metagenomic studies that can significantly impact downstream analyses. To ensure reproducibility and minimize bias, we followed the manufacturer's instructions for DNA extraction using commercially available kits. Specifically, we employed the kits shown in Figure 1, which have been extensively validated and optimized for high yield and purity of DNA from a diverse range of microbial samples. However, these kits and computer-based approaches have not been extensively tested for human intestinal biopsies.

2.2.1 Total DNA/RNA extraction

Total DNA was isolated using AllPrep DNA/RNA Mini Kit (Qiagen) as described before (Yilmaz et al., 2019). 600 μ L of RLT Plus Buffer containing 6 μ L beta-mercaptoethanol and a 3 mm bead were added into the tube. Biopsies were homogenized by the Retsch Tissue Lyser (Qiagen) at 30/revolution for 5 min. Supernatants were transferred into AllPrep DNA mini spin column and centrifuged at

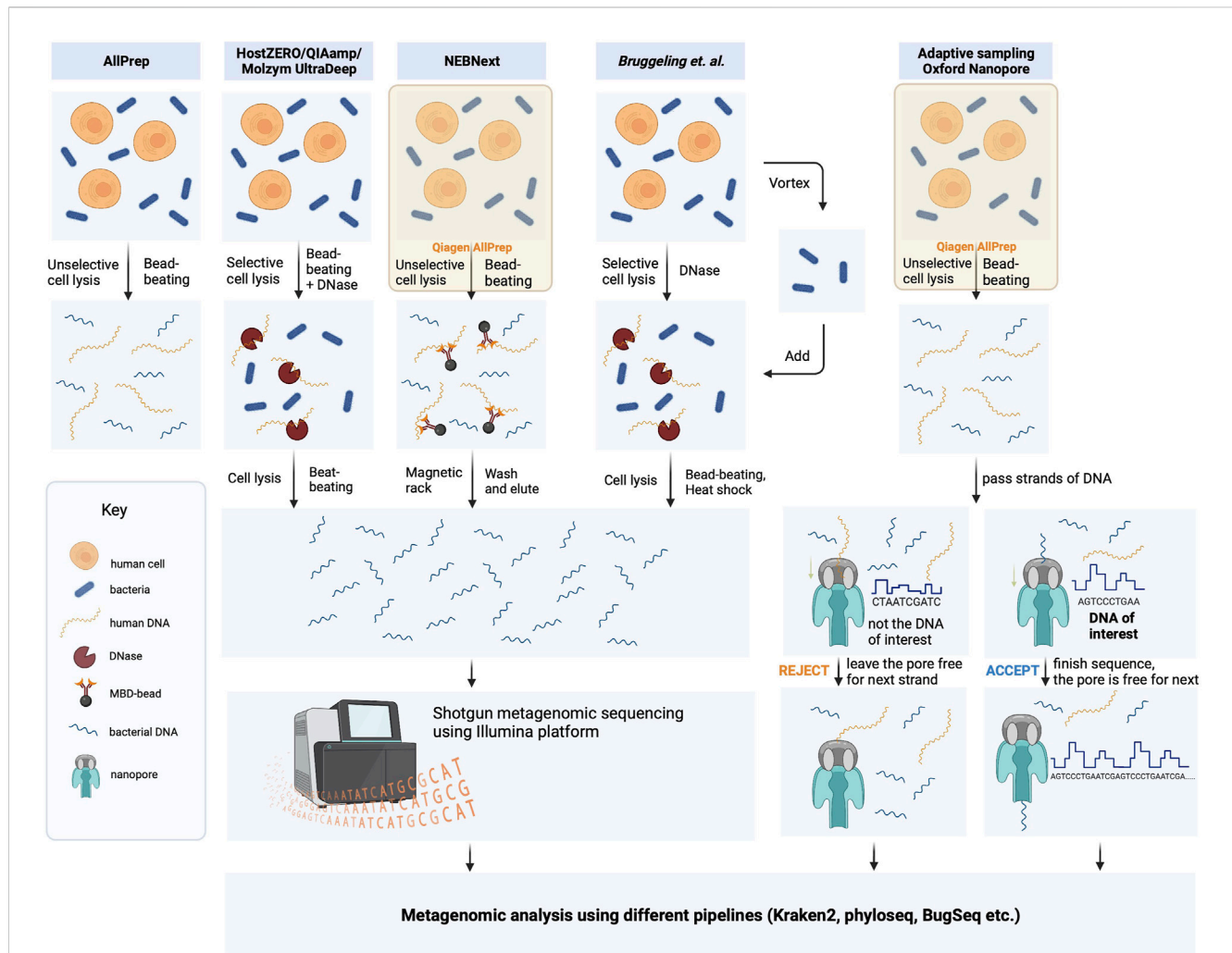


FIGURE 1

Schematic representation of DNA isolation protocol strategies used in this study. The bacterial DNA was extracted from biopsies using five different methods, which included protocols provided by the manufacturers and the laboratory-optimized protocol developed by Bruggeling and colleagues. (Bruggeling et al., 2021). These methods comprised unselective cell lysis kits (Qiagen AllPrep and NEBNext) and selective cell lysis kits (HostZero/QIAamp Microbiome and Molzym UltraDeep), with or without microbiome enrichment. The resulting DNA samples were then sequenced using Illumina NovaSeq 6,000 in 150 bp paired-end mode. Additionally, DNA samples obtained through the Qiagen AllPrep DNA/RNA extraction kit were also analyzed using adaptive sequencing technologies from Oxford Nanopore.

9000 g for 30 s. DNA attached to spin columns was subjected to clean-up using 500 µL of Buffer AW1 and AW2 afterwards. As a last step, DNA samples were eluted with 30 µL nuclease-free water into 1.5 mL microfuge tubes and stored (-20°C) until proceeding with downstream steps. The concentration and purity of the isolated DNA samples were evaluated by NanoDrop® (Thermo Scientific). Of note, RNA was extracted following the protocol instructions even though not used in our study.

2.2.1.1 HostZERO microbial DNA kit

This kit initially applies the physical homogenization of tissue samples with bead-beating, followed by selective chemical lysis of eukaryotic host cells using the Host Depletion Solution, with the intention to keep microbial cells intact. In the host depletion part of the protocol, 200 µL Host Depletion was added to each sample and incubated at room temperature for 15min. Following centrifugation

(10000g, 5min), the supernatant was discarded, and 100 µL of Microbial Selection Buffer and 1 µL of Microbial Selection Enzyme were added to each tube for incubation at 37°C for 30 min. To enhance the depletion of host DNA, 20 µL of Proteinase K were added to the sample and incubated at 55°C for 30 min 100 µL of DNA/RNA Shield™ (2X Concentrate) was added. For microbial DNA isolation, each sample was treated with ZR BashingBead™ Lysis Tube and 750 µL of ZymoBIOMICS® Lysis Solution. The Retsch Tissue Lyser (Qiagen) was used for 5 min at a frequency of 30/min. Next, 400 µL of supernatant was transferred to another collection tube. After adding 1200 µL of ZymoBIOMICS® DNA Binding Buffer to the tubes, mixing was done by thoroughly pipetting the entire volume up and down five times. Each sample was then transferred to the Zymo-Spin™ IC-Z Column, and ZymoBIOMICS® DNA Wash Buffer 1 and Wash Buffer 2 were applied in the washing step. After the final centrifugation (10000g,

2 min) of the washing steps, 30 μ L DNase-free water was applied to the center of the Zymo-Spin™ IC-Z Column. DNA was eluted by centrifugation at 10000g for 1 min and stored at -20°C .

2.2.2 QIAamp DNA microbiome kit

This kit also works based on the principle of lysing host cells first and depleting host DNA enzymatically while keeping microbial cells intact for the downstream microbial DNA extraction process. Briefly, 500 μ L of Buffer AHL was added to each tube, followed by incubation at room temperature for 30 min. After a centrifugation step at 10000 g for 10 min, the supernatant was removed. 190 μ L of Buffer RDD and 2.5 μ L of Benzonase were added into each tube and incubated at 37°C for 30 min on a thermomixer with shaking at 600rpm. After the addition of 20 μ L Proteinase K, samples were again incubated at 56°C for 30 min on a thermomixer (600rpm). Then, 200 μ L Buffer ATL was added to each sample and transferred into Pathogen Lysis Tube L. After lysis, samples were heated to 95°C for 5min. Following centrifugation, 40 μ L Proteinase K was added to each sample, vortexed, and incubated at 56°C for 30 min. Next, 200 μ L Buffer APL was added to each tube followed by incubation at 70°C for 10 min. Afterwards, 200 μ L ethanol was added to the lysate and mixed by pulse-vortexing for 15–30 s. 700 μ L of the mixture was then transferred into the QIAamp UCP Mini spin column. Washing steps using 500 μ L of Buffer AW1 and AW2 were done following the instructions of the protocol. After the final centrifugation of the washing steps, 30 μ L DNase-free water was applied to the center of the membrane. DNA was eluted by centrifuging at 10000 g for 1 min and stored at -20°C .

2.2.3 Molzym ultra-deep microbiome prep kit

This kit utilizes a combination of mechanical and enzymatic lysis to effectively release DNA from cells and includes a bead-beating step for efficient cell lysis. This allows the degradation of free-floating and human DNA and isolates the genomic DNA of microbes. The DNA was then purified using silica-based spin column technology. To reduce the interference of host DNA in microbial DNA sequencing, this method selectively lyzed human cells using CM buffer, followed by degradation of host-released DNA using human DNase (MolDNase B), leaving bacterial cells intact. Bacterial cells were then concentrated by centrifugation, and DNA was extracted using enzymes that specifically target bacterial cell walls. To ensure consistency and reproducibility, all steps of the protocol were conducted precisely following the manufacturer's instructions.

2.2.4 NEBNext microbiome DNA enrichment kit

In contrast to the three kits described above, the starting material of this kit is total DNA. The NEBNext Microbiome DNA enrichment kit (New England Biolabs) contains magnetic beads that selectively bind to the CpG-methylated host DNA (Feehery et al., 2013). This step facilitates the enrichment of bacterial DNA and the depletion of host DNA. Briefly, total DNA was extracted using the AllPrep DNA/RNA Mini Kit (Qiagen) using a slow bead beating step at a frequency of 10/min for 5 min. Host methylated DNA was captured using 160 μ L of MBD2-Fc-bound magnetic beads and prepared according to the manufacturer's instructions. The undiluted bind/wash Buffer (5X) was added to make the final

concentration 1X and tubes were agitated by rotating at room temperature for 15 min. Each tube was then placed on the magnetic rack for 5 min until the beads were collected on the wall of the tube and the solution was clear. The supernatant was removed without disturbing the beads and transferred to a clean microcentrifuge tube. This supernatant contains the target microbial DNA. Afterwards, the ethanol precipitation protocol was followed to elute captured host DNA. 2.5X pure ethanol was added to each sample and then incubated on ice for 10 min. Afterwards, ethanol was removed by a centrifugation step at 16000 g for 30 min and pellets were air-dried.

2.2.5 Host DNA depletion by the method proposed by Bruggeling et al.

Bruggeling et al. (2021) proposed a bacterial DNA isolation method optimized for the human gut biopsy tissue. We followed the protocol described in this manuscript. Briefly, bacteria loosely bound to the surface of the biopsy were separated by vortexing and transferred to another microfuge tube. 20 μ L proteinase K and 0.0125% saponin were then used for digestion and lysis of human cells while keeping bacterial cells intact. The resulting cell suspension thus contained lyzed human cells and intact bacterial cells. Next, DNase treatment was applied for human DNA depletion. In the end, each sample had reduced human DNA content and intact bacteria. Bacterial cells were then lyzed by utilizing a specialized bead-beating protocol, using 0.5 KU/mL mutanolysin (Sigma) and a brief heat shock to ensure susceptibility for mechanical lysis.

We altered the protocol slightly to increase the yield of microbial DNA: In step 3 of the original protocol, instead of vortexing the tubes for 5 min samples were put in a bead-beater at 10 Hz for 5 min without beads. Further, instead of adding 2 μ L TurboDNase in 10x Turbo DNase buffer, we used 2.5 μ L DNase I in Buffer RDD of the RNase Free DNase Set by Qiagen. Moreover, in the original protocol, 20 μ L mutanolysin per sample was used, but we reduced it to half. HostZERO Microbial DNA Kit and the QIAamp DNA Microbiome Kit were tested separately for bead-beating and microbial DNA extraction.

2.2.6 Library preparation and shotgun metagenomic sequencing

Due to a low DNA concentration in some groups of samples, we prepared the libraries using the Nextera XT kit, which requires a minimum of 2 ng DNA as the starting material. DNA libraries were prepared according to the Nextera DNA Library Prep (Illumina) as instructed in the manufacturer's protocol and sequenced on NovaSeq 6,000 (Illumina, 150bp, PE mode). The metaWRAP-Read_qc module was applied to filter out the human genome-contaminated reads, remove adaptor sequences, and low-quality reads, and produce quality reports for each of the sequenced samples prior to the microbial abundance estimation (Uritskiy et al., 2018). This pipeline contains the FASTQC (Andrews, 2015) and the BMTagger modules (Rotmistrovsky, 2011).

Every sample was subjected to Illumina sequencing, while an additional 5 samples from the initial group that were extracted using AllPrep later underwent ONT sequencing as well.

Before conducting any subsequent diversity and taxonomy analysis, the read counts of each sample were divided by the total

number of reads in that sample, using the library size normalization method. The taxonomy profile was assessed using the Kraken2 pipeline with a custom RefSeq database following the developer's guideline (Wood et al., 2019) and the Kraken report was generated in <https://github.com/DerrickWood/kraken2/blob/master/docs/MANUAL.markdown#custom-databases>. To generate the CustomDB taxonomy directory with the necessary information, we executed the `kraken2-build` command with the "--download-taxonomy" option. This allowed us to obtain the accession number to taxon maps, taxonomic names, and tree information from NCBI. However, the information was limited to complete genomes of archaea, bacteria, fungi, plants, protozoa, and virus. For taxonomy classifications, we utilized Kraken2, with the following command line serving as a representative example: `kraken2 --use-names--db/home/ubelix/dbmr/terziev/CustomDB--fastq-input--report-zero-counts--confidence 0.1 --threads 12 --minimum-base-quality 0 --paired--gzip-compressed {input_1.fastq.gz} {input_2.fastq.gz} --output {output.reads} --report {output.report} > input.kraken`.

2.2.7 Library preparation and sequencing with Oxford Nanopore Technologies (ONT)

1 µg of high molecular weight DNA samples obtained using AllPrep DNA/RNA kit were used for Oxford Nanopore adaptive sequencing. Library preparation was performed using the SQK-LSK110 kit (Oxford Nanopore Technologies, Oxford) following the genomic DNA ligation protocol (<https://community.nanoporetech.com/protocols/genomic-dna-by-ligation-sqk-lsk110/>). Finally, the libraries were loaded separately onto different Nanopore R9.4.1 flow cells (FLO-MIN106), one for sequencing with Adaptive sampling (AS) and one for control sequencing. Both flow cells were run simultaneously on a GridION X5 device (MinKNOW version 21.11.7; Guppy 5.1.13; Oxford Nanopore Technologies [ONT]) for up to 72 h.

2.2.8 Bioinformatic analysis of Oxford Nanopore Technologies (ONT) adaptive sampling (AS) sequencing data

The output of AS sequencing runs consists of nanopore reads in a FASTQ format, accompanied by a csv file that lists the classification of each read made by the ReadUntil API (https://github.com/nanoporetech/read_until_api) available in the MinKNOW interface on the GridION machine. This classification is based on read matching to the user-provided reference sequence(s), as follows: Under a "depletion" AS runs as follows: the initial 400–600 bases of a strand that translocate through a given pore are used to classify the reads by the ONT's ReadUntil API. Each read that passed through the nanopore was aligned to the human reference genome while it was being sequenced. The alignment occurred at intervals of several bases, and three types of decisions were made: 1) "no_decision"—the read was continued and realigned to the reference(s) after several bases ("no decision"), 2) "stop_receiving"—the read was fully sequenced and accepted ("accepted"), and 3) "unblock"—the sequencing was immediately terminated, and the read was rejected. In the "rejected" case, the voltage is reversed at the pore level and the DNA will be expelled from the pore, preventing further sequencing Payne et al. (2021).

Taxonomy classification and quality control analysis of long-read sequences were performed using the BugSeq workflow which uses a combination of tools and databases to classify reads into taxonomic groups and also identify potential contaminant sequences (Fan et al., 2021; Chandrakumar et al., 2022). Briefly, the reads were quality-controlled using *fastp* with a minimum read length of 100bp and minimum average read quality of Phred 7 (Chen et al., 2018). Then, the reads were mapped to BugSeq's curated database containing microbial sequences, the human genome, and contaminants using *minimap2* in "map-ont" mode with the "-a" flag (Li, 2018). The alignments were then reassigned using *Pathoscope* based on a Bayesian statistical framework and the lowest common ancestor of alignments was taken (Francis et al., 2013). Finally, the lowest common ancestor of the reassigned reads was calculated using *Recentrifuge* (Marti, 2019). The output obtained with this pipeline was saved in csv files which were then used to prepare the corresponding tables and figures in this study using GraphPad Prism Version 9.5.1.

2.2.9 Statistical analysis

All statistical analyses were performed using R version 3.6.1 or Prism 9 (GraphPad Software, San Diego, CA). Differences between the groups after library size normalization were evaluated using one-way ANOVA (parametric), followed by Tukey's honest significant difference test or the two-stage step-up method of Benjamini, Krieger, and Yekutieli, as a *post hoc* test. The effect size was calculated using Cohen's *d* in Excel using the following formula = ABS (AVERAGE (group1) - AVERAGE (group2))/ (SQRT (((COUNT (group1) - 1) * STDEV. S (group1, group2)² + (COUNT (group2) - 1) * STDEV. S (group2, group1)²)/(COUNT (group1) + COUNT (group2) - 2))).

The computation of alpha and beta diversity was carried out using different metrics, such as the Shannon index for alpha diversity and Aitchison distance for beta diversity. These computations were performed using the *phyloseq* package in R. (McMurdie and Holmes, 2013; Callahan et al., 2016). Statistical analyses were performed using Mann-Whitney U tests for alpha diversity and Adonis (PERMANOVA) for beta diversity with pairwise comparison (Benjamini-Hochberg false discovery rate correction) using *pairwiseAdonis* R package to confirm the strength (McMurdie and Holmes, 2013; Callahan et al., 2016). Microbial changes were tested using multivariate analysis by linear models (MaAsLin2) R package (Morgan et al., 2012; Mallick et al., 2021). Differences of $p < 0.05$ or $\text{adj-}p < 0.05$ were considered significant in all statistical analyses.

3 Results

3.1 Commercial kits can enrich the microbial DNA but cannot entirely deplete host DNA from intestinal biopsy samples

Depletion of host DNA from biopsies is crucial for identifying the bacteria present in a specific region of the human intestine, as well as for characterizing the genetic features of these bacteria *via* metagenomic analysis. We evaluated several commercially available DNA depletion kits, including the HostZERO Microbial DNA kit,

TABLE 1 Quality control of DNA extracted by the multiple kits from endoscopic biopsy samples. DNA concentration, purity, and DNA integrity number (DIN) are recorded for each sample. The average values with standard deviations (\pm) are shown for each category.

Extraction Method	Sample Size	DNA concentration (ng/ μ L)	260/280 ratio	260/230 ratio	DN
AllPrep	10	173.28 \pm 204.3	1.90 \pm 0.01	2.35 \pm 0.05	8.25 \pm 0.05
NEBNext	6	19.4 \pm 6.8	2.01 \pm 0.30	2.11 \pm 0.30	8.35 \pm 0.15
HostZero	8	1.3 \pm 1.1	1.75 \pm 0.21	0.39 \pm 0.21	7.55 \pm 0.13
Mozym	10	53.23 \pm 46.1	2.10 \pm 0.25	0.85 \pm 0.05	8.55 \pm 0.35
QIAamp	9	2.92 \pm 1.9	2.07 \pm 0.22	0.47 \pm 0.11	7.25 \pm 0.42
AllPrep2	9	285.4 \pm 185.1	1.85 \pm 0.05	2.15 \pm 0.10	8.15 \pm 0.11
QIAamp + Lab-optimized	8	15.48 \pm 35.6	2.25 \pm 0.27	1.25 \pm 0.55	8.25 \pm 0.15
HostZero + Lab-optimized	6	1.8 \pm 1.7	1.85 \pm 0.40	0.49 \pm 0.01	7.69 \pm 0.13

Molzym Ultra-Deep Microbiome Prep Kit, NEBNext Microbiome DNA Enrichment Kit, and QIAamp DNA Microbiome Kit, to assess their suitability for extracting DNA from intestinal biopsy samples. These kits employ different strategies for lysing host cells and enzymatically degrading host DNA, except for the NEBNext kit, which utilizes a different approach by selectively removing CpG-methylated host DNA from total DNA extracted from AllPrep DNA/RNA Mini Kit (Figure 1).

The concentration and purity of the isolated DNA samples were analyzed by NanoDrop[®] (Thermo Scientific). The purity with different extraction kits was in an acceptable range with a 260 nm/280 nm absorption ratio varying between 1.75 and 2.10 (Table 1). In addition, the DNA was not fragmented with any of the extraction kits, as indicated by a Bioanalyzer 2,100 measured DNA integrity number (DIN) between 7.25 and 8.55. Additional tests indicated some impurities with Molzym, QIAamp Microbiome, and HostZERO tests revealing relatively low 260/230 ratios. This low ratio is likely due to traces of residual guanidine from the lysis buffer used in column-based kits. However, such impurities typically do not affect downstream sequencing analysis, as stated by the manufacturers' application note. Notably, seven samples were excluded from the analysis due to low DNA yield, leaving 43 samples for further analysis.

Metagenomic sequencing using the 150 bp PE mode on an Illumina NovaSeq platform yielded a total of 1'002'860'693 reads from 43 extracted DNA samples (Supplementary Table S1). Of these reads, only 97'015'448 were assigned to the microbial portion using the Kraken2 pipeline with a custom dataset containing microbial and human databases (Supplementary Table S1). Notably, 66'469'799 reads failed to match any of the databases and hence they were not further analyzed. These reads are likely derived from plant DNA in line with our recent study, which showed that Kraken2 could effectively identify plant DNA even in samples with abundant host material (Yilmaz et al., 2022). Since plant DNA was not the focus of this study, these reads were not further analyzed.

The most effective kit for microbial DNA enrichment was the QIAamp DNA microbiome kit, which enriched the

microbial DNA population to ~28% on average. Similarly, the NEBNext kit enriched the microbial DNA population, yielding an average of ~24% microbial DNA (Figure 2A). This demonstrates that some of the commercially available microbial DNA enrichment kits are effective in reducing host DNA from intestinal tissue samples, specifically the NEBNext and QIAamp methods. However, not all extraction kits were able to enrich microbial DNA. Based on the host DNA/microbial DNA ratios, we concluded that the HostZero and Molzym kits did not affect the host DNA content: Extraction with the control AllPrep kit yielded on average ~1.0% microbial DNA, whereas the Molzym and HostZero kits resulted in an average of only ~0.2% and ~7.0% bacterial DNA, respectively (Figure 2A).

We previously demonstrated that investigating SNPs and structural variants in the most abundant taxa requires over ~90% of microbial DNA in a given sample with more than 50 million reads (Yilmaz et al., 2022). However, as shown above, none of the commercial kits was able to reduce the host DNA content but did not achieve the desired purity of \geq 90% (Figure 2A). Therefore, we attempted to optimize the commercial kit protocols (laboratory-optimized protocol) by adding additional vortexing steps as well as saponin incubation steps as described by Bruggeling et al. (2021) before performing all extraction steps of the respective kit. We tested this modification for the QIAamp and HostZero kits and compared the results to those obtained using the AllPrep kit. Our results showed that the AllPrep kit with optimization yielded on average ~3.0% microbial DNA, slightly higher than without the optimization steps (Figure 2). While for the QIAamp kit, the optimization did not significantly enrich microbial DNA (~13%), the HostZero kit with laboratory-based optimization increased microbial DNA yield to up to 20% (Figure 2B). Overall, these results suggest that while commercial kits can reduce host DNA contents in samples, they may not be sufficient for advanced microbial genetic analysis. However, since simple additional steps such as extra vortexing and saponin supplementation improved the efficacy, suggesting that there may be room for further improvement in microbial DNA enrichment protocols.

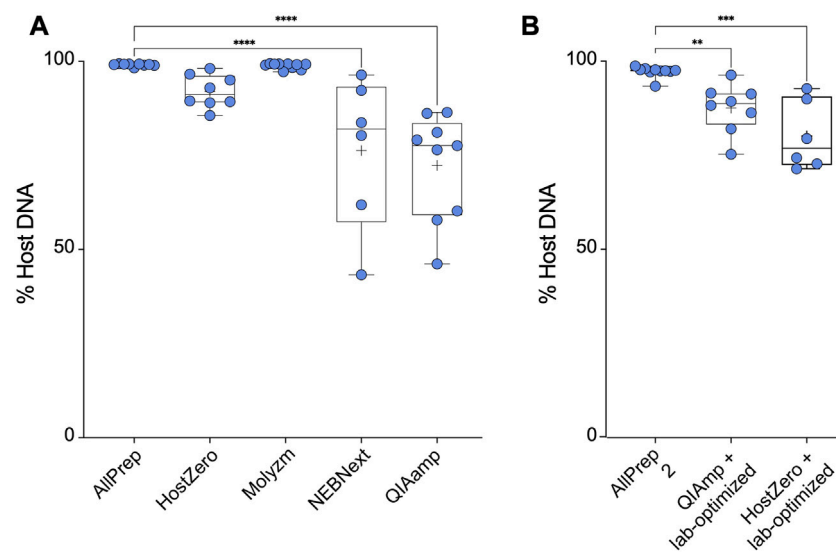


FIGURE 2

Microbial enrichment kits reduce host DNA in extracted DNA samples. **(A)** Percentage of host DNA in samples prepared using different microbiome DNA enrichment methods was compared to the percentage of host DNA extracted using a total DNA extraction kit (AllPrep). **(B)** As in **(A)**, only the methods of the indicated microbiome DNA enrichment kits were modified as described by *Bruggeling et al.* Asterisks for p -values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$. The Cohen's d value for each comparison is as follows: HostZero versus AllPrep ($d = 1.22$), Molyzm versus AllPrep ($d = 0.44$), NEBNext versus AllPrep ($d = 1.36$), QIAamp versus AllPrep ($d = 1.72$), QIAamp + Lab-optimized versus AllPrep ($d = 1.52$) and HostZero + Lab-optimized versus AllPrep ($d = 1.79$).

3.2 Impact of host depletion with different extraction kits on microbial community abundance and composition

The level of microbial DNA enrichment differs among these kits. Although none of these enrichments were adequate for conducting a comprehensive genetic profile of the most prevalent taxa in the gut, we next investigated the similarity of the microbial community composition in samples extracted with different kits (Figure 3). The microbiota of the HostZero and QIAamp kits show a greater diversity of species, even with the laboratory-optimized optimization protocol (Figure 3A). Further, relative composition differences of the intestinal microbiota were found with NEBNext, HostZero and QIAamp kits compared to standard AllPrep kit, assessed by PCA with Aitchison distance ($p < 0.01$) (Figure 3B). These findings remained robust, also when the laboratory-optimized protocol was utilized ($p < 0.05$). Furthermore, our analysis of all samples collectively revealed a positive correlation between microbial enrichment and alpha diversity, with statistical significance observed ($p < 0.05$) (Supplementary Figure S1A). Additionally, Supplementary Figure S1B reveals a clustering pattern among the samples based on their microbial DNA abundance.

Variations in host depletion methods affected certain phyla and families of bacteria much stronger than others. Specifically, we observed significant changes in the Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria phyla (Figures 3C, D; Supplementary Table S2). Within the Proteobacteria phylum, the Pseudomonadales and Enterobacterales orders, as well as the Xanthomonadaceae (*Stenotrophomonas* genus), Burkholderiaceae, Enterobacteriaceae, and Sphingomonadaceae families, were more enriched after

subjecting to extraction using HostZero, NEBNext, and QIAamp kits compared to the standard AllPrep kit. Moreover, HostZero and QIAamp kits yielded in an increase in the relative abundance of the Micrococcaceae family (*Micrococcus luteus* species) from the Actinobacteria phylum, as well as the Staphylococcaceae family (*Staphylococcus aureus* species) and Streptococcaceae family (*Streptococcus* genus) of the Firmicutes phylum, even though to a lesser extent (Figure 3C).

Taxa shifts could also be demonstrated when the lab-based optimization protocol was applied before the usage of the HostZero and QIAamp kits (Figure 3D). Changes in taxa observed with the HostZero lab-optimized approach were similar to those without the lab-optimized protocol shown in Figure 3C. Furthermore, under these conditions, the HostZero kit led to the enrichment of the Bacteroidaceae family (*Bacteroides* genus) of Bacteroidetes and the Lachnospiraceae family (*Lachnoclostridium* and *Blautia* genera) of Firmicutes, but the QIAamp kit did not show this effect (Figure 3D). These findings suggest that shifts in the observed bacterial compositional due to the usage of different host DNA depletion kits affect only a relatively small number of taxa, primarily from the Proteobacteria and Firmicutes phyla and less from Actinobacteria and Bacteroidetes.

3.3 Bacterial sequence enrichment in human intestinal samples using ONT adaptive sampling

Our findings showed that the ability to detect bacterial strains in intestinal biopsies with high levels of host DNA (>98%) is rather insufficient or ineffective with the existing wet-lab procedures of

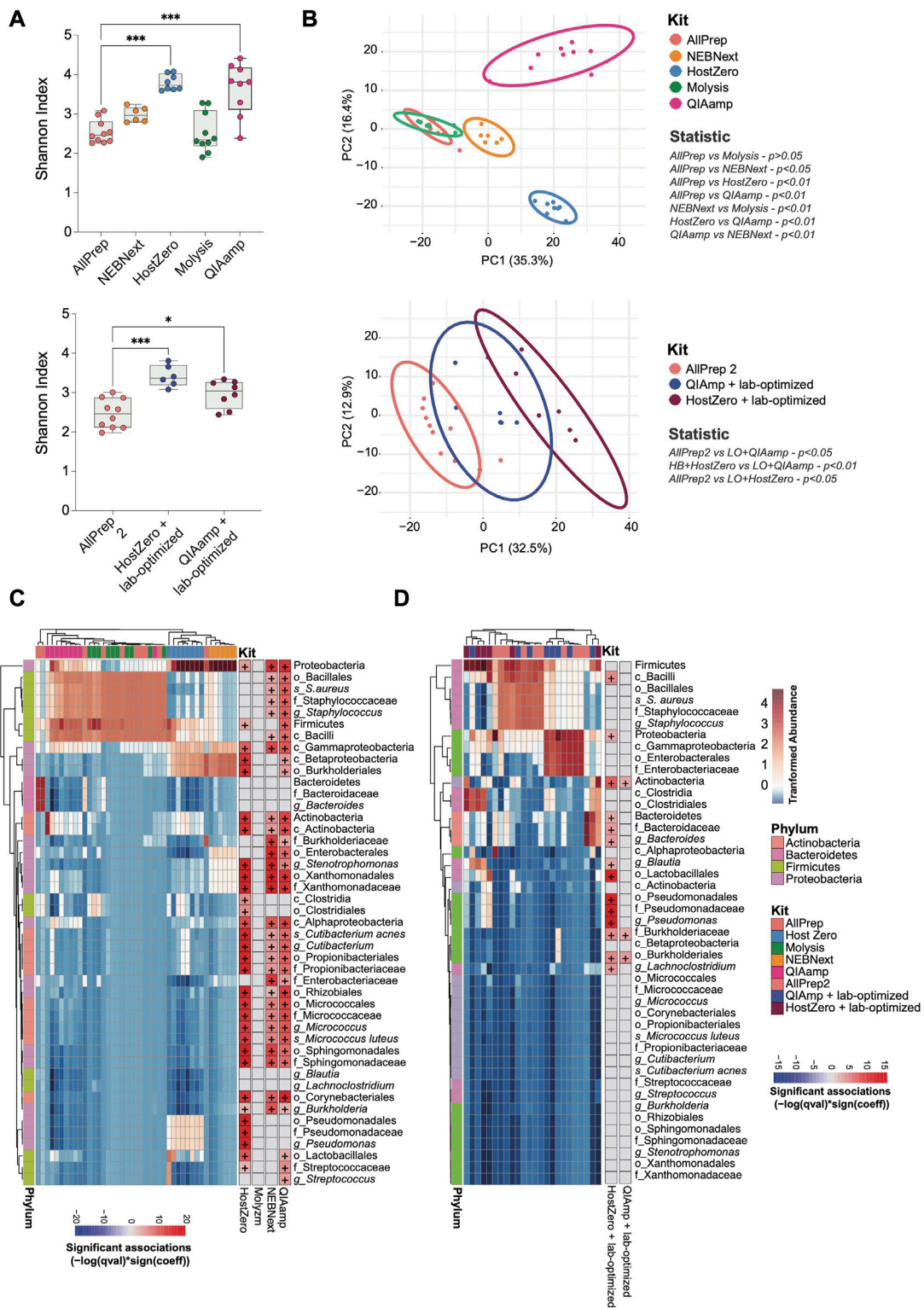


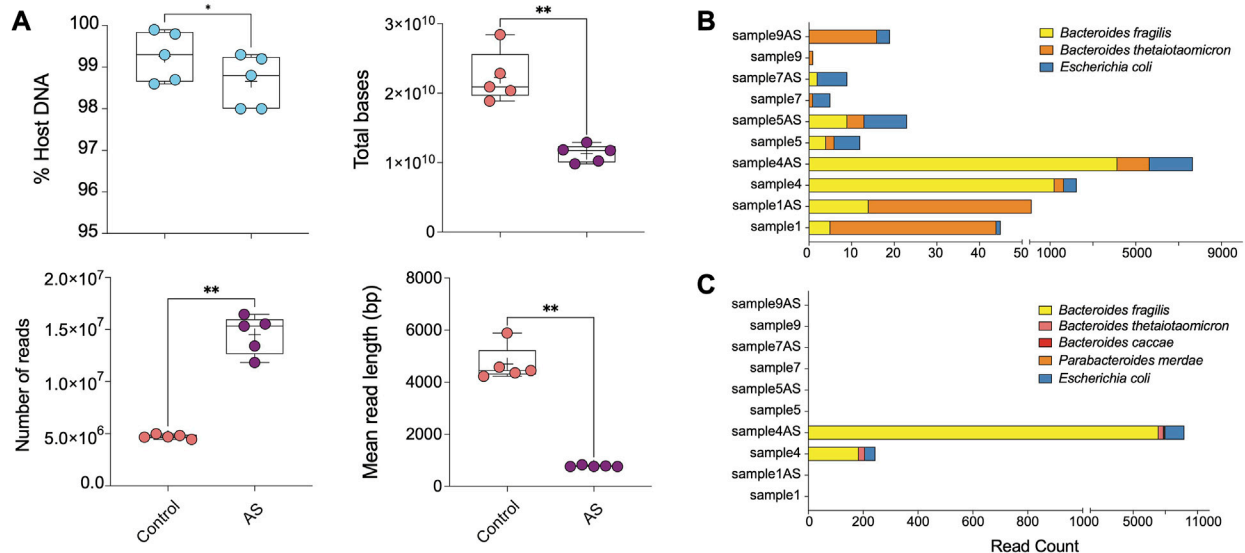
FIGURE 3

Shifts in the observed bacterial composition induced by various host depletion kits. The bacterial DNA of human intestinal biopsies was analyzed by shotgun metagenomic analysis. (A) Alpha diversity between different host depletion kits was measured using the Shannon index and presented on box-and-whisker plots displaying quartiles, range, and standard deviations. (B) Differences in microbial composition between the groups were analyzed with Aitchison distance. Ellipsoids represent the 95% confidence interval of the position of each group. The non-parametric Mann-Whitney U-test and the Adonis test were used to determine statistically significant differences between groups regarding alpha diversity (A) and beta diversity (B), respectively. (C) A heatmap was generated to show the relative abundance of each taxon that differed between host depletion kits compared to the standard AllPrep kit. (D) A similar heatmap was generated for host depletion kits combined with the lab-optimized approach compared to the standard AllPrep kit. A p -value less than 0.05 was considered significant, and significant taxa are shown on the right panel of each heatmap with the color representing associations calculated using $-\log(q\text{-value}) \cdot \text{sign}(\text{coefficient})$, where "+" represents

(Continued)

FIGURE 3 (Continued)

an adj-*p*-value <0.05. Asterisks for *p*-values: **p* < 0.05, ***p* < 0.01, ****p* < 0.001 and *****p* < 0.001. The Cohen's *d* value for each comparison is as follows for (A): HostZero versus AllPrep (*d* = 1.33), Molyris versus AllPrep (*d* = 1.82), NEBNext versus AllPrep (*d* = 0.03), QIAamp versus AllPrep (*d* = 1.47), QIAamp + Lab-optimized versus AllPrep (*d* = 1.15) and HostZero + Lab-optimized versus AllPrep (*d* = 1.65).



selective lysis of host and microbial cells or selective removal of CpG-methylated host DNA. Therefore, we carried out an alternative to lab-based depletion or enrichment approaches which is based on a software-controlled enrichment method by depletion of unwanted DNA during sequencing with providing a target DNA sequence (Martin et al., 2022) and it is called adaptive sampling (AS). ONT AS method allows the currently being sequenced DNA fragment in a given pore to be compared instantly with provided references to determine whether to sequence the DNA further (accepted or undecided) or reject it from the pore (rejected), which increases the sequencing capacity for molecules of interest (Loose et al., 2016).

To test the capacity of host DNA depletion with AS, we used five DNA samples primarily extracted with the AllPrep kit and sequenced on Illumina platform using 150 bp PE mode (Figures 2A, 3). We re-sequenced these samples using ONT, with one group serving as a "control group" without adaptive sampling and the other one as a "AS group" with potential reduction or depletion of host DNA reads (Figure 4).

Overall, ONT adaptive sampling yielded 50% less total bases of raw data (average of ~20'915'948'394 bp in the control group and ~11'745'064'707 bp in AS group) and 3 times more reads (average of 4'700'719 reads in the control group and 15'331'338 reads in the depletion group) (Figure 4A; Supplementary Table S3). Despite

efforts to reduce it, the prevalence of host DNA reads remained relatively high in all samples. However, the length distribution of host DNA reads was restricted to approximately 500 base pairs, suggesting a discernible impact on both the mean read length and the read length N50.

Our assessment of AS efficiency by comparing the host and microbial read numbers revealed a statistically significant, albeit relatively modest depletion rate (between 0.5% and 0.7%) compared to traditional wet-lab-based approaches (Figure 4A). Due to the removal of human reads from the sequencing pool, which comprised the majority of reads in human biopsy samples, sequencing with the AS approach yielded shorter human DNA read lengths. Additionally, the ejection of DNA strands from the pore requires a recovery time, leading to a lower number of reads generated with the AS approach. Consequently, the total output of the AS approach was reduced due to both the lower number and shorter length of reads, while allowing for a substantially greater number of non-human bases to be sequenced in our compartment of interest (Supplementary Table S3; Figure 4A).

Taxonomic profiling revealed AS led to a 2–5-fold increase in the number of reads corresponding to *Escherichia coli* in analyzed samples compared to traditional ONT sequencing without AS

(Figure 4B; Supplementary Table S4). Additionally, we observed a modest enrichment of *Bacteroides fragilis* and *Bacteroides thetaiotaomicron* in sample 4AS. In a supplementary analysis, we investigated the efficacy of AS approach in generating better metagenomic assemblies from human intestinal biopsy samples. Our findings revealed that the AS enabled us to assemble more bacterial contigs with greater completeness compared to the non-AS approach (Supplementary Table S5; Figure 4C). Specifically, in sample 4AS, we were able to assemble approximately 57% (3.5 Mbp) of the *B. fragilis* genome with 4X coverage, while only 3% (166 Kbp) of its genome could be assembled from sequence generated without AS (Supplementary Table S5; Figure 4C).

Moreover, our results demonstrated that the greater genome completeness achieved with AS allowed us to recover antimicrobial resistance markers in sample 4AS, which were not detectable in the non-AS sample 4. Specifically, we were able to identify a CepA beta-lactamase and a tetracycline-resistant ribosomal protection protein in sample 4AS. In contrast, these markers were not identified in the non-AS sample 4. AS also enabled the identification of two plasmids in sample 4AS, which were not identified in sample 4 (Table 2). Overall, these results demonstrate the potential utility of the AS approach for the targeted sequencing of specific microbial taxa in complex samples containing high amounts of host DNA and highlight the potential advantages of the AS approach for generating high-quality metagenomic assemblies and identifying important biological features, such as antimicrobial resistance markers and plasmids.

4 Discussion

Metagenomic shotgun sequencing of bacterial populations and advanced downstream analysis techniques are powerful techniques to assess the impact of environmental insults or host-derived factors. However, obtaining sufficient bacterial DNA from intestinal tissues can be challenging due to the presence of high amounts of host DNA, which vastly outnumbers microbial DNA. Therefore, substantially decreasing the amount of human DNA is crucial for the successful application of metagenomic sequencing. Host DNA depletion kits are a recent development in the field aiming to enrich microbial DNA in host-associated samples such as blood, feces, urine, saliva, or biopsies. These kits and previous work have advanced a number of strategies; however, a systematic comparison of these approaches alone and in combination has not been done. To address this, we tested host DNA depletion from human intestinal biopsies using i) wet-lab approaches using commercial kits and a protocol inspired by Bruggeling et al. (2021) and ii) a software-based enrichment protocol using a nanopore sequencing platform.

Kit-based microbial DNA enrichment approaches are designed to be effective across most hosts, regardless of the type of microbe present in the sample. They rely on a series of predetermined steps that are optimized to extract the DNA of interest. In contrast, a software-based adaptive sampling method in ONT requires prior knowledge of the host. It is based on the electrical properties of DNA molecules as they pass through tiny pores. The adaptive sampling approach involves selecting specific areas of interest within the host

genome and sequencing only those areas in order to identify the areas of interest accurately. Therefore, the key difference between these two approaches is that kit-based microbial DNA enrichment approaches are more robust and can be used across most hosts, while adaptive sampling methods used in ONT require prior knowledge of the host for accurate results. However, the overall efficacy of different assays varied, and some methods yielded acceptable results with up to 28% of host DNA depletion. However, no method depleted 90% of the host DNA, which is required for highly sophisticated analyses such as bacterial genome analyses. Further, potential biases, such as the preferential enrichment of specific microbial taxa remain a concern.

In this study, we first assessed the capacity and performance of various commercially available host DNA depletion kits, such as the HostZERO Microbial DNA kit, the Molzym Ultra-Deep Microbiome Prep Kit, the NEBNext Microbiome DNA Enrichment Kit, and the QIAamp DNA Microbiome Kit for extracting DNA from intestinal biopsy samples (Figure 1). These kits employ different techniques for lysing host cells and enzymatically degrading host DNA. The NEBNext Microbiome DNA Enrichment Kit and QIAamp DNA Microbiome Kit depleted host DNA by up to ~28% (Figure 2). The NEBNext kit uses a distinct approach by selectively eliminating CpG-methylated host DNA from total DNA extracted using the AllPrep DNA/RNA Mini Kit, and our results confirm the potential utility of this approach. In a previous study with complex respiratory samples, NEBNext also showed an effective host DNA depletion (Thoendel et al., 2016). However, in previous analyses, this method showed poorer results, such as no effective host DNA reduction from saliva samples (Marotz et al., 2018), and relatively low host DNA depletion from resected arthroplasty components and sonicated fluids from prosthetic joint infections (Nelson et al., 2019).

The tested wet-lab-based enrichment methods are not without limitations and biases. Host DNA depletion can introduce a bias toward the identification of specific microorganisms. The kits are designed to remove host-associated DNA and proteins but may also remove microorganisms that are either closely associated with host cells or show DNA characteristics similar to mammalian DNA. This can lead to an underrepresentation of specific microorganisms in the final enriched sample. When analyzing shotgun metagenomic datasets for shifts in richness and taxonomy profile, we observed that all kits, except for the Molzym kit, affected certain bacterial phyla and families stronger than others. Specifically, significant changes were observed in the Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria phyla. Within the Proteobacteria phylum, the Pseudomonadales and Enterobacterales orders, as well as the Xanthomonadaceae, Burkholderiaceae, Enterobacteriaceae, and Sphingomonadaceae families, were more enriched after the extraction using HostZero, NEBNext, and QIAamp kits compared to the standard AllPrep kit (Figure 3). The reasons for the selective enrichment or decrease for the mentioned taxa are unclear. On the other hand, the microbiota of the HostZero and QIAamp kits showed a greater diversity of species compared to the standard AllPrep kit. Therefore, these kits have a higher potential to detect bacteria of lower abundance in intestinal biopsies, which might be beneficial in some situations.

A potential alternative to host DNA depletion kits is the use of Oxford Nanopore Technologies' (ONT) adaptive sampling (AS) feature,

TABLE 2 ONT adaptive sampling improves the detection of plasmids. Two plasmids were identified in sample 4 with adaptive sampling but not without AS. Cluster IDs reflect unique taxonomic identifiers for plasmids and are stable over time. *MRF*: Metapair information and *oriT*: origin of transfer.

Plasmid cluster ID	MPF type	oriT type(s)	Median size across samples (bp)	Coverage	Predicted host range	Nearest NCBI accession	Replicon type(s)	Relaxase type(s)
AA336	MPF _F	MOB _F	103819 bp	5X	Enterobacterales	LT985217	IncFIA	MOB _F
AG294	—	—	17450 bp	3X	<i>Escherichia</i>	CP019906	—	—

which has been shown to increase sequencing depth in bacterial sequences without altering microbial composition when a human reference genome is provided. In fact, recent studies have demonstrated that ONT's adaptive sampling method reliably increases overall diversity and sequencing depth in clinical metagenomic samples such as ~113-fold increase in clinical samples of respiratory tract infections (Gan et al., 2021), ~30-fold enrichment of 148 human genes associated with hereditary cancers (Kovaka et al., 2021), a ~14-fold increase of the least abundant species in a mock community (Martin et al., 2022), and a 40-fold enrichment of a ZymoBIOMICS mock metagenomic community (Payne et al., 2021). In our hands, adaptive sampling yielded a modest enrichment of bacterial sequences, nonetheless, the overall higher number of bacterial reads enabled a more complete assembly of bacterial genomes and a better identification of bacterial plasmids. However, even with adaptive sampling, complete genomic assembly of even the most abundant species has not been possible, indicating the need for further improvement.

It is worth noting, however, that ONT AS approach also changed the overall identified bacterial composition by depleting some microbial sequences. Particularly, sequencing of the genetic material of the Enterobacteriaceae family, such as *Escherichia coli*, was favoured by this method, similar to the results observed with the DNA depletion kits (Figure 4B). On the other hand, one sample by ONT AS had enriched sequences of two *Bacteroides* taxa, which had not been affected by DNA depletion kits (Figures 4B,C). These observations highlight potential limitations of adaptive sampling, and further investigations are warranted to determine the optimal sequencing approach for different types of microbiome studies. One limitation of our study was that we did not evaluate the wet-lab technique using a host DNA depletion kit and subsequently employing adaptive sampling in ONT. This might help to increase the number of bacterial reads in a tested sample; however, this will also increase the overall sequencing cost per sample. Furthermore, we did not examine the potential usage of formalin-fixed paraffin-embedded (FFPE) tissue, which could serve as a valuable resource for microbial characterization of tissue sections studied previously. However, microbial DNA extracted FFPEs might be affected by various factors, such as damage caused during the fixation and embedding process, and a potentially low quantity of microbial DNA in the sample. Generally, the yield of DNA from FFPE is lower than those obtained from fresh or frozen tissue, which suggests that the tested protocols in this study may not necessarily lead to improved microbial enrichment for FFPEs.

Host DNA depletion up to 30%–50% can be considered an acceptable range for taxonomy classification using shotgun metagenomic approaches. However, it is important to note that having over half of the reads assigned to host DNA still poses a challenge in bacterial genome assembly and identification of SNPs and structural variants. In a recent study, we collected stoma content

from the small intestine of six cured colorectal cancer patients who fasted overnight and consumed a standardized breakfast over 6–10 h (Yilmaz et al., 2022). Our results showed that microbial reads obtained from Illumina 150 pb paired-end sequencing increased significantly as time progressed after feeding. However, the sub-strains of the blooming bacteria could only be identified once microbial reads reached over 78% of the total DNA within a given sample. These findings highlight the importance of effective host DNA depletion to enhance microbial read recovery and downstream metagenomic analysis. Based on the previous and current findings, we conclude that current commercial kits have the ability to partially deplete host DNA; however, the depletion rates achieved are not sufficient for sub-strain analysis or structural analysis of the bacterial genomes, even of abundant bacterial species.

In conclusion, the present study provides valuable insights into the efficacy and limitations of host DNA depletion methods for microbiome studies in human intestinal samples. It also highlights the necessity for further development of more effective methods to optimize bacterial DNA yields. With current technologies, researchers designing an analysis pipeline involving microbial DNA enrichment steps must scrutinize several factors when selecting the most suitable approach for their research. Specifically, parameters such as the sample type, the bacterial species or pathogen(s) of interest, cost, and the level of enrichment should be considered. It is crucial to note that the starting proportion of bacterial DNA in a sample highly influences the enrichment factor. Therefore, a sample with lower microbial content will experience a higher fold enrichment than a sample with a higher initial microbial DNA content, provided that an equal amount of host DNA is removed. For instance, it is unfeasible to achieve a 500-fold enrichment if microbial DNA initially constituted only ≤1% of the total DNA (Shi et al., 2022). In these situations, it may be worth exploring the use of alternative technologies for bacterial genome analysis, such as single-cell or long-read sequencing, which may be less impacted by host DNA contamination.

Data availability statement

The datasets generated for this study are available through the Sequence Read Archive NCBI. The Nanopore datasets for this study can be found in the BioProject ID: PRJNA943380 and BioSample: SAMN33717862 (<http://www.ncbi.nlm.nih.gov/bioproject/943380>).

Ethics statement

The studies involving human participants were reviewed and approved by Bern Cantonal Ethics Commission (Ref: KEK-BE: 251/

14). The patients/participants provided their written informed consent to participate in this study.

Author contributions

DM and BY designed the experiments. PJ collected the biopsy samples. DM and JL sequenced the biopsy samples. BY and JL analyzed the data. DM, BM, and BY wrote the manuscript. All authors contributed to revisions of the manuscript, and read, and approved the submitted version.

Funding

This work was supported by SNF Ambizione Grant (PZ00P3_185880) and Novartis Foundation for Medical-Biological Research (#19A013) to BY. BY has also received funding from SNF Starting Grant TMSGI3_211300. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Acknowledgments

We also thank the University of Bern Next-Generation Sequencing (NGS) Platform at VetSuisse Faculty for sequencing and the Interfaculty Bioinformatics Unit for computational infrastructure. We thank all healthy subjects who donate the intestinal biopsies for their commitment. We also thank the staff of the University Hospital of Visceral Medicine and Surgery, for helping PJ with endoscopic sample collections. We are also grateful to Dr. Alban Ramette (Institute for Infectious Diseases, University of Bern) and Dr. Loïc Borcard (Institute for Infectious Diseases, University of Bern) for their help with sequencing our samples in Nanopore with the adaptive sampling approach. Analysis of the

Nanopore dataset was supported by the BugSeq Platform (<https://bugseq.com>) and we thank to Dr. Sam Chorlton for guiding us with the interpretation of the data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1184473/full#supplementary-material>

SUPPLEMENTARY FIGURE S1

Diversity differences based on host DNA abundance. The alpha and beta diversity differences were quantified using the Shannon index and the Aitchison distance, respectively, in relation to host DNA abundance. (A) The correlation between host DNA abundance and Shannon index was shown for all five kits used in the study with the default setting (left panel), as well as for three kits with lab-optimized protocols (right panel). Each sample is represented by a dot, and statistical significance was observed in both panels (p -value < 0.05). (B) The differences in host DNA abundance with Aitchison distance, with the left panel displaying results for all five kits used with the manufacturer's instruction and the right panel showing results for three kits with lab-optimized protocols.

References

- Andrews, S. (2015). *A quality control tool for high throughput sequence data*. [Online].
- Backhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., Nagy, A., et al. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. U. S. A.* 101, 15718–15723. doi:10.1073/pnas.040707610
- Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., et al. (2016). Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* 533, 543–546. doi:10.1038/nature17645
- Bruggeling, C. E., Garza, D. R., Achouti, S., Mes, W., Dutilh, B. E., and Boleij, A. (2021). Optimized bacterial DNA isolation method for microbiome analysis of human tissues. *Microbiologyopen* 10, e1191. doi:10.1002/mbo3.1191
- Callahan, B. J., Sankaran, K., Fukuyama, J. A., Mcmurdie, P. J., and Holmes, S. P. (2016). Bioconductor workflow for microbiome data analysis: From raw reads to community analyses. *F1000Res.* 5, 1492. doi:10.12688/f1000research.8986.2
- Chandrakumar, I., Gauthier, N. P. G., Nelson, C., Bonsall, M. B., Locher, K., Charles, M., et al. (2022). BugSplit enables genome-resolved metagenomics through highly accurate taxonomic binning of metagenomic assemblies. *Commun. Biol.* 5, 151. doi:10.1038/s42003-022-03114-4
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi:10.1093/bioinformatics/bty560
- Cheng, Y., Ling, Z., and Li, L. (2020). The intestinal microbiota and colorectal cancer. *Front. Immunol.* 11, 615056. doi:10.3389/fimmu.2020.615056
- Costantini, L., Magno, S., Albanese, D., Donati, C., Molinari, R., Filippone, A., et al. (2018). Characterization of human breast tissue microbiota from core needle biopsies through the analysis of multi hypervariable 16S-rRNA gene regions. *Sci. Rep.* 8, 16893. doi:10.1038/s41598-018-35329-z
- Cox, L. M., Yamanishi, S., Sohn, J., Alekseyenko, A. V., Leung, J. M., Cho, I., et al. (2014). Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* 158, 705–721. doi:10.1016/j.cell.2014.05.052
- De Albuquerque, G. E., Moda, B. S., Serpa, M. S., Branco, G. P., Defelicibus, A., Takenaka, I., et al. (2022). Evaluation of bacteria and fungi DNA abundance in human tissues. *Genes (Basel)* 13, 237. doi:10.3390/genes13020237
- Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., et al. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. U. S. A.* 107, 11971–11975. doi:10.1073/pnas.1002601107
- Fan, J., Huang, S., and Chorlton, S. D. (2021). BugSeq: A highly accurate cloud platform for long-read metagenomic analyses. *BMC Bioinforma.* 22, 160. doi:10.1186/s12859-021-04089-5
- Feehery, G. R., Yigit, E., Oyola, S. O., Langhorst, B. W., Schmidt, V. T., Stewart, F. J., et al. (2013). A method for selectively enriching microbial DNA from contaminating vertebrate host DNA. *PLoS One* 8, e76096. doi:10.1371/journal.pone.0076096
- Feng, Q., Liang, S., Jia, H., Stadlmayr, A., Tang, L., Lan, Z., et al. (2015). Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.* 6, 6528. doi:10.1038/ncomms7528
- Francis, O. E., Bendall, M., Manimaran, S., Hong, C., Clement, N. L., Castro-Nallar, E., et al. (2013). Pathoscope: Species identification and strain attribution with unassembled sequencing data. *Genome Res.* 23, 1721–1729. doi:10.1101/gr.150151.112

- Franzosa, E. A., Mciver, L. J., Rahnnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. doi:10.1038/s41592-018-0176-y
- Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat. Microbiol.* 4, 293–305. doi:10.1038/s41564-018-0306-4
- Gan, M., Wu, B., Yan, G., Li, G., Sun, L., Lu, G., et al. (2021). Combined nanopore adaptive sequencing and enzyme-based host depletion efficiently enriched microbial sequences and identified missing respiratory pathogens. *BMC Genomics* 22, 732. doi:10.1186/s12864-021-08023-0
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inf.* 42, 377–381. doi:10.1016/j.jbi.2008.08.010
- Hasan, M. R., Rawat, A., Tang, P., Jithesh, P. V., Thomas, E., Tan, R., et al. (2016). Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *J. Clin. Microbiol.* 54, 919–927. doi:10.1128/JCM.03050-15
- Heravi, F. S., Zakrzewski, M., Vickery, K., and Hu, H. (2020). Host DNA depletion efficiency of microbiome DNA enrichment methods in infected tissue samples. *J. Microbiol. Methods* 170, 105856. doi:10.1016/j.mimet.2020.105856
- Hilt, E. E., Mckinley, K., Pearce, M. M., Rosenfeld, A. B., Zilliox, M. J., Mueller, E. R., et al. (2014). Urine is not sterile: Use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *J. Clin. Microbiol.* 52, 871–876. doi:10.1128/JCM.02876-13
- Hooper, L. V., and Macpherson, A. J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat. Rev. Immunol.* 10, 159–169. doi:10.1038/nri2710
- Ji, X. C., Zhou, L. F., Li, C. Y., Shi, Y. J., Wu, M. L., Zhang, Y., et al. (2020). Reduction of human DNA contamination in clinical cerebrospinal fluid specimens improves the sensitivity of metagenomic next-generation sequencing. *J. Mol. Neurosci.* 70, 659–666. doi:10.1007/s12031-019-01472-z
- Korem, T., Zeevi, D., Suez, J., Weinberger, A., Avnit-Sagi, T., Pompan-Lotan, M., et al. (2015). Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science* 349, 1101–1106. doi:10.1126/science.aac4812
- Kovaka, S., Fan, Y., Ni, B., Timp, W., and Schatz, M. C. (2021). Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat. Biotechnol.* 39, 431–441. doi:10.1038/s41587-020-0731-9
- Lau, J. T., Whelan, F. J., Herath, L., Lee, C. H., Collins, S. M., Bercik, P., et al. (2016). Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Med.* 8, 72. doi:10.1186/s13073-016-0327-7
- Leite, G. G. S., Weitsman, S., Parodi, G., Celly, S., Sedighi, R., Sanchez, M., et al. (2020). Mapping the segmental microbiomes in the human small bowel in comparison with stool: A reimagine study. *Dig. Dis. Sci.* 65, 2595–2604. doi:10.1007/s10620-020-06173-x
- Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., and Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. U. S. A.* 102, 11070–11075. doi:10.1073/pnas.0504978102
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191
- Liang, Q., Chiu, J., Chen, Y., Huang, Y., Higashimori, A., Fang, J., et al. (2017). Fecal bacteria act as novel biomarkers for noninvasive diagnosis of colorectal cancer. *Clin. Cancer Res.* 23, 2061–2070. doi:10.1158/1078-0432.CCR-16-1599
- Lloyd-Price, J., Arze, C., Ananthkrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. doi:10.1038/s41586-019-1237-9
- Loman, N. J., Constantinidou, C., Chan, J. Z., Halachev, M., Sergeant, M., Penn, C. W., et al. (2012). High-throughput bacterial genome sequencing: An embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* 10, 599–606. doi:10.1038/nrmicro2850
- Loose, M., Malla, S., and Stout, M. (2016). Real-time selective sequencing using nanopore technology. *Nat. Methods* 13, 751–754. doi:10.1038/nmeth.3930
- Lynch, J. B., and Hsiao, E. Y. (2019). Microbiomes as sources of emergent host phenotypes. *Science* 365, 1405–1409. doi:10.1126/science.aay0240
- Macpherson, A. J., Yilmaz, B., Limenitakis, J. P., and Ganal-Vonarburg, S. C. (2018). IgA function in relation to the intestinal microbiota. *Annu. Rev. Immunol.* 36, 359–381. doi:10.1146/annurev-immunol-042617-053238
- Mallick, H., Rahnnavard, A., Mciver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* 17, e1009442. doi:10.1371/journal.pcbi.1009442
- Marotz, C. A., Sanders, J. G., Zuniga, C., Zaramela, L. S., Knight, R., and Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome* 6, 42. doi:10.1186/s40168-018-0426-3
- Marti, J. M. (2019). Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLoS Comput. Biol.* 15, e1006967. doi:10.1371/journal.pcbi.1006967
- Martin, R., Heilig, H. G., Zoetendal, E. G., Jimenez, E., Fernandez, L., Smidt, H., et al. (2007). Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women. *Res. Microbiol.* 158, 31–37. doi:10.1016/j.resmic.2006.11.004
- Martin, S., Heavens, D., Lan, Y., Horsfield, S., Clark, M. D., and Leggett, R. M. (2022). Nanopore adaptive sampling: A tool for enrichment of low abundance species in metagenomic samples. *Genome Biol.* 23, 11. doi:10.1186/s13059-021-02582-x
- McMurdie, P. J., and Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *Plos One* 8, e61217. doi:10.1371/journal.pone.0061217
- Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79. doi:10.1186/gb-2012-13-9-r79
- Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z., and Dominguez-Bello, M. G. (2015). The infant microbiome development: Mom matters. *Trends Mol. Med.* 21, 109–117. doi:10.1016/j.molmed.2014.12.002
- Nelson, M. T., Pope, C. E., Marsh, R. L., Wolter, D. J., Weiss, E. J., Hager, K. R., et al. (2019). Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell. Rep.* 26, 2227–2240. doi:10.1016/j.celrep.2019.01.091
- Olivares, M., Walker, A. W., Capilla, A., Benitez-Paez, A., Palau, F., Parkhill, J., et al. (2018). Gut microbiota trajectory in early life may predict development of celiac disease. *Microbiome* 6, 36. doi:10.1186/s40168-018-0415-6
- Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., and Loose, M. (2021). Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* 39, 442–450. doi:10.1038/s41587-020-00746-x
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi:10.1038/nature08821
- Ranjan, R., Rani, A., Metwally, A., Mcgee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977. doi:10.1016/j.bbrc.2015.12.083
- Robinson, S. L., Piel, J., and Sunagawa, S. (2021). A roadmap for metagenomic enzyme discovery. *Nat. Prod. Rep.* 38, 1994–2023. doi:10.1039/d1np00006c
- Rotmistrovsky, K. A., R. (2011). BMTagger: Best match tagger for removing human reads from metagenomics datasets.
- Saffarian, A., Mulet, C., Regnault, B., Amiot, A., Tran-Van-Nhieu, J., Ravel, J., et al. (2019). Crypt- and mucosa-associated core microbiotas in humans and their alteration in colon cancer patients. *mBio* 10, e01315–e01319. doi:10.1128/mBio.01315-19
- Schriefer, A. E., Cliften, P. F., Hibberd, M. C., Sawyer, C., Brown-Kennerly, V., Burcea, L., et al. (2018). A multi-amplicon 16S rRNA sequencing and analysis method for improved taxonomic profiling of bacterial communities. *J. Microbiol. Methods* 154, 6–13. doi:10.1016/j.mimet.2018.09.019
- Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* 14, e1002533. doi:10.1371/journal.pbio.1002533
- Shi, Y., Wang, G., Lau, H. C., and Yu, J. (2022). Metagenomic sequencing for microbial DNA in human samples: Emerging technological advances. *Int. J. Mol. Sci.* 23, 2181. doi:10.3390/ijms23042181
- Smith, K., Mccoy, K. D., and Macpherson, A. J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin. Immunol.* 19, 59–69. doi:10.1016/j.smim.2006.10.002
- Stecher, B., Macpherson, A. J., Hapfelmeier, S., Kremer, M., Stallmach, T., and Hardt, W. D. (2005). Comparison of *Salmonella enterica* serovar Typhimurium colitis in germfree mice and mice pretreated with streptomycin. *Infect. Immun.* 73, 3228–3241. doi:10.1128/IAI.73.6.3228-3241.2005
- Stewart, C. J., Ajami, N. J., O'brien, J. L., Hutchinson, D. S., Smith, D. P., Wong, M. C., et al. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562, 583–588. doi:10.1038/s41586-018-0617-x
- Suez, J., Zmora, N., Zilberman-Schapira, G., Mor, U., Dori-Bachash, M., Bashardes, S., et al. (2018). Post-antibiotic gut mucosal microbiome reconstitution is impaired by probiotics and improved by autologous FMT. *Cell.* 174, 1406–1423. doi:10.1016/j.cell.2018.08.047
- Thoendel, M., Jeraldo, P. R., Greenwood-Quaintance, K. E., Yao, J. Z., Chia, N., Hanssen, A. D., et al. (2016). Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J. Microbiol. Methods* 127, 141–145. doi:10.1016/j.mimet.2016.05.022
- Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R., and Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031. doi:10.1038/nature05414
- Uchimura, Y., Fuhrer, T., Li, H., Lawson, M. A., Zimmermann, M., Yilmaz, B., et al. (2018). Antibodies Set boundaries limiting microbial metabolite penetration and the resultant mammalian host response. *Immunity* 49, 545–559. doi:10.1016/j.immuni.2018.08.004

- Uritskiy, G. V., Diruggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6, 158. doi:10.1186/s40168-018-0541-1
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. doi:10.1186/s13059-019-1891-0
- Yap, M., Feehily, C., Walsh, C. J., Fenelon, M., Murphy, E. F., Mcauliffe, F. M., et al. (2020). Evaluation of methods for the reduction of contaminating host reads when performing shotgun metagenomic sequencing of the milk microbiome. *Sci. Rep.* 10, 21665. doi:10.1038/s41598-020-78773-6
- Yilmaz, B., Fuhrer, T., Morgenthaler, D., Krupka, N., Wang, D., Spari, D., et al. (2022). Plasticity of the adult human small intestinal stoma microbiota. *Cell. Host Microbe* 30, 1773–1787. e6. doi:10.1016/j.chom.2022.10.002
- Yilmaz, B., Juillerat, P., Oyas, O., Ramon, C., Bravo, F. D., Franc, Y., et al. (2019). Microbial network disturbances in relapsing refractory Crohn's disease. *Nat. Med.* 25, 323–336. doi:10.1038/s41591-018-0308-z
- Yilmaz, B., Mooser, C., Keller, I., Li, H., Zimmermann, J., Bosshard, L., et al. (2021). Long-term evolution and short-term adaptation of microbiota strains and sub-strains in mice. *Cell. Host Microbe* 29, 650–663. e9. doi:10.1016/j.chom.2021.02.001
- Yilmaz, B., Portugal, S., Tran, T. M., Gozzelino, R., Ramos, S., Gomes, J., et al. (2014). Gut microbiota elicits a protective immune response against malaria transmission. *Cell* 159, 1277–1289. doi:10.1016/j.cell.2014.10.053
- Young, V. B. (2017). The role of the microbiome in human health and disease: An introduction for clinicians. *BMJ* 356, j831. doi:10.1136/bmj.j831
- Yu, J., Feng, Q., Wong, S. H., Zhang, D., Liang, Q. Y., Qin, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 66, 70–78. doi:10.1136/gutjnl-2015-309800
- Zeevi, D., Korem, T., Godneva, A., Bar, N., Kurilshikov, A., Lotan-Pompan, M., et al. (2019). Structural variation in the gut microbiome associates with host health. *Nature* 568, 43–48. doi:10.1038/s41586-019-1065-y
- Zhou, Z., Sun, B., Yu, D., and Zhu, C. (2022). Gut microbiota: An important player in type 2 diabetes mellitus. *Front. Cell. Infect. Microbiol.* 12, 834485. doi:10.3389/fcimb.2022.834485