Check for updates

#### **OPEN ACCESS**

EDITED BY Mariusz Cycoń, Medical University of Silesia, Poland

REVIEWED BY Lauren M. Lui, Berkeley Lab (DOE), United States Jingjie Du, Cornell University, United States Jintao He, Zhejiang University, China

\*CORRESPONDENCE Ksenia Klimina ⊠ ppp843@yandex.ru

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 27 February 2025 ACCEPTED 05 May 2025 PUBLISHED 20 May 2025

#### CITATION

Strokach A, Zoruk P, Boldyreva D, Morozov M, Olekhnovich E, Veselovsky V, Babenko V, Selezneva O, Zakharevich N, Larin A, Koldman S, Koldman V, Odorskaya M, Yunes R, Pavlov V, Kudryavtseva A, Danilenko V and Klimina K (2025) Comparative evaluation of sequencing technologies and primer sets for mouse gut microbiota profiling. *Front. Microbiol.* 16:1584359. doi: 10.3389/fmicb.2025.1584359

#### COPYRIGHT

© 2025 Strokach, Zoruk, Boldyreva, Morozov, Olekhnovich, Veselovsky, Babenko, Selezneva, Zakharevich, Larin, Koldman, Koldman, Odorskava, Yunes, Pavlov, Kudryavtseva, Danilenko and Klimina. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.

## Comparative evaluation of sequencing technologies and primer sets for mouse gut microbiota profiling

Aleksandra Strokach<sup>1†</sup>, Polina Zoruk<sup>1†</sup>, Daria Boldyreva<sup>1</sup>, Maxim Morozov<sup>1</sup>, Evgenii Olekhnovich<sup>1</sup>, Vladimir Veselovsky<sup>1</sup>, Vladislav Babenko<sup>1</sup>, Oksana Selezneva<sup>1</sup>, Natalia Zakharevich<sup>1</sup>, Andrey Larin<sup>1</sup>, Severina Koldman<sup>1,2</sup>, Vail Koldman<sup>1,2</sup>, Maya Odorskaya<sup>3</sup>, Roman Yunes<sup>3</sup>, Vladislav Pavlov<sup>4</sup>, Anna Kudryavtseva<sup>4</sup>, Valeriy Danilenko<sup>3</sup> and Ksenia Klimina<sup>1,3</sup>\*

<sup>1</sup>Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia, <sup>2</sup>Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency, Moscow, Russia, <sup>3</sup>Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia, <sup>4</sup>Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

**Background:** Advancements in sequencing technologies, such as Illumina and Oxford Nanopore Technologies (ONT), have significantly improved microbiome research. However, variations in sequencing platforms, primer selection, and DNA quality may influence microbial diversity assessments, particularly in studies of gut microbiota. This study systematically evaluates these factors in mouse gut microbiota analysis, comparing 16S rRNA gene sequencing and metagenome sequencing (MS) across both platforms.

**Results:** Our findings highlight the critical influence of primer selection on 16S rRNA sequencing results, with certain primer combinations detecting unique taxa that others miss. Despite these variations in taxonomic resolution, all tested primer sets consistently revealed significant differences between experimental groups, indicating that key microbial shifts induced by bacterial cultures remain detectable regardless of primer choice. A comparative analysis of Illumina and ONT 16S rRNA sequencing revealed notable differences in microbial diversity profiling, with ONT capturing a broader range of taxa. In contrast, MS on both platforms showed a high degree of correlation, indicating that ONT sequencing errors have minimal impact on taxonomic diversity estimations. Furthermore, the type of extracted DNA (high molecular weight vs. standard DNA) had little on microbial diversity outcomes, underscoring the robustness of these sequencing technologies.

**Conclusion:** These results highlight the advantages and limitations of different sequencing strategies in microbiota research. While 16S rRNA sequencing remains a cost-effective tool for assessing bacterial diversity, MS provides superior taxonomic resolution and more precise species identification. Our study advocates for a hybrid approach that combines multiple sequencing technologies to achieve a more comprehensive and accurate representation of microbial communities.

#### KEYWORDS

gut microbiota, sequencing technologies, metagenome sequencing, Oxford Nanopore Technologies, Illumina sequencing, microbial diversity, high molecular weight DNA, primer combination

### Background

In recent decades, advancements in DNA sequencing technologies have transformed microbiome research, providing profound insights into the intricate microbial ecosystems inhabiting various environments. These breakthroughs have significantly contributed to our understanding of microbiota, particularly the gut microbiota (GM), which plays a crucial role in host health and disease.

The human gut microbiome is a diverse and dynamic community of bacteria, viruses, archaea, and microeukaryotes that play a crucial role in maintaining host health. This complex ecosystem influences numerous physiological processes, such as nutrient metabolism and immune system development (Wampach et al., 2017; Hooper and Mac Pherson, 2010; Blaut and Clavel, 2007). Additionally, the GM contributes to colonization resistance, acting as a natural barrier against pathogenic microorganisms by competing for resources and producing antimicrobial compounds (Van Der Waaij et al., 1971; Stecher and Hardt, 2008). Disruptions in the composition of the GM have been increasingly associated with a variety of diseases, including inflammatory bowel diseases, obesity, and cancer (Blaut and Clavel, 2007; Iida et al., 2013). Notably, alterations in the GM can influence cancer progression and response to therapy by modulating the tumor microenvironment, underscoring the essential role of microbiota-host interactions in health and disease (Iida et al., 2013).

Murine models are among the most widely used systems for studying GM due to their physiological and microbial similarities to humans (Lloyd-Price et al., 2016). These models have been instrumental in unraveling the complex interactions between GM and various aspects of health and disease, including metabolic disorders and immune responses, thereby deepening our understanding of the microbiome's role in these processes and paving the way for the development of targeted therapeutics strategies (Turnbaugh et al., 2007; Round and Mazmanian, 2009).

Traditionally, microbiota composition is analyzed using 16S rRNA gene sequencing, which targets specific variable regions of the gene to assess microbial diversity. This approach has been widely adopted across platforms such as Illumina, which uses short reads, and Oxford Nanopore Technologies (ONT), which enables full-length 16S sequencing through long-read capabilities (Klindworth et al., 2013; Li et al., 2021). While 16S sequencing remains a cornerstone of microbiome research, it provides limited information on functional potential and species-level resolution. MS addresses these limitations by enabling species-and strain-level identification as well as functional annotation. Both 16S and MS approaches can be performed using short read (e.g., Illumina) and long read (e.g., ONT) platforms, each with its own advantages and trade-offs (Callahan et al., 2016b; Jünemann et al., 2017).

Importantly, it is now well established that both DNA extraction protocols and the choice of 16S rRNA gene primers significantly influence MS outcomes. DNA extraction method can bias the representation of certain bacterial taxa, especially those with more resilient cell walls, such as Gram-positive organisms (Galla et al., 2024; Maropola et al., 2015; Sui et al., 2020), while factors like DNA quality, fragmentation, and contaminants can alter the results of metagenomic analysis (Demkina et al., 2023). Likewise, primer selection strongly affects taxonomic resolution and diversity estimates, as some primer sets preferentially amplify specific taxa or exclude others, and certain variable regions (e.g., V3–V4) may not allow for species-level classification (Abellan-Schneyder et al., 2021). Specific primers may also reduce amplification of unwanted host or organellar sequences (Nakano, 2018), or improve archaeal detection (Bahram et al., 2019).

Previous studies have primarily focused on comparing 16S rRNA sequencing results between Illumina and ONT, assessing taxonomic resolution and rare taxa detection while largely overlooking MS data. For example, ONT-based sequencing has demonstrated superior species-level classification and improved detection of rare taxa compared to Illumina in GM studies (Szoboszlay et al., 2023). Similarly, microbial profiling discrepancies have been reported between ONT full-length 16S rRNA sequencing and Illumina V3-V4 region sequencing in head and neck cancer tissues (Yeo et al., 2024). Other studies have examined taxonomic resolution differences between platforms in human nasal microbiota and assessed ONT's EPI2ME pipeline performance (Heikema et al., 2020). Additionally, some research has investigated the impact of different PCR conditions on sequencing outcomes. Although these studies provide valuable insights, they primarily focus on isolated factors-such as platform performance or PCR parameters-without addressing their combined effects or broader implications for microbiome research (Fujiyoshi et al., 2020).

Our study fills this gap by providing a comprehensive assessment of key methodological variables in MS. We systematically evaluate the impact of primer selection and sequencing platforms on microbial diversity while comparing MS strategies across long-read (ONT) and short-read (Illumina) technologies. Notably, our MS analysis revealed a high degree of correlation between the Illumina and ONT platforms, highlighting their complementarity in microbial composition assessment. Additionally, we examined the influence of different DNA extraction methods on microbial diversity profiling, shedding light on how extraction protocols shape sequencing outcomes. By integrating these factors, our study offers a holistic perspective on microbiome sequencing variability, informing best practices and enhancing the reliability, comparability, and complementarity of different methods. This integrated approach not only evaluates the influence of individual parameters on sequencing outcomes but also reveals their interactions, ultimately informing best practices for microbiome research.

### **Methods**

#### Experimental design

The study was conducted on 27 female C57BL/6 mice obtained from the Stolbovaya Branch of the Federal State Budgetary Institution of Science, "Scientific Center of Biomedical Technologies of the Federal Medical and Biological Agency." All animals were certified as healthy and possessed veterinary certificates attesting to their health status. Prior to the experiment, a 14-day quarantine period was observed for the animals. Subsequently, they were randomly allocated into three groups: the control group (designated as "Control"), mice administered with lactobacilli ("Lacto" group), and mice administered with bifidobacteria ("Bifido" group). The mice were housed in standard cages (n = 9 per cage) maintained at a temperature range of +20–23°C and a humidity level of 60–65%, under natural light conditions with forced ventilation. Sterilized wood shavings served as bedding material. Throughout the duration of the experiment, the mice had ad libitum access to both water and certified briquetted feed.

During the experiment, each mouse was intragastrically injected with 0.3 mL (10^8 CFUs/ml for *Lacticaseibacillus rhamnosus* K32 and 10^7 CFUs/ml *Bifidobacterium adolescentis* 150) of culture daily for 5 days (Figure 1A). The control group was injected with 0.3 mL of phosphate-buffered saline (PBS). Disposable sterile probes were used for injection.

Fecal samples from mice were collected in 1.5 mL Eppendorf tubes in the morning, following bedding replacement. Feces were taken at days 0, 5, 8, 12, 15, 19, 21, 25, 28 and were stored at  $-80^{\circ}$ C until use. All collected time points were used for the comparison of sequencing platforms and primer combinations (Figure 1B).

#### Bacterial strains and growing conditions

The strains *L. rhamnosus* K32 (GenBank accession number JNNV00000000) and *B. adolescentis* 150 (LBHQ00000000) were cultivated for 18 h under anaerobic conditions (10% CO<sub>2</sub> atmosphere, Anaerobic System Mark II, HiMedia, India) at +37°C in Man-Rogosa–Sharpe (MRS) medium (HiMedia, India). For *B. adolescentis* 

150, the medium was additionally supplemented with 0.05% cysteine (Dyachkova et al., 2015).

### Lyophilization of cultures

Bacterial cells in the stationary growth phase were harvested and washed with sterile PBS and were then dissolved to a solution composed 1% gelatin and 10% sucrose. The resulting suspension was dispensed into ampicillin vials in 5 mL aliquots. This mixture was incubated for 24 h at  $-20^{\circ}$ C, followed by drying in a 2.5-L Labconco freeze dryer (Labconco, USA) under a pressure of 0.42 mBar and a temperature of  $-52^{\circ}$ C for 48 h. The resulting lyophilized products were stored in vials at +4°C. The lyophilized strains *L. rhamnosus* K32 and *B. adolescentis* 150 were then dissolved in 5 mL PBS before injection. The CFUs/ml of *L. rhamnosus* K32 was  $6 \times 10^{9}$  before lyophilization and  $3.2 \times 10^{8}$  after, while for *B. adolescentis* 150, the CFUs/ml counts were  $5 \times 10^{8}$  before lyophilization and  $1.08 \times 10^{7}$  after. The difference in CFU concentrations between *L. rhamnosus* K32 and *B. adolescentis* 150 is due to inherent strain-specific growth properties and differences in viability following lyophilization under standardized conditions.

### DNA extraction for 16S rRNA analysis

Nucleic acids were extracted using the PureLink<sup>TM</sup> Microbiome DNA Purification Kit (Thermo Fisher Scientific, USA). To each of the six stool samples collected from mice, 400  $\mu$ L of PBS was added. Subsequently, the samples were transferred to bead tubes from the kit for homogenization using MagNA Lyser (Roche, Switzerland). Following homogenization, the samples were centrifuged for 1 min at 7,000 g, after which the supernatant was



transferred to a new tube and subjected to extraction using the manufacturer's protocol. The resulting DNA was then quantified using the Qubit 4 fluorometer with the Quant-iT dsDNA BR Assay Kit (Thermo Fisher Scientific, USA).

# Extraction and purification of high molecular weight DNA

Nucleic acids were extracted using the Wizard Genomic DNA Purification Kit (Promega, USA) with protocol modifications. Six stool samples were transferred to 15 mL tubes, and 2 mL of 50 mM EDTA along with 500 µL of 20 mg/mL Lysozyme (Sigma-Aldrich, USA) were added to each sample. The samples were incubated at a thermal shaker (Allsheng, China) at  $37^\circ C$   $65\,g$  for  $1.5\,h$  with intermittent pipetting for homogenization. After incubation, the samples were centrifuged at 3,000 g for 2 min using a Centrifuge 5,804 R (Eppendorf, Germany), and the supernatant was transferred to a new tube. The remaining pellet was resuspended in 1 mL of Nuclei Lysis Buffer, homogenized by pipetting, and incubated in a thermal shaker at 56°C and 300 rpm for 1.5 h. After incubation 400 µL of Protein Precipitation Solution was added and samples were vortexed briefly before being incubated on ice for 10 min. The samples were then centrifuged at 3,700 g for 10 min, and the resulting supernatant was transferred to a new 15 mL tube containing 1.4 mL of isopropanol. The samples were incubated at  $-20^\circ\text{C}$  for 1 h, followed by centrifugation at 3,700 g for 20 min. The supernatant was discarded, and the DNA pellet was washed twice with 80% ethanol, air-dried for 10–15 min, and resuspended in 200  $\mu L$  of Low TE buffer.

For additional purification, an equal volume of 2% CTAB solution (100 mM Tris pH 8.0, 20 mM EDTA, 1.4 M NaCl, pre-warmed to 55-65°C) was added to the extracted DNA. The mixture was gently inverted to mix and incubated at 65°C for 10 min. An equal volume of chloroform was then added, vortexed briefly (~30 s), and centrifuged at 3,500 g for 5 min to separate the phases. The upper aqueous layer was transferred to a fresh tube, and 1.1 volumes of 1% CTAB solution (50 mM Tris pH 8.0, 10 mM EDTA, pre-warmed to 55-65°C) was added, mixed thoroughly, and centrifuged at 3,500 g for 5 min, resulting in visible DNA precipitation. The pellet was dissolved in 0.5 mL of 10 mM Tris pH 8.0, 0.1 mM EDTA, and 1 M NaCl by heating at 65°C for 30 min, followed by the addition of 0.6 volumes of isopropanol. The solution was gently inverted multiple times and centrifuged in a microfuge for 8 min. The DNA pellet was washed twice with 0.5 mL of 70% ethanol, air-dried, and resuspended in 100 µL of Low TE buffer, followed by incubation at 65°C for 30 min to ensure complete solubilization. DNA concentration and purity were assessed using a Qubit 4 Fluorometer and a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA).

# 16S rRNA gene sequencing on the MinION platform

The extracted DNA (1–5 ng) was amplified using the forward primers: 27F (AGAGTTTGATYMTGGCTCAG) (Satokari et al., 2002), bif27F (GGGTTCGATTCTGGCTCAG) (Park et al., 2021),

and 8F (AGAGTTTGATCCTGGCTCAG) (Turner et al., 1999), in various combinations, with 1492R (GGTTACCTTGTTAYGACTT) (Turner et al., 1999) used as the reverse primer. PCR amplification was performed using the Tersus Plus PCR kit (Eurogen, Russia) in a total reaction volume of 25  $\mu$ L. Amplification was performed with the following PCR conditions: initial denaturation at 95°C for 2 min, (95°C for 1 min, 60°C for 1 min, and 72°C for 3 min), 27 cycles, followed by a final extension at 72°C for 2 min and 4°C – cooling (Fujiyoshi et al., 2020). The quality of the amplicons was checked by electrophoresis in 1.5% agarose gel. The final amplicons were purified using KAPA HyperPure Beads (Roche, Switzerland) according to the manufacturer's protocol.

Libraries were prepared according to the manufacturer's protocol (Ligation sequencing amplicons) with modification. The amplicons were processed with NEBNext® Ultra™ II End Repair/ dA-Tailing Module (NEB, USA). Barcodes [Native Barcoding Kit 96 (SQK-NBD109.96) were ligated with Blunt/TA Ligase Master Mix (NEB)]. Barcoded libraries were purified using KAPA Pure Beads (Roche, Switzerland). Library concentrations were measured using the Quant-iT dsDNA Assay Kit, High Sensitivity (Thermo Fisher Scientific, USA) and samples were mixed equimolarity. Final adapter Adapter Mix II Expansion (Oxford Nanopore Technologies, UK) was ligated to the pooled library using the NEBNext Quick Ligation Module (NEB). The prepared DNA library (12 µL) was mixed with 37.5 µL of Sequencing Buffer, 25.5 µL of Loading Beads, loaded onto the R9.4.1 flow cell (FLO-MIN106; Oxford Nanopore Technologies), and sequenced on the MinION Mk1B. MINKNOW software ver. 22.12.7 (Oxford Nanopore Technologies) was used for data acquisition. Reads were basecalled using dorado v.7.6.7 using default parameters [high accuracy (HAC) model, minimum quality value  $\geq 7$ ].

### 16S rRNA gene sequencing and analysis on the Illumina platform

For the amplification of extracted DNA (1-5 ng), standard 16S rRNA gene primers targeting the V3-V4 region and incorporating 5'-Illumina adapter sequences (16S Amplicon PCR Forward Primer: 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CCTACGGGNGGCWGCAG; 16S Amplicon PCR Reverse Primer: 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGAC TACHVGGGTATCTAATCC) were used. These primers were obtained from Evrogen (Russia). PCR amplification was performed with the Tersus Plus PCR kit (Evrogen, Russia) in a total volume of 25 µL. Library preparation and sequencing on the Illumina platform were performed as described in our previous work (Larin et al., 2024). DNA libraries were sequenced on a MiSeq instrument (Illumina, USA) using the 500-cycle MiSeq reagent kit v2 (Illumina, USA). The results of 16S rRNA gene sequencing were processed and analyzed using a standardized workflow. Fastp v0.23.4 was used for trimming low-quality and filtering of technical sequences (Chen et al., 2018). Processed sequences were analyzed using the DADA2 pipeline v1.26.0 (Callahan et al., 2016a) with taxonomic classification based on the SILVA database v138 (Quast et al., 2013). Differences between sequencing technologies were assessed using LefSe (Segata et al., 2011) from microbiomeMarker package (v1.10.0).<sup>1</sup>

# Bioinformatics analysis of 16S rRNA sequencing data from Oxford Nanopore

Raw sequencing reads were processed using Porechop v 0.2.4<sup>2</sup> to remove adapters, with default parameter settings. Quality filtering was performed using Chopper v 0.6.0 (De Coster and Rademakers, 2023), applying a Phred score threshold of 10 and selecting sequences within a length range of 100 to 1,800 bp. Taxonomic classification was conducted using the Emu pipeline v 3.4.5 (Curry et al., 2022), while NanoStat v 1.6.0 (De Coster and Rademakers, 2023) was employed to generate read quality statistics. The processed data were imported into RStudio (version 2023.12.0 + 369, R 4.3.2) for further analysis using the MicrobiotaProcess package v 1.16.1 (Xu et al., 2023). Alpha diversity was assessed using Shannon indices, with statistical comparisons performed via the Wilcoxon rank-sum test. Statistical significance between experimental groups was determined using distance-based permutational multivariate analysis of variance (mp\_ adonis function), with a threshold of p < 0.05 after 10,000 permutations. Differential abundance analysis was conducted using the mp\_diff\_analysis function, with visualizations generated via mp\_ plot\_diff\_res, and mp\_plot\_diff\_cladogram (Segata et al., 2011). The ComplexUpset package v1.3.3 (Lex et al., 2014) was used to generate an Upset plot, while additional visualizations were created using ggplot2 v3.5.1 and gplots v3.2.0 (Wickham, 2016). The raw data have been deposited in the NCBI GenBank database: BioProject accession PRJNA1069621 (the 16S rRNA sequences of the mice fecal microbiota).

# Comparative analysis of 16S rRNA gene sequencing using Illumina and ONT

Illumina V3-V4 16S rRNA sequencing data were processed using the DADA2 pipeline (v1.26.0) (Callahan et al., 2016a), with taxonomic classification performed against the SILVA database 138 (Quast et al., 2013). For ONT full-length 16S rRNA sequencing data, the Emu pipeline v3.4.5 was used with the same reference database to ensure consistency in taxonomic assignment (Curry et al., 2022).

# Library preparation and sequencing for Illumina

100 ng of the extracted DNA was used for library preparation using the KAPA HyperPlus Kit (Roche, Switzerland) according to the manufacturer's protocol. The library underwent a final cleanup using the KAPA HyperPure Beads (Roche, Switzerland) after which the library size distribution and quality were assessed using a high sensitivity DNA chip (Agilent Technologies). Libraries were subsequently quantified by Quant-iT DNA Assay Kit, High Sensitivity (Thermo Fisher Scientific). The DNA libraries underwent sequencing using the HiSeq 2,500 platform (Illumina, USA), in accordance with the manufacturer's recommendations. For this purpose, we employed the following reagent kits: HiSeq Rapid PE Cluster Kit v2, HiSeq Rapid SBS Kit v2 (200 cycles), and HiSeq Rapid PE FlowCell v2. Additionally, a 2% PhiX spike-in control was included in the process. The raw data have been deposited in the NCBI GenBank database: BioProject accession PRJNA1070000 (metagenomic sequences of the fecal microbiota of mice).

## Metagenomic sequencing on PromethION platform

DNA (1  $\mu$ g) was used to prepare libraries for ONT according to the manufacturer's protocol. The long reads were generated with PromethION sequencing (Oxford Nanopore Technologies, UK). The sequencing libraries were prepared using the ligation sequencing kit SQK-LSK109, native barcoding expansion kit EXP-NBD196 and run in a R9.4.1 (FLO-PRO002) flow cell. Reads were basecalled using Guppy v6.5.7 using default parameters (high accuracy (HAC) model, minimum quality value > 7).

### Bioinformatics analysis of Illumina and ONT metagenomic sequencing data

Oxford Nanopore reads were pre-processed in a similar way as 16S data. Illumina metagenomic sequencing data was processed as follows. Quality control of Illumina reads was performed using FastQC.<sup>3</sup> Fastp v0.23.4 tool was used for removing adapters and low-quality sequences from raw data (Chen et al., 2018). Removing host reads from trimmed data was performed using HiSAT2 v2.2.1 aligner (Kim et al., 2019) and mice genome version GRCm39.<sup>4</sup> For comparison of different sequencing technologies, Kraken2 v2.1.3 tool with default bacterial database (assembled in November 2022) was used for taxonomic annotation of 16S and MS Illumina and Oxford Nanopore pre-filtered reads (Lu et al., 2022). The resulting taxonomic tables were filtered by summary relative abundance > 0.01%.

### Results

# Assessment of microbial consistency in mouse gut microbiota using diverse 16S rRNA primer pairs

To assess the impact of primer selection on metagenomic analysis, we tested five distinct 16S rRNA primer sets: (1) 27F - 1492R, for broad bacterial profiling; (2) 8F - 1492R, targeting bacterial 16S rRNA genes without degenerate positions; (3) bif27F - 1492R, designed to amplify *Bifidobacterium* species; (4) 27F + bif27F - 1492R and (5)

<sup>1</sup> https://github.com/yiluheihei/microbiomeMarker

<sup>2</sup> https://github.com/rrwick/Porechop

<sup>3</sup> https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

<sup>4</sup> https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\_000001635.27

8F + bif27F - 1492R incorporating a *Bifidobacterium*-specific primer to enhance detection of this genus alongside general bacterial populations. Detection of *Bifidobacterium* species in metagenomic samples can be challenging due to mismatches between universal primers and the 16S rRNA gene sequences of this genus. For instance, the commonly used 27F primer has a known mismatch in the binding region—specifically, a G-to-A substitution frequently found in *Bifidobacterium* (Figure 2).

DNA extracted from the fecal samples of 27 mice underwent fulllength 16S rRNA sequencing using the ONT platform, resulting in a total of 2,136,829 reads. After filtering, 1,944,663 high-quality reads were retained, with sequencing distribution details provided in Supplementary Figure S1 and Supplementary Table S1. Despite variations in primer pairs, statistical analysis of alpha diversity confirmed consistent microbial composition across groups (Figure 3A), suggesting minimal impact of primer choice on overall GM diversity in mice. However, beta diversity analysis via PCoA revealed distinct grouping patterns depending on the primer combination. The '27F - 1492R', '27F + bif27-1492R', '8F - 1492R', and '8F + bif27F - 1492R' sets cluster closely together, indicating similar microbial detection profiles (Figure 3B). In contrast, the 'bif27F -1492R' combination formed a distinct cluster, likely dues to its preferential amplification of specific taxa that are less efficiently targeted by universal primers (Supplementary Figure S2). While this primer set did not detect the highest number of unique species, its clustering pattern suggests it may complement general primers by capturing specific groups within the GM.

An UpSet plot analysis illustrated the distribution of unique and shared bacterial species across different primer combinations (Figure 3C). Combining all five primers detected 174 common species, accounting for 56% of all identified species, indicating that the multi-primer approach provides broad coverage. Examining individual primer sets, '27F - 1492R' and 'bif27F - 1492R' detect 53 (18%) and 19 (8%) species, respectively. However, when the 'bif27F - 1492R' primer is removed from the analysis, the intersection drops to only eight species (3%). This reduction in the common intersection

suggests that the other primers detect largely distinct subsets of the bacterial community with relatively little overlap among themselves. In other words, its removal exposes the limited shared species among the remaining primer sets, emphasizing that each primer tends to capture a different facet of the bacterial diversity. Due to the distinct sequence profiles generated by 'bif27F - 1492R' primers and its divergence from other primer sets, it was excluded from further analysis to maintain dataset consistency.

Since our study included three distinct experimental groups (see Materials and Methods), the use of four selected primer combinations allowed us to observe clear group-specific differences in microbial composition. Statistical analysis using the Shannon index revealed significant differences between the control group and the groups receiving lyophilized bacterial cultures, with an increase in alpha diversity observed in the experimental groups. This suggests a potential beneficial effect on the intestinal microbiota (Figure 4A). Beta diversity analysis further demonstrated that the microbiota structure significantly differed between experimental groups (PERMANOVA adj.p < 0.0001, 10,000 permutations, Bray–Curtis dissimilarity metric) (Figure 4B).

# Comparative evaluation of 16S rRNA gene sequencing across Illumina and ONT

To assess differences in microbial community profiling between Illumina and ONT sequencing platforms, we compared their taxonomic diversity outputs. Despite a similar read count (Figure 5A), we observed considerable variation in the Shannon index, which measures genus diversity within the community (Figure 5B). Illumina sequencing identified 73 genera, whereas ONT sequencing identified 86. The Venn diagram (Figure 5C) shows that 43 genera (37%) were detected by both platforms. Additionally, our findings show that Illumina sequencing predominantly identifies *Bacteroides* and related bacteria, whereas ONT sequencing captures a broader range of *Firmicutes* (Supplementary Figure S3).





# Comparison of sequencing technologies for mouse microbiota analysis

To compare different sequencing approaches for mouse microbiota analysis, we utilized both high-molecular-weight (HM) DNA and DNA extracted using standard methods (Figure 1B). MS was performed on both Illumina and ONT platforms, with results labeled as Illumina\_WGS and ONT\_WGS, respectively. Additionally, the V3-V4 region of the 16S rRNA gene was sequenced using the Illumina MiSeq platform for both standard and HM DNA (Illumina\_16S and Illumina\_16S\_HM, respectively). Full-length 16S rRNA gene sequencing was conducted on the ONT platform for both DNA types (ONT\_16S and ONT\_16S\_HM, respectively). To ensure consistency, bacterial diversity assessments were performed using libraries prepared from the same extracted DNA across all sequencing methods. The number of reads obtained from each sequencing approach is summarized in Figure 6A. The average read lengths observed for each sequencing approach were as follows:



#### FIGURE 4

Microbial community analysis of experimental groups. (A) Raincloud plot of the alpha diversity index. The x-axis represents the experimental groups (Control, Bifido, and Lacto), while the y-axis represents the alpha diversity index (Shannon). Colors indicate the different groups. (B) Bray-Curtis distance-based PCoA plot for each group. Each point represents a single sample, with colors corresponding to group names. The first and second components are shown.



Illumina\_WGS – 197.3 bp, ONT\_WGS – 5332.8 bp, Illumina\_16S – 501.5 bp, Illumina\_16S\_HM – 501.6 bp, ONT\_16S – 1445.5 bp, and ONT\_16S\_HM – 1470.7 bp.

MS using both Illumina and ONT platforms (ONT\_WGS, Illumina\_WGS) yielded the highest read counts. Conversely, full-length 16S rRNA gene libraries derived from HM DNA sequenced on the ONT platform (ONT\_16S\_HM) had the lowest read counts. Despite this, bacterial diversity remained relatively high, suggesting that ONT sequencing effectively captures bacterial diversity even with fewer reads (Figure 6B). However, when sequencing the 16S rRNA gene from more fragmented DNA (ONT\_16S), the results obtained with ONT tend to be less comprehensive compared to those achieved with Illumina.

Next, we analyzed the similarity in metagenomic composition among samples sequenced using different technologies (Figure 7). Clustering analysis based on sequencing platform revealed a statistically significant correlation between taxonomic class detection and sequencing technology (Figure 7A). Across all platforms, 68 species were identified, but some were unique to specific sequencing methods (Figure 7B). For example, bacteria of the *Bacilli* class are more effectively identified using 16S rRNA methods, whereas certain



*Bacteroidetes* were more effectively identified via MS (Figure 7C). A higher number of *Bacilli* genera were observed by 16S rRNA gene analysis (full-length and V3–V4) compared to MS. Furthermore, results obtained using the same sequencing methods for both HM and standard DNA exhibited strong similarity.

In our study, we performed pairwise comparisons of microbial composition derived from different sequencing technologies, considering the quality of extracted DNA (Figure 8). Our results showed a strong correlation within each sequencing technology, though variability was influenced by the initial DNA quality. A detailed comparison of 16S rRNA gene sequencing data from illumina and ONT platforms revealed only a moderate correlation between the two (Rho = 0.49 and Rho = 0.48, respectively). In contrast, MS results from Illumina and ONT exhibited a high degree of similarity (Rho = 0.86), suggesting that combining different sequencing approaches may enhance the detection and characterization of bacterial diversity in biological samples.

### Discussion

The comparative analysis of five primer combinations revealed a significant impact of primer choice on the 16S rRNA profiling of mouse GM using ONT sequencing. The PCR conditions used in this study were based on a previously optimized protocol (Fujiyoshi et al., 2020). While all primer combinations exhibited similar alpha diversity, the 'bif27F - 1492R' combination notably identified a greater number of unique species. This variability may stem from variations in the GC content of target species, potentially introducing biases in sequencing outcomes (Park et al., 2021). The distinct results obtained with the 'bif27F - 1492R' primers, compared to others, suggest a limitation in fully capturing bacterial diversity within the samples. Nonetheless, these primers are particularly effective at detecting specific bacteria of interest, such as *Catabacter hongkongensis* — a Gram-positive anaerobic coccobacillus associated with gastrointestinal diseases and potential acute intestinal conditions (Lau et al., 2012; Lau et al., 2007;

Valentin et al., 2022). Furthermore, the identification of *Acetatifactor muris* with these primers underscores its potential role in affecting obesity or inflammation through changes in microbiota composition, highlighting the clinical significance of precise microbial profiling in therapeutic settings (Lee et al., 2019). However, while primer selection influences the detection of specific taxa, our results indicate that any chosen primer pair still provides a representative view of the experimental groups. Both alpha and beta diversity analyses revealed significant differences in microbiota composition between groups, suggesting that the key microbial shifts induced by bacterial cultures remain detectable regardless of the primer set used.

Further analysis of 16S rRNA gene sequencing data from both Illumina and ONT platforms revealed significant differences in sensitivity and specificity, particularly in microbiota composition analysis. These differences predominantly stem from the distinct read lengths utilized by each platform. Illumina sequencing targets the V3-V4 (~460 bp) region of the 16S rRNA gene, constrained by its read length limitations, whereas ONT has the capability to sequence the full-length gene (~1,480 bp). This capacity to capture full-length 16S rRNA genes endows ONT with the potential to furnish a more nuanced and comprehensive bacterial profile through sequence alignment with databases, thereby offering a broader and potentially more precise portrayal of bacterial communities within the same sample. Although ONT has a higher per-read error rate, this has limited impact on diversity estimates, as relative abundance patterns - rather than single-read accuracy - play a more critical role in alpha and beta diversity analyses. Moreover, recent improvements in ONT chemistry (e.g., Q20 + with R10.4.1 flow cells) have significantly enhanced sequencing accuracy (Yoon et al., 2017), making ONT increasingly suitable for species-level classification, detection of rare taxa, and richness estimation (Szoboszlay et al., 2023).

A comparative study investigating 16S rRNA gene sequencing of human nasal microbiota across ONT and Illumina platforms found that diversity profiles at the genus level exhibited strong similarities between both platforms. However, ONT yielded significantly fewer *Corynebacterium* compared to Illumina, potentially due to primer



mismatches during nanopore sequencing (Heikema et al., 2020). Similarily, Teahyen Cha et al. (2023) examined infant GM that taxonomic profiles identified by ONT closely paralleled those detected by Illumina at the genus level, further supporting ONT's reliability in identifying bacterial genus identification.

Analyzing sequencing data from 16S rRNA and MS across both Illumina and ONT platforms underscores the value of integrating diverse sequencing methodologies to enhance the resolution of intestinal microbiota composition. In metagenomic analysis, the generation of longer contigs plays a crucial role in achieving highquality results, which are foundational for various downstream applications. These applications include but are not limited to taxonomic assignments (Patil et al., 2011; Ciuffreda et al., 2021), gene annotation, operon identification (frequently surpassing 10 kb in length), and detecting structural variations (Ye et al., 2022). Notably, initial DNA quality appears to have minimal impact on sequencing outcomes. While MS provided a more detailed view of microbial diversity than 16S rRNA gene sequencing (Brumfield et al., 2020; Lewis et al., 2021), its effectiveness in characterizing low-abundance microbial populations is limited. MS often prioritizes highly represented species, potentially overlooking less prevalent taxa. Although increasing metagenomic read coverage could mitigate this issue, it remains economically impractical. Conversely, 16S amplicon sequencing is more resource-efficient, requiring fewer reads while capturing the "long tail" of relative abundance distributions. However, despite its advantages, 16S rRNA gene sequencing may suffer from reduced taxonomic resolution and primer-induced biases. Our study advocates for a hybrid approach, combining multiple sequencing techniques to achieve a more comprehensive understanding of microbial community diversity, acknowledging that taxonomic identifications can vary depending on sequencing technology. While recognizing potential limitations, such as misclassification with tools like Kraken2 (Lu et al., 2022), existing literature supports the utility of this method for taxonomic annotation across Illumina and ONT platforms using 16S rRNA gene sequencing and MS data (Curry et al., 2022; Lu and Salzberg, 2020; Matsuo et al., 2021).



### Conclusion

In this study, we conducted a comprehensive comparison of Illumina and ONT sequencing platforms to evaluate their performance in profiling mouse GM using 16S rRNA gene sequencing and metagenomic approaches. Our results demonstrate that the choice of primer pairs for 16S rRNA gene sequencing on the ONT platform does not significantly affect microbial diversity, though specific primers can identify unique taxa. Comparative analysis between ONT and Illumina platforms revealed differences in microbial diversity profiling, which are largely attributed to the difference in amplicon length rather than the sequencing platform itself, as ONT's ability to sequence the full-length 16S rRNA gene provides higher taxonomic resolution compared to the partial V3-V4 region targeted by Illumina. Additionally, MS showed a high degree of correlation between the two platforms, emphasizing their complementarity in assessing microbial composition. Importantly, the quality of DNA (high molecular weight versus standard DNA) did not substantially impact microbial diversity outcomes, highlighting the robustness of these sequencing technologies.

### Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

### Ethics statement

The animal study was approved by Burnasyan Federal Medical Biophysical Center of Federal Medical Biological Agency. The study was conducted in accordance with the local legislation and institutional requirements.

### Author contributions

AS: Formal analysis, Methodology, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. PZ: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. DB: Formal analysis, Investigation, Methodology, Validation, Writing original draft, Writing - review & editing. MM: Formal analysis, Methodology, Software, Validation, Visualization, Writing - review & editing. EO: Methodology, Software, Supervision, Validation, Visualization, Writing - original draft. VV: Formal analysis, Investigation, Methodology, Visualization, Writing - review & editing. VB: Investigation, Methodology, Validation, Writing - review & editing. OS: Methodology, Software, Visualization, Writing - review & editing. NZ: Formal analysis, Investigation, Methodology, Software, Writing original draft, Writing - review & editing. AL: Methodology, Validation, Visualization, Writing - review & editing. SK: Conceptualization, Data curation, Investigation, Visualization, Writing - original draft, Writing review & editing. VK: Conceptualization, Data curation, Investigation, Visualization, Writing – original draft, Writing – review & editing. MO: Methodology, Visualization, Writing – review & editing. RY: Methodology, Validation, Writing – review & editing. VP: Methodology, Writing – review & editing. AK: Methodology, Resources, Writing – review & editing. VD: Methodology, Resources, Writing – review & editing. KK: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing.

### Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The financial support for this study was provided by the Russian Science Foundation under grant number 23-75-10125, which can be accessed at https://rscf.ru/project/23-75-10125/.

### Acknowledgments

This work was performed using the core facilities of the Lopukhin FRCC PCM "Genomics, proteomics, metabolomics" (http://rcpcm. org/?p=2806). Additionally, our gratitude extends to Roman Yunes for his assistance in editing the English text.

### **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.

### References

Abellan-Schneyder, I., Matchado, M. S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., et al. (2021). Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *mSphere* 6:e01202-20. doi: 10.1128/MSPHERE.01202-20

Bahram, M., Anslan, S., Hildebrand, F., Bork, P., and Tedersoo, L. (2019). Newly designed 16S rRNA metabarcoding primers amplify diverse and novel archaeal taxa from the environment. *Environ. Microbiol. Rep.* 11, 487–494. doi: 10.1111/1758-2229.12684

Blaut, M., and Clavel, T. (2007). Metabolic diversity of the intestinal microbiota: implications for health and disease 1. J. Nutr. 137, 751S–755S. doi: 10.1093/JN/137.3.751S

Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L., and Leddy, M. B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One* 15:e0228899. doi: 10.1371/JOURNAL.PONE.0228899

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016a). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., and Holmes, S. P. (2016b). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research* 5:1492. doi: 10.12688/f1000research.8986.2

Cha, T., Kim, H. H., Keum, J., Kwak, M. J., Park, J. Y., Hoh, J. K., et al. (2023). Gut microbiome profiling of neonates using Nanopore min ION and Illumina MiSeq sequencing. *Front. Microbiol.* 14:1148466. doi: 10.3389/FMICB.2023.1148466

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an ultra-fast allin-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/BIOINFORMATICS/BTY560

Ciuffreda, L., Rodríguez-Pérez, H., and Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Comput. Struct. Biotechnol. J.* 19, 1497–1511. doi: 10.1016/J.CSBJ.2021.02.020

### **Generative AI statement**

The authors declare that no Gen AI was used in the creation of this manuscript.

### 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/fmicb.2025.1584359/ full#supplementary-material

#### SUPPLEMENTARY FIGURE S1

Number of reads for different primer pairs.

#### SUPPLEMENTARY FIGURE S2

Bacterial species uniquely identified by the primer combination "bif27F - 1492R" compared to other primers using LefSe.

#### SUPPLEMENTARY FIGURE S3

Differential taxonomic abundance analysis between 16S Illumina and ONT sequencing using LefSe. Red colors represent taxa identified by Illumina 16S rRNA gene sequencing, while blue colors denote taxa identified by ONT sequencing.

**SUPPLEMENTARY TABLE S1** Reads information.

Curry, K. D., Wang, Q., Nute, M. G., Tyshaieva, A., Reeves, E., Soriano, S., et al. (2022). Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* 19, 845–853. doi: 10.1038/s41592-022-01520-4

De Coster, W., and Rademakers, R. (2023). Nano pack 2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39:btad311. doi: 10.1093/BIOINFORMATICS/BTAD311

Demkina, A., Slonova, D., Mamontov, V., Konovalova, O., Yurikova, D., Rogozhin, V., et al. (2023). Benchmarking DNA isolation methods for marine metagenomics. *Sci. Rep.* 13, 1–18. doi: 10.1038/s41598-023-48804-z

Dyachkova, M. S., Klimina, K. M., Kovtun, A. S., Zakharevich, N. V., Nezametdinova, V. Z., Averina, O. V., et al. (2015). Draft genome sequences of *Bifdobacterium angulatum* GT102 and *Bifdobacterium adolescentis* 150: focusing on the genes potentially involved in the gut-brain Axis. *Genome Announc.* 3, 709–724. doi: 10.1128/GENOMEA.00709-15

Fujiyoshi, S., Muto-Fujita, A., and Maruyama, F. (2020). Evaluation of PCR conditions for characterizing bacterial communities with full-length 16S rRNA genes using a portable nanopore sequencer. *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-69450-9

Galla, G., Praeg, N., Rzehak, T., Sprecher, E., Colla, F., Seeber, J., et al. (2024). Comparison of DNA extraction methods on different sample matrices within the same terrestrial ecosystem. *Sci. Rep.* 14, 1–14. doi: 10.1038/s41598-024-59086-4

Heikema, A. P., Horst-Kreft, D., Boers, S. A., Jansen, R., Hiltemann, S. D., de Koning, W., et al. (2020). Comparison of Illumina versus Nanopore 16S rRNA gene sequencing of the human nasal microbiota. *Genes* 11:1105. doi: 10.3390/GENES11091105

Hooper, L. V., and Mac Pherson, A. J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat. Rev. Immunol.* 10, 159–169. doi: 10.1038/nri2710

Iida, N., Dzutsev, A., Stewart, C. A., Smith, L., Bouladoux, N., Weingarten, R. A., et al. (2013). Commensal bacteria control cancer response to therapy by modulating the tumor microenvironment. *Science* 342, 967–970. doi: 10.1126/SCIENCE.1240527

Jünemann, S., Kleinbölting, N., Jaenicke, S., Henke, C., Hassa, J., Nelkner, J., et al. (2017). Bioinformatics for NGS-based metagenomics and the application to biogas research. J. Biotechnol. 261, 10–23. doi: 10.1016/J.JBIOTEC.2017.08.012

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/S41587-019-0201-4

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and nextgeneration sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/NAR/GKS808

Larin, A. K., Klimina, K. M., Veselovsky, V. A., Olekhnovich, E. I., Morozov, M. D., Boldyreva, D. I., et al. (2024). An improved and extended dual-index multiplexed 16S rRNA sequencing for the Illumina HiSeq and MiSeq platform. *BMC Genom. Data* 25, 1–7. doi: 10.1186/S12863-024-01192-3

Lau, S. K. P., Fan, R. Y. Y., Lo, H. W., Ng, R. H. Y., Wong, S. S. Y., Li, I. W. S., et al. (2012). High mortality associated with *Catabacter hongkongensis* bacteremia. *J. Clin. Microbiol.* 50, 2239–2243. doi: 10.1128/JCM.00128-12

Lau, S. K. P., McNabb, A., Woo, G. K. S., Hoang, L., Fung, A. M. Y., Chung, L. M. W., et al. (2007). *Catabacter hongkongensis* gen. Nov., sp. nov., isolated from blood cultures of patients from Hong Kong and Canada. *J. Clin. Microbiol.* 45, 395–401. doi: 10.1128/JCM.01831-06

Lee, C., Hong, S. N., Paik, N. Y., Kim, T. J., Kim, E. R., Chang, D. K., et al. (2019). CD1d modulates colonic inflammation in NOD2–/– mice by altering the intestinal microbial composition comprising *Acetatifactor muris. J. Crohns Colitis* 13, 1081–1091. doi: 10.1093/ECCO-JCC/JJZ025

Lewis, S., Nash, A., Li, Q., and Ahn, T. H. (2021). Comparison of 16S and whole genome dog microbiomes using machine learning. *BioData Min.* 14, 1–15. doi: 10.1186/S13040-021-00270-X

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: visualization of intersecting sets. *IEEE Trans. Vis. Comput. Graph.* 20, 1983–1992. doi: 10.1109/TVCG.2014.2346248

Li, Y., Jin, Y., Zhang, J., Pan, H., Wu, L., Liu, D., et al. (2021). Recovery of human gut microbiota genomes with third-generation sequencing. *Cell Death Dis.* 12, 1–11. doi: 10.1038/s41419-021-03829-y

Lloyd-Price, J., Abu-Ali, G., and Huttenhower, C. (2016). The healthy human microbiome. *Genome Med.* 8, 1–11. doi: 10.1186/S13073-016-0307-Y

Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., et al. (2022). Metagenome analysis using the kraken software suite. *Nat. Protoc.* 17, 2815–2839. doi: 10.1038/s41596-022-00738-y

Lu, J., and Salzberg, S. L. (2020). Ultrafast and accurate 16S rRNA microbial community analysis using kraken 2. *Microbiome* 8, 1–11. doi: 10.1186/S40168-020-00900-2

Maropola, M. K. A., Ramond, J. B., and Trindade, M. (2015). Impact of metagenomic DNA extraction procedures on the identifiable endophytic bacterial diversity in *Sorghum bicolor* (L. Moench). *J. Microbiol. Methods* 112, 104–117. doi: 10.1016/J.MIMET.2015.03.012

Matsuo, Y., Komiya, S., Yasumizu, Y., Yasuoka, Y., Mizushima, K., Takagi, T., et al. (2021). Full-length 16S rRNA gene amplicon analysis of human gut microbiota using min ION<sup>™</sup> nanopore sequencing confers species-level resolution. *BMC Microbiol.* 21, 1–13. doi: 10.1186/S12866-021-02094-5

Nakano, M. (2018). 16S rRNA gene primer validation for bacterial diversity analysis of vegetable products. *J. Food Prot.* 81, 848–859. doi: 10.4315/0362-028X.JFP-17-346

Park, C., Kim, S. B., Choi, S. H., and Kim, S. (2021). Comparison of 16S rRNA gene based microbial profiling using five next-generation sequencers and various primers. *Front. Microbiol.* 12:715500. doi: 10.3389/FMICB.2021.715500 Patil, S. B., Hurd, T. W., Ghosh, A. K., Murga-Zamalloa, C. A., and Khanna, H. (2011). Functional analysis of retinitis Pigmentosa 2 (RP2) protein reveals variable pathogenic potential of disease-associated missense variants. *PLoS One* 6:e21379. doi: 10.1371/journal.pone.0021379

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/NAR/GKS1219

Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323. doi: 10.1038/nri2515

Satokari, R. M., Vaughan, E. E., Favier, C. F., Doré, J., Edwards, C., and De Vos, W. M. (2002). Diversity of Bifdobacterium and Lactobacillus spp. in breast-fed and formulafed infants as assessed by 16S rDNA sequence differences. *Microb. Ecol. Health Dis.* 14, 97-105. doi: 10.1080/08910600260081748

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, 1–18. doi: 10.1186/GB-2011-12-6-R60

Stecher, B., and Hardt, W. D. (2008). The role of microbiota in infectious disease. *Trends Microbiol.* 16, 107–114. doi: 10.1016/J.TIM.2007.12.008

Sui, H. Y., Weil, A. A., Nuwagira, E., Qadri, F., Ryan, E. T., Mezzari, M. P., et al. (2020). Impact of DNA extraction method on variation in human and built environment microbial community and functional profiles assessed by shotgun metagenomics sequencing. *Front. Microbiol.* 11:536420. doi: 10.3389/FMICB.2020.00953

Szoboszlay, M., Schramm, L., Pinzauti, D., Scerri, J., Sandionigi, A., and Biazzo, M. (2023). Nanopore is preferable over Illumina for 16S amplicon sequencing of the gut microbiota when species-level taxonomic classification, accurate estimation of richness, or focus on rare taxa is required. *Microorganisms* 11:804. doi: 10.3390/MICROORGANISMS11030804

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244

Turner, S., Pryer, K. M., Miao, V. P. W., and Palmer, J. D. (1999). Investigating deep phylogenetic relationships among Cyanobacteria and plastids by small subunit rRNA sequence analysis 1. *J. Eukaryot. Microbiol.* 46, 327–338. doi: 10.1111/J.1550-7408.1999.TB04612.X

Valentin, M., Stéphane, C., Anne-Laure, C., Aurélie, G., Stéphanie, D., Jaafar, B., et al. (2022). A fatal case associated with *Catabacter hongkongensis* bacteremia in lung cancer patient: a case report. *IDCases* 29:e01535. doi: 10.1016/J.IDCR.2022.E01535

Van Der Waaij, D., Berghuis-de Vries, J. M., and Lekkerkerk-Van Der Wees, J. E. C. (1971). Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *J. Hyg.* 69, 405–411. doi: 10.1017/S0022172400021653

Wampach, L., Heintz-Buschart, A., Hogan, A., Muller, E. E. L., Narayanasamy, S., Laczny, C. C., et al. (2017). Colonization and succession within the human gut microbiome by archaea, bacteria, and microeukaryotes during the first year of life. *Front. Microbiol.* 8:241593. doi: 10.3389/FMICB.2017.00738

Wickham, H. (2016) ggplot2: Elegant graphics for data analysis. New York:Springer-Verlag. Available at: https://ggplot2.tidyverse.org/authors.html#citation.

Xu, S., Zhan, L., Tang, W., Wang, Q., Dai, Z., Zhou, L., et al. (2023). MicrobiotaProcess: A comprehensive R package for deep mining microbiome. *The Innovation*. 4:100388. doi: 10.1016/J.XINN.2023.100388

Ye, L., Dong, N., Xiong, W., Li, J., Li, R., Heng, H., et al. (2022). High-resolution metagenomics of human gut microbiota generated by Nanopore and Illumina hybrid metagenome assembly. *Front. Microbiol.* 13:801587. doi: 10.3389/FMICB.2022.801587

Yeo, K., Connell, J., Bouras, G., Smith, E., Murphy, W., Hodge, J. C., et al. (2024). A comparison between full-length 16S rRNA Oxford nanopore sequencing and Illumina V3-V4 16S rRNA sequencing in head and neck cancer tissues. *Arch. Microbiol.* 206, 1–14. doi: 10.1007/S00203-024-03985-7

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and wholegenome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/IJSEM.0.001755