



OPEN ACCESS

EDITED BY

Abhijeet Singh,
Swedish University of Agricultural Sciences,
Sweden

REVIEWED BY

Shuheng Gan,
University of Texas Southwestern Medical
Center, United States
Tianyuan Zhang,
Chinese Academy of Agricultural Sciences,
China

*CORRESPONDENCE

Polina Zoruk
✉ z-polly@mail.ru

RECEIVED 19 August 2025

REVISED 28 November 2025

ACCEPTED 28 November 2025

PUBLISHED 16 December 2025

CITATION

Zoruk P, Morozov M, Veselovsky V,
Strokach A, Babenko V and Klimina K (2025)
Impact of DNA extraction techniques and
sequencing approaches on microbial
community profiling accuracy.
Front. Microbiomes 4:1688681.
doi: 10.3389/fmmbi.2025.1688681

COPYRIGHT

© 2025 Zoruk, Morozov, Veselovsky, Strokach,
Babenko 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.

Impact of DNA extraction techniques and sequencing approaches on microbial community profiling accuracy

Polina Zoruk*, Maxim Morozov, Vladimir Veselovsky,
Aleksandra Strokach, Vladislav Babenko and Ksenia Klimina

Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical
Biological Agency, Moscow, Russia

Background: Quality control in metagenomic data analysis is crucial for ensuring the accuracy and reliability of research results. Among the key steps in microbiome research, DNA extraction plays a critical role, as it directly determines DNA yield, integrity, and representation of microbial taxa.

Results: We compared three commercial DNA extraction kits and our protocol specifically developed for the recovery of high molecular weight (HMW) DNA from complex microbial communities, using the ZymoBIOMICS Gut Microbiome Standard. The PureLin™ Microbiome DNA Purification Kit and our custom protocol provided superior recovery of DNA from Gram-positive bacteria, while the Wizard® kit and our protocol yielded HMW DNA suitable for long-read Oxford Nanopore sequencing. Among sequencing approaches, metagenomic sequencing on the Illumina platform provided the most accurate representation of the reference composition. However, all methods showed limited ability to detect taxa below 0.5% of relative abundance. Additionally, taxonomic classification based on 16S rRNA gene amplicon sequencing data misclassified closely related species due to high gene homology, a limitation not observed with metagenomic approaches.

Conclusions: Our study establishes that a customized DNA extraction protocol is optimal for comprehensive microbiome studies utilizing long-read sequencing technologies. We show that metagenomic sequencing outperforms 16S rRNA gene amplicon sequencing for species-level accuracy, providing a validated benchmark for future gut microbiome research.

KEYWORDS

16S rRNA gene amplicon sequencing, DNA extraction methods, illumina sequencing, metagenomic sequencing, microbiota profiling, Oxford Nanopore Technologies, ZymoBIOMICS gut microbiome standard

Background

The study of microbiota has attracted increasing attention due to its profound implications in both fundamental and applied sciences. Microbial communities play crucial roles in human health, agriculture, and industrial biotechnology. Metagenomic analysis, which enables comprehensive identification and functional profiling of microbial communities, is a key tool for advancing public health, sustainable environmental practices, and biotechnological innovation (Gu et al., 2019; Taş et al., 2021).

Among the essential steps in sequencing, DNA extraction directly impacts genomic coverage and taxonomic resolution. Inefficient lysis – particularly of Gram-positive bacteria – or DNA fragmentation can substantially alter the apparent community composition (Karstens et al., 2021; Kazantseva et al., 2021; Rajar et al., 2022; Wright et al., 2023; Lourenço et al., 2024; Pu et al., 2025). Recognizing this, initiatives such as the MicroBiome Quality Control (MBQC) project and the International Human Microbiome Standards (IHMS) emphasize the importance of reliable DNA extraction for reproducible outcomes (Costea et al., 2017; Leigh Greathouse et al., 2019; Nouws et al., 2020; Nielsen et al., 2022).

Commercial DNA extraction kits vary in performance, each with specific advantages and limitations. Protocols such as MagicPure® Stool and Soil Genomic DNA Kit and PureLink™ Microbiome DNA Purification Kit rely on mechanical lysis, which can effectively disrupt resilient cells but often results in fragmented DNA that is suboptimal for long-read sequencing (Thomas et al., 2012; Yuan et al., 2012). In contrast, the Wizard® Genomic DNA Purification Kit excludes mechanical disruption and enables the recovery of high molecular weight (HMW) DNA, which is essential for Oxford Nanopore Technologies (ONT) and other long-read sequencing technologies.

16S rRNA gene amplicon sequencing remains a widely used approach for microbiota profiling. On the Illumina platform, sequencing of short regions such as V3–V4 or V4–V5 of 16S rRNA gene (~450 bp) provides high read accuracy (> 99%) but limited resolution, generally at the genus level (Ranjan et al., 2016; Larin et al., 2024). By contrast, ONT enables sequencing of the full-length 16S rRNA gene amplicon (V1–V9), allowing for species-level identification (Cha et al., 2023). Nevertheless, amplicon-based methods are less precise in characterizing microbial diversity than whole metagenome sequencing (MS) (Khachatryan et al., 2020). Comprehensive metagenomic analysis allows not only the identification of microorganisms in a sample but also the characterization of their potential biological and ecological roles, provided that sufficient genome coverage is achieved.

The ZymoBIOMICS Gut Microbiome Standard has become a widely used benchmark for evaluating DNA extraction methods, sequencing platforms, and bioinformatics pipelines. This synthetic microbial community mimics the diversity of gut microbiota and can help in evaluating consistency and reproducibility across studies (Mori et al., 2023; Gulyás et al., 2024; Isokääntä et al., 2024). In this study, we compared three commonly used DNA extraction kits — MagicPure® Stool and Soil Genomic DNA Kit

(TransGen Biotech, China), Wizard® Genomic DNA Purification Kit (Promega, USA), and PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, USA) — alongside our protocol specifically developed for the extraction of HMW DNA from fecal microbiota, suitable for subsequent sequencing on platforms such as ONT. Our aim was to assess how different DNA extraction techniques influence the composition and quality of microbiota profiles obtained through sequencing.

Methods

DNA extraction

In this study, we used ZymoBIOMICS Gut Microbiome Standard (D6331, Zymo Research), a mixture of cells from pure cultures of 18 bacterial strains, 2 fungal strains, and 1 archaeal strain in staggered abundances to simulate a typical gut microbiome. Nucleic acids were extracted using three commercial kits – PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, USA), MagicPure® Stool and Soil Genomic DNA Kit (TransGen Biotech, China), and Wizard® Genomic DNA Purification Kit (Promega, USA) – and a custom protocol. For clarity, the following abbreviations are used throughout the text:

- ZymoMS_Thermo – PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, USA).
- ZymoMS_Promega – Wizard® Genomic DNA Purification Kit (Promega, USA).
- ZymoMS_KingFisher – MagicPure® Stool and Soil Genomic DNA Kit (TransGen Biotech, China).
- ZymoMS_Modified – Custom protocol.

PureLink™ microbiome DNA purification kit

DNA extraction from ZymoBIOMICS Gut Microbiome Standard was performed according to our previous protocol (Strokach et al., 2025). Briefly, 75 µL of the sample was supplemented with 400 µL of phosphate-buffered saline (PBS), transferred to bead tubes and homogenized (30 s 7,000 x g) using MagNA Lyser (Roche, Switzerland). The homogenate was centrifuged for 1 min at 7,000 × g (Dlab D3024, Dlab Scientific co. Ltd., Beijing, China), and the resulting supernatant was used for DNA extraction following the manufacturer's instructions.

MagicPure® stool and soil genomic DNA kit

The sample (75 µL) was mixed with 400 µL of PBS and transferred to MagNA Lyser Green Beads (Roche, Switzerland) bead tubes for homogenization on a MagNA Lyser (30 s 7,000 x g) (Roche, Switzerland). After homogenization, the sample was centrifuged for 1 min at 7,000 × g (Dlab D3024, Dlab Scientific

co. ltd., Beijing, China). The supernatant was transferred to a new tube, 20 μ L of proteinase K was added, and the mixture was incubated for 20 min at 65°C. DNA extraction was then performed according to the manufacturer's protocol for the MagicPure Stool and Soil Genomic DNA Kit (TransGen Biotech, China) using a KingFisherTM Purification System (Thermo Fisher Scientific, USA).

Wizard[®] genomic DNA purification kit

To 75 μ L of the sample, 480 μ L of 50 mM EDTA and 120 μ L of lysozyme (10 ng/ μ L) were added. The sample was incubated at 37°C for 60 min, then centrifuged for 2 min at 20,000 $\times$ g (Dlab D3024, Dlab Scientific co. ltd., Beijing, China), and the supernatant was discarded. DNA was extracted from the resulting pellet according to the manufacturer's instructions for the Wizard[®] Genomic DNA Purification Kit (Promega, USA).

Custom protocol

DNA extraction was performed as described previously (Strokach et al., 2025). The sample (75 μ L) was transferred to a 15 mL tube, and 2 mL of 50 mM EDTA along with 500 μ L of 20 mg/mL lysozyme (Sigma-Aldrich, USA) was added. The sample was incubated at a thermal shaker (Allsheng, China) at 37°C 65 $\times$ g for 1.5 h with continuous mixing. After incubation, the sample was centrifuged at 3,000 $\times$ g for 2 min using a Centrifuge 5804 R (Eppendorf, Germany), and the supernatant was transferred to a new tube. The remaining pellet was resuspended in 1 mL of Nuclei Lysis Solution (Wizard[®] Genomic DNA Purification Kit), homogenized by pipetting, and incubated in a thermal shaker (Allsheng, China) at 56°C and 300 $\times$ g for 1.5 h.

Following incubation 400 μ L of Protein Precipitation Solution (Wizard[®] Genomic DNA Purification Kit) was added, vortexed briefly (1 s), and incubated on ice for 10 min. The mixture was centrifuged at 3,700 $\times$ g for 10 min (Dlab D3024, Dlab Scientific co. ltd., Beijing, China), and the supernatant was transferred to a new 15 mL tube containing 1.4 mL of isopropanol. After incubation at –20°C for 1 h, the sample was centrifuged at 3,700 $\times$ g for 20 min (Centrifuge 5804 R, Eppendorf, Germany), and the DNA pellet was washed twice with 80% ethanol, air-dried for 10–15 min, and resuspended in 200 μ L of Low TE buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA).

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 DNA, mixed gently, and incubated at 65°C for 10 min. An equal volume of chloroform was then added, vortexed for 30 s, and centrifuged at 3,500 $\times$ g for 5 min (Dlab D3024, Dlab Scientific co. ltd., Beijing, China). The upper aqueous phase was transferred to a new tube, and mixed with 1.1 volumes of 1% CTAB solution (50 mM Tris pH 8.0, 10 mM EDTA, pre-warmed to 55–65°C), and centrifuged at 3,500 $\times$ g for 5 min

(Dlab D3024, Dlab Scientific co. ltd., Beijing, China), 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 mixture was gently inverted and centrifuged at 3,500 $\times$ g for 8 min (Dlab D3024, Dlab Scientific co. ltd., Beijing, China). The DNA pellet was washed twice with 0.5 mL of 70% ethanol, air-dried, and resuspended in 100 μ L of Low TE buffer, followed by incubation at 65°C for 30 min to ensure complete solubilization.

DNA extracted with each protocol was quantified using a Qubit 4 Fluorometer and a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, USA).

16S rRNA gene amplicon sequencing on the Illumina platform

For amplification of extracted DNA (1–5 ng), 16S rRNA gene amplicon primers targeting the V3–V4 region and incorporating 5'-Illumina adapter sequences (Forward Primer: 5' TCGTC GGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGG NGGCWGCAG; Reverse Primer: 5' GTCTCGTGGGCTC GGAGATGTGTATAAGAGACAGGACTACHVGGG TATCTAATCC) were used (Evrogen Russia). PCR amplification was performed using the Tersus Plus PCR kit (Evrogen, Russia) in a total volume of 25 μ L. Library preparation and sequencing on the Illumina platform were carried out as described previously (Larin et al., 2024). The libraries were sequenced on a MiSeq instrument (Illumina, USA) using the 500-cycle MiSeq reagent kit v2 (Illumina, USA) with 20% PhiX spike-in.

16S rRNA gene amplicon sequencing on the MinION platform

PCR amplification, library preparation, and sequencing on the MinION platform were performed according to our previous protocol (Strokach et al., 2025). Briefly, the full-length 16S rRNA gene amplicon was amplified using primers 27F (AGAGTTTGTATYMTGGCTCAG) and 1492R (GGTTACCT TGTTAYGACTT). PCR amplification was performed using the Tersus Plus PCR kit (Eurogen, Russia) in a total reaction volume of 25 μ L. The cycling conditions were as follows: 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. Amplicon quality was verified by electrophoresis in 1.5% agarose gel. The PCR products were purified using KAPA HyperPure Beads (Roche, Switzerland) according to the manufacturer's protocol.

For sequencing the prepared library (12 μ L) was mixed with 37.5 μ L of sequencing buffer and 25.5 μ L of loading beads, loaded onto an R10.4.1 flow cell (FLO-MIN114; Oxford Nanopore Technologies), and sequenced using the MinION Mk1B device. Data acquisition was performed using MINKNOW software v24.06.14.

Metagenome sequencing on the Illumina platform

Extracted DNA (100 ng) was used for library preparation with the KAPA HyperPlus Kit (Roche, Switzerland) following the manufacturer's instructions. Library underwent a final cleanup using KAPA HyperPure Beads (Roche, Switzerland), after which the library size distribution and quality were assessed using a High Sensitivity DNA chip (Agilent Technologies). Quantification was performed using the Quant-iT DNA Assay Kit, High Sensitivity (Thermo Fisher Scientific, USA). Sequencing was performed on the NextSeq 1000 (Illumina, USA) using the NextSeq 1000/2000 P2 Reagents kit (200 cycles) v3, with 2% PhiX (Illumina) added as an internal control.

Metagenome sequencing on the PromethION platform

HMW DNA (1 µg) was used for library preparation according to the manufacturer's protocol (Oxford Nanopore Technologies, UK). Sequencing libraries were prepared using the Ligation sequencing kit (SQK-LSK114-XL), native barcoding expansion kit (SQK-NBD114.96) and run on an R10.4.1 (FLO-PRO114M) flow cell using a PromethION device. Long reads were generated and basecalled using Dorado v7.6.7 using default parameters (high accuracy (HAC) model, minimum quality value > 7).

Bioinformatics analysis of 16S rRNA gene amplicon data from ONT and Illumina platforms

Reads obtained from ONT sequencing were processed using Porechop v0.2.4 (<https://github.com/rrwick/Porechop>) to remove adapters and PCR primers, with default parameter settings. Reads with Phred quality score below 10 and those shorter than 1,300 nucleotides (nt) or longer than 1,800 nt were filtered out using Chopper v0.6.0 (De Coster and Rademakers, 2023) with the settings -l 1300 and -maxlength 1800. Taxonomic classification was performed with the Emu pipeline v3.4.5 (Curry et al., 2022) with default parameters. Read quality statistics were generated using NanoStat v1.6.0 (De Coster and Rademakers, 2023). The processed data were then imported into RStudio (version 2024.04.2 + 764, R 4.4.1) for downstream analysis using the MicrobiotaProcess package v1.16.1 (Xu et al., 2023).

Raw reads obtained from Illumina sequencing were filtered and trimmed using Fastp v0.23.4 (Chen et al., 2018), followed by further processing with the DADA2 pipeline v1.26.0 (Callahan et al., 2016) was used to carry out further processing. Taxonomic annotation was carried out using the SILVA database v138.1 (Quast et al., 2013). The MicrobiotaProcess package v1.16.1 (Xu et al., 2023) was then used to evaluate the collected data.

To enable direct comparison, quality-filtered ONT reads and merged Illumina paired-end reads were aligned against a reference

database containing the reference genomes of the organisms included in the ZymoBIOMICS Gut Microbiome Standard (<https://s3.amazonaws.com/zymp-files/BioPool/D6331.refseq.zip>). Alignment was performed using BLAST v2.16.0 with the following parameters: e-value of 0.00001, minimum query coverage per high-scoring pair of 70%, and a maximum of one target sequence per query (-max_target_seqs 1).

Bioinformatics analysis of metagenomic data from Illumina and ONT platforms

ONT reads were preprocessed using Porechop v0.2.4 (<https://github.com/rrwick/Porechop>) with default parameter settings and Chopper v0.6.0 (De Coster and Rademakers, 2023) for filtering reads with Phred score below 10. To remove host (human) sequences, reads were aligned to the human reference genome GRCh38 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.40) using Minimap2 v2.26 (Li, 2018). Taxonomic classification of the remaining reads was performed using the Emu pipeline v3.4.5 (Curry et al., 2022) with default parameters. Filtered reads were also mapped to the ZymoBIOMICS Gut Microbiome Standard reference (<https://s3.amazonaws.com/zymp-files/BioPool/D6331.refseq.zip>) using the BWA software package v0.7.18 (Li, 2013) with its default settings, and mapping statistics were calculated using SAMtools v1.21 (Danecek et al., 2021).

Illumina metagenomic reads were quality-checked with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Low-quality sequences and adapters were removed from the raw data using Fastp v0.23.4. Human reads were filtered out by aligning against the GRCh38 genome using HiSAT2 v2.2.1 (Kim et al., 2019). Taxonomic profiling was performed with MetaPhlAn4 v4.0.6 (Blanco-Míguez et al., 2023). The MicrobiotaProcess package v1.16.1 (Xu et al., 2023) was used to analyze the collected data.

Additionally, quality-filtered reads were aligned to the ZymoBIOMICS Gut Microbiome Standard reference database comprising the reference genomes of the organisms included in the Standard (<https://s3.amazonaws.com/zymp-files/BioPool/D6331.refseq.zip>). Alignment was performed using BLAST v2.16.0 with the following parameters: e-value of 0.00001, minimum query coverage per high-scoring pair of 70%, and a maximum of one target sequence per query (-max_target_seqs 1). Alignments with > 95% identity were retained, and average percent identity per reference species was calculated.

Results

Comparison of the efficiency of DNA extraction methods

Three commercial kits and a custom protocol were evaluated based on several key parameters, including total processing time, cell wall disruption method, DNA quantification, final DNA yield,

TABLE 1 Comparison of DNA extraction methods from three commercial kits and custom protocol.

Extraction kits	Protocol duration (min)	Cell wall disruption method	DNA concentration method	Sample concentration, ng/ μ L	DNA quality	A260/A280	A260/A230	N50 (Illumina/ONT)	Mean Q-value (Illumina/ONT)
ZymoMS_Thermo	55	Mechanical	Silica Spin Column	1.1	Fragmented	Was not measured	Was not measured	101/Was not measured	33/Was not measured
ZymoMS_KingFisher	50	Chemical	Magnetic beads	2.67	Fragmented	Was not measured	Was not measured	101/Was not measured	33/Was not measured
ZymoMS_Promega	160	Chemical	Isopropanol precipitation	1.17	High-molecular	1.98	1.86	101/1,368	33/11
ZymoMS_Modified	400	Mechanical	Isopropanol precipitation	1.3	High-molecular	2.01	1.85	101/10,142	33/13

purity and optical density (Table 1). The ZymoMS_Modified protocol had the longest processing time (400 min), as the extended duration of certain steps promotes the extraction of HMW DNA. The highest DNA concentration was obtained using the ZymoMS_KingFisher protocol (2.67 ng/ μ L), while the other methods yielded comparable concentrations ranging from 1.1 to 1.3 ng/ μ L. DNA purity was assessed using A260/A280 and A260/A230 spectrophotometric ratios. ZymoMS_Promega and ZymoMS_Modified showed no evidence of contamination, with A260/A280 values of 1.98 and 2.01, and A260/A230 values of 1.86 and 1.85, respectively. Notably, ZymoMS_Modified demonstrated slightly higher DNA purity compared to ZymoMS_Promega. All samples were evaluated by electrophoresis; however, those extracted using ZymoMS_Thermo and ZymoMS_KingFisher lacked HMW DNA and exhibited pronounced fragmentation. Consequently, these samples were not subjected to spectrophotometric analysis and were excluded from ONT sequencing. For the further sequencing, all DNA samples were pooled in equimolar amounts to ensure comparable read depths across samples (Supplementary Table S1).

For every sequencing technique, alpha diversity metrics, including the Shannon and Simpson indices, were calculated (Supplementary Table S2). The Shannon diversity of the V3–V4 region (Illumina) and full-length 16S rRNA gene (ONT) showed a statistically significant difference as determined by the Kruskal-Wallis test ($p = 0.0163$). However, no significant difference was observed between MS on the Illumina and ONT platform ($p = 0.4561$). Importantly, no significant differences in Shannon indices were detected when comparing variations within each sequencing technique across different DNA extraction protocols. Similarly, Simpson diversity index showed no significant differences between sequencing techniques or among extraction protocols. These findings suggest that, although the observed diversity may be influenced by the choice of sequencing technology, alpha diversity in this artificial community is not affected by the choice of extraction protocol within a given sequencing strategy.

16S rRNA gene amplicon sequencing of the V3–V4 region on the Illumina platform

The sequencing data obtained for the V3–V4 region of the 16S rRNA gene amplicon using the Illumina platform are summarized in Supplementary Table S1. Sequencing of this region provided reliable genus-level profiles. All bacterial genera present in the ZymoBIOMICS Gut Microbiome Standard at an abundance above 0.5% were successfully identified (Figure 1). In contrast, taxa with relative abundances below 0.5% — *Methanobrevibacter smithii*, *Salmonella enterica*, *Enterococcus faecalis*, and *Clostridium perfringens* — were below the reliable detection threshold of the DADA2 pipeline and were not consistently detected across all replicates (Supplementary Table S3).

The presence of these low-abundance species was confirmed by BLAST alignment of reads against reference genomes. For example, the ZymoMS_Thermo protocol detected 0.005% (*M. smithii*, 28

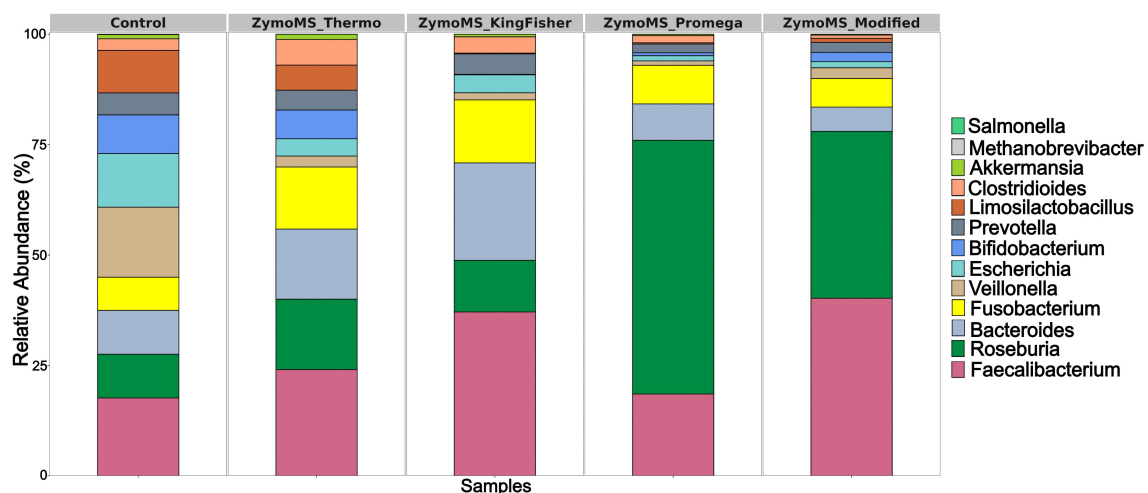


FIGURE 1

Relative abundances of bacterial genera identified using the V3–V4 region of the 16S rRNA gene amplicon sequencing on the Illumina platform. Bar plots show the relative abundances (%) of bacterial genera detected across four DNA extraction methods in comparison with the theoretical composition of the ZymoBIOMICS Gut Microbiome Standard (Control). All detected taxa are shown; however, those with relative abundances below 0.5% are difficult to distinguish due to their low representation. Sequencing of the V3–V4 region on the Illumina platform enables reliable genus-level profiling, confirming the presence of major bacterial taxa included in the standard. Detailed abundance data are provided in [Supplementary Table S3](#).

reads), 0.095% (*S. enterica*, 483 reads), 0.007% (*E. faecalis*, 40 reads), and 0.021% (*C. perfringens*, 110 reads), while other extraction methods yielded comparable but slightly lower detection rates. The ZymoMS_Modified protocol identified 0.0004% (2 reads), 0.005% (21), 0.001% (4), and 0.004% (20) reads, while the ZymoMS_Promega protocol yielded 0.001% (4), 0.012% (56), 0.001% (4), and 0.004% (20) for the same taxa. These findings indicate that Illumina-based 16S rRNA gene amplicon sequencing exhibits limited sensitivity for taxa with very low relative abundances (< 0.5%) in complex microbial communities.

Full-length 16S rRNA gene amplicon sequencing on the ONT platform

The number of reads obtained from 16S rRNA gene amplicon sequencing on the ONT platform is summarized in [Supplementary Table S1](#). Of the 18 bacterial species included in the ZymoBIOMICS Gut Microbiome Standard, 11 were identified. Eukaryotic organisms (*Candida albicans* and *Saccharomyces cerevisiae*) were not detected, as they lack 16S rRNA genes. Similarly, low-abundance microorganisms (< 0.5%) such as *M. smithii*, *S. enterica*, *E. faecalis*, and *C. perfringens* were also not detected using the Emu pipeline (Figure 2). However, BLAST alignment of ONT reads to the reference genomes of these minor species confirmed their presence at low frequencies in several extraction protocols. For *S. enterica*, 0.078% (63 reads) were detected in ZymoMS_KingFisher, 0.025% (14) in ZymoMS_Promega, 0.106% (203) in ZymoMS_Thermo, and 0.039% (21) in ZymoMS_Modified. For *E. faecalis*, 0.005% (4 reads) were detected in ZymoMS_KingFisher, 0.007% (4) in ZymoMS_Promega, 0.027% (52) in ZymoMS_Thermo, and

0.022% (12) in ZymoMS_Modified. Similarly, *C. perfringens* was identified with 0.018% (10 reads) in ZymoMS_Promega and 0.058% (110) in ZymoMS_Thermo. Gram-positive bacteria, including *Bifidobacterium adolescentis* and *Limosilactobacillus fermentum*, were extracted more efficiently using the ZymoMS_Thermo protocol, likely due to the bead-beating step that facilitating cell wall disruption.

Analysis of full-length 16S rRNA gene amplicon sequencing data obtained on the ONT platform revealed the presence of the major bacterial species included in the ZymoBIOMICS Gut Microbiome Standard, except for *Veillonella rogosae*. This species was misidentified as *V. parvula* across all four extraction methods. The misidentification is attributed to the high sequence similarity between the 16S rRNA genes of these species, which share 98% homology (Arif et al., 2008). However, their whole genome shares only 67% homology, illustrating the limitations of 16S rRNA gene amplicon sequencing in distinguishing closely related taxa. These findings highlight the importance of complementing amplicon-based approaches with MS to achieve higher species-level accuracy. Among the tested extraction methods, the microbial composition derived from the ZymoMS_Thermo protocol most closely matched the theoretical composition of the ZymoBIOMICS Gut Microbiome Standard ([Supplementary Table S3](#)).

Metagenome sequencing for comprehensive microbial profiling

The number of reads obtained from MS on the Illumina platform is shown in [Supplementary Table S1](#). The observed difference in read counts between samples extracted using the ZymoMS_Promega and ZymoMS_Modified protocols can be

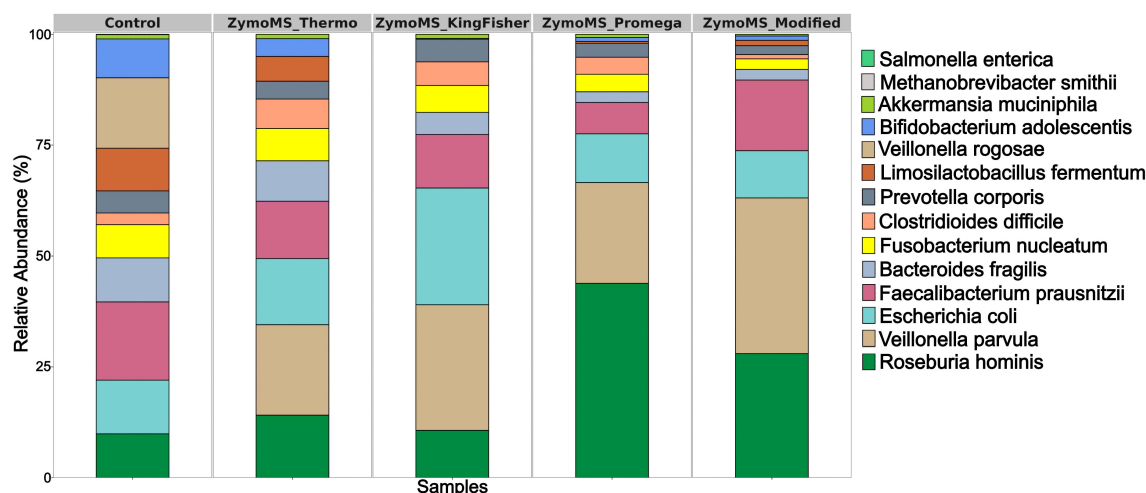


FIGURE 2

Relative abundances of bacterial species identified based on full-length 16S rRNA gene amplicon sequencing on the ONT platform. Bar plots show the relative abundances (%) of bacterial species detected across four DNA extraction methods in comparison with the ZymoBIOMICS Gut Microbiome Standard (Control). All detected taxa are shown, however, species with relative abundances below 0.5% are difficult to distinguish due to their low representation. Full-length 16S rRNA gene sequencing on the ONT platform improved species-level resolution for dominant taxa, while the presence of low-abundance species was confirmed by BLAST alignment. Detailed results are provided in [Supplementary Table S3](#).

attributed to the substantially longer average read length in the latter (10,142 bp vs. 1,327 bp), resulting in lower number of reads but longer fragments ([Table 1](#)). MS on the Illumina platform detected all major bacterial species present in the ZymoBIOMICS Gut Microbiome Standard, while species with theoretical abundance below 0.5% mostly remained undetected. A notable exception was *M. smithii*, which was successfully identified in the ZymoMS_KingFisher and ZymoMS_Modified samples ([Figure 3](#), [Supplementary Table S4](#)).

Moreover, some reads obtained with the ZymoMS_KingFisher protocol were assigned to *Escherichia marmotae* (0.002%), a bacterial species not included in the ZymoBIOMICS Gut

Microbiome Standard ([Supplementary Table S4](#)). This low-level false-positive assignment likely resulted from a bioinformatic classification artifact. Notably, *V. rogosae* was correctly identified using MS, in contrast to 16S rRNA gene amplicon sequencing, which misidentified it as *V. parvula*. These results underscore the superior accuracy of MS in identifying major bacterial species and highlight its value for accurate and comprehensive microbiome profiling.

Among the tested methods, the custom ZymoMS_Modified protocol for HMW DNA extraction provided the most comprehensive representation, achieving a complete match with the theoretical composition of the ZymoBIOMICS Gut Microbiome

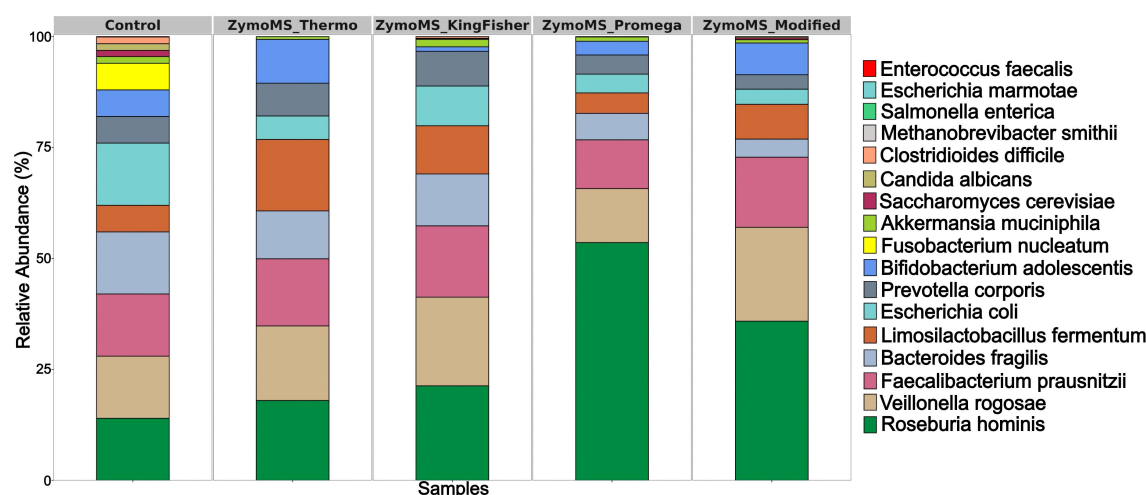


FIGURE 3

Relative abundance (%) of bacteria at the species level identified by MS on the Illumina platform.

TABLE 2 Average percent identity between obtained reads and reference genomes (alignments > 95% identity).

Genome	Strain ID	GenBank	Sample			
			ZymoMS_Thermo	ZymoMS_KingFisher	ZymoMS_Promega	ZymoMS_Modified
<i>Faecalibacterium prausnitzii</i>	AP34BHI	GCA_028743395.1	98.9	98.7	98.7	98.8
<i>Veillonella rogosae</i>	AC2811 AN NA 2	GCA_028743475.1	99.6	99.6	99.5	99.5
<i>Roseburia hominis</i>	OB EAV1-11 DCM	GCA_028743455.1	98.1	97.8	97.7	97.7
<i>Bacteroides fragilis</i>	OB EAV1-11 D6 FAA	GCA_028743275.1	99.7	99.6	99.6	99.6
<i>Prevotella corporis</i>	OB21 FMU 4	GCA_028743775.1	99.2	99.1	99.1	99.1
<i>Bifidobacterium adolescentis</i>	LMG 10502	GCA_028743295.1	98.5	98.4	98.4	98.5
<i>Fusobacterium nucleatum</i>	2/1/50A	GCA_028743415.1	98.4	97.8	97.8	96.2
<i>Lactobacillus fermentum</i>	B-1840	GCA_030770375.1	99.3	99.4	99.4	99.4
<i>Clostridium difficile</i>	P4D3A1-1	GCA_028743315.1	99.1	99.3	99.3	99.4
<i>Akkermansia muciniphila</i>	OB21 FAA NB 28	GCA_028743255.1	99.7	99.2	99.0	99.3
<i>Methanobrevibacter smithii</i>	DSM 861	GCA_000016525.1	99.9	98.9	99.6	99.4
<i>Salmonella enterica</i>	B-4212	GCA_028743635.1	99.2	99.6	99.6	99.6
<i>Enterococcus faecalis</i>	IP101412 AER FAA 2	no number	not detected	100.0	99.2	99.9
<i>Clostridium perfringens</i>	OB21 TSA 19	GCA_028743735.1	not detected	100.0	not detected	not detected
<i>Escherichia coli</i>	JM109	GCF_009916115.1	99.1	99.4	99.4	99.4
<i>Escherichia coli</i>	B-3008	GCA_028743355.1	99.6	98.8	99.1	98.9
<i>Escherichia coli</i>	B-2207	GCA_028743335.1	98.9	98.9	98.9	98.9
<i>Escherichia coli</i>	B-766	GCA_028743755.1	99.7	99.6	99.6	99.6
<i>Escherichia coli</i>	B-1109	GCA_028743555.1	98.3	99.1	99.2	99.1
<i>Candida albicans</i>	IHEM 3108	no number	not detected	not detected	not detected	not detected
<i>Saccharomyces cerevisiae</i>	Y-567	GCA_030867715.1	not detected	not detected	not detected	not detected

Standard. In contrast, while the ZymoMS_Thermo protocol yielded a bacterial composition similar to the control, it failed to detect the two eukaryotic species, *C. albicans* and *S. cerevisiae*. This was likely because the lysis conditions optimized for bacterial cells in the ZymoMS_Thermo were insufficient to disrupt the more robust fungal cell walls. Consequently, the ZymoMS_Modified protocol, with its more rigorous and extended lysis steps, proved superior for the simultaneous extraction of DNA from both bacterial and eukaryotic members of a microbial community. Furthermore, both the ZymoMS_Thermo and ZymoMS_Modified protocols demonstrated enhanced efficiency in extracting nucleic acids from Gram-positive bacteria (*B. adolescentis* and *L. fermentum*), as

evidenced by their higher relative abundances compared to other methods (Supplementary Tables S3, S4).

Although many bioinformatics tools allow strain-level identification (De Coster and Rademakers, 2023), no strains identical to those declared in the ZymoBIOMICS Gut Microbiome Standard were detected in strain-level analysis. BLAST alignments of reads to the reference strains included in the standard showed high sequence homology, confirming the accuracy of species-level identification (Table 2). Sequences obtained using the ZymoMS_KingFisher protocol exhibited 97.8% to 100% homology with NCBI references, while those from ZymoMS_Modified ranged from 96.2% to 99.9%. Reads from the

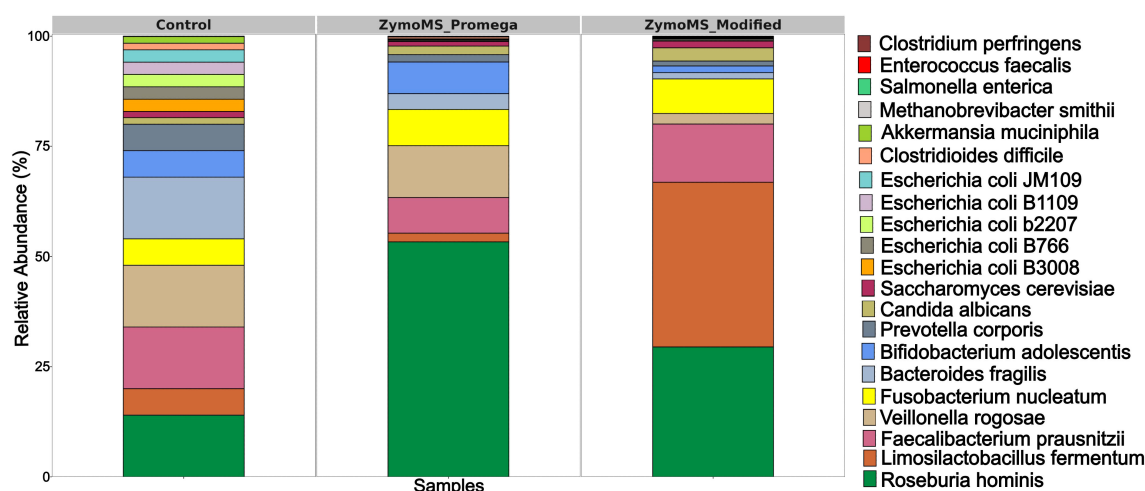


FIGURE 4
Relative abundance (%) of bacteria at the species level obtained with MS on the ONT platform.

ZymoMS_Promega and ZymoMS_Thermo protocols demonstrated 97.7% to 99.6% and 98.1% to 99.9% homology, respectively.

The high alignment rates to bacterial strains in the ZymoBIOMICS Gut Microbiome Standard underscore the utility of MS for comprehensive and accurate microbiome profiling, while emphasizing the importance of optimized extraction protocols to enhance the detection of low-abundance taxa.

The ZymoMS_Promega and ZymoMS_Modified protocols were selected for sequencing on the ONT platform because of their ability to recover HMW DNA. Analysis of the sequencing data using the Emu pipeline did not detect the low-abundance bacterial species present in the standard. To confirm their presence, a direct read mapping approach was employed against the corresponding reference genomes. This sensitive alignment-based analysis successfully identified reads aligning to all bacterial species included in the ZymoBIOMICS Gut Microbiome Standard (Figure 4, Supplementary Table S5), confirming the presence of their DNA in the sequencing libraries. The primary challenge, therefore, lies not in the wet-lab protocol but in the bioinformatic detection limits for very low-abundance taxa.

Notably, the quality of reads mapped to minor bacteria was, for the most part, significantly lower than that of reads mapped to dominant taxa (Supplementary Tables S6 and S7), likely reflecting their initially low abundance in the sample. Detecting low-abundance bacteria remains challenging with current sequencing technologies. A higher sequencing depth and greater number of contiguous genome fragments were interpreted as indicators of superior DNA integrity and extraction efficiency, whereas lower coverage and increased genome fragmentation suggested degraded or incomplete DNA recovery. A statistical comparison of the coverage depth between the two extraction kits revealed a significant difference (Wilcoxon rank sum test, p -value < 0.05), supporting the conclusion that the choice of extraction method significantly influences DNA yield and quality (Supplementary Figure S1).

Discussion

Studies utilizing ZymoBIOMICS Gut Microbiome Standard have demonstrated its effectiveness in benchmarking DNA extraction protocols, particularly in assessing their ability to recover diverse microbial communities without bias. Comparative analyses using the ZymoBIOMICS Gut Microbiome Standard have shown that mechanical lysis methods often introduce fragmentation, while enzymatic approaches preserve DNA integrity for long-read sequencing. The frequent use of this standard across various platforms underscores its critical role in addressing technical variability and advancing microbiome research (Bertolo et al., 2024; Lin et al., 2024).

Our findings confirmed that mechanical lysis methods, such as those employed in the MagicPure[®] and PureLink[™] kits, lead to significant fragmentation. In contrast, enzymatic methods, including those in the Wizard[®] kit and the custom protocol, successfully preserved HMW DNA suitable for MS on the ONT platform. The ability to preserve DNA integrity is crucial for applications requiring long reads, such as species-level identification or strain-level analysis. The primary strength of using the ZymoBIOMICS Gut Microbiome Standard, as demonstrated in this study, is the ability to make definitive, quantitative assessments of DNA extraction and sequencing protocols against a known ground truth. This approach unequivocally identified the performance characteristics of each method. To complement these controlled findings, we also sought to verify the practical utility of the top-performing protocols. In a parallel study using complex, biological mouse gut microbiota samples, the ZymoMS_Thermo and ZymoMS_Modified protocols were applied and yielded high-quality sequencing data (Strokach et al., 2025). The successful application of these methods in a real-world scenario demonstrates their robustness for DNA recovery from complex biological matrices, supporting their broader applicability beyond synthetic communities.

Despite these advances, both Illumina and ONT sequencing approaches demonstrated limitations in identifying taxa with abundances below 0.5%, which remains a common challenge in microbiome research. Low-abundance species are often overlooked due to insufficient sequencing depth, extraction or algorithmic constraints. Increasing sequencing coverage can improve the detection of rare taxa, but this inevitably raises costs and may limit the feasibility of large-scale studies. A recent study systematically evaluated the influence of sequencing depth on microbial diversity estimates (Veselovsky et al., 2025) and demonstrated that increasing the number of reads per sample improved taxonomic resolution up to a certain threshold, beyond which additional sequencing provided minimal benefit. Taxa with relative abundances below 0.5% should be interpreted with caution.

Recent advances in nanopore sequencing technologies have significantly enhanced the accuracy and taxonomic resolution of full-length 16S rRNA gene amplicon analyses. Parameters that can influence the detection of bacterial species, particularly those present in low abundance, include basecalling mode, sequencing depth, read quality, and bioinformatic post-processing. The choice of basecalling mode in ONT workflows - specifically the high-accuracy “hac” and super-accuracy “sup” modes - can have measurable effects on per-base accuracy and, consequently, on the reliable detection of rare taxa. Methodological developments have demonstrated that ONT’s R10.4.1 chemistry, combined with the latest basecallers, can achieve Q20 and even Q28 accuracy levels (Zhang et al., 2023, 2024), narrowing the performance gap between nanopore and short-read platforms such as Illumina. This improvement may be particularly beneficial for distinguishing low-abundance or closely related bacterial species that were previously unresolved due to read-level errors. Nevertheless, evidence from recent comparative studies indicates that these improvements in basecalling accuracy may not always translate into substantially different microbiome diversity outcomes. For instance, (Aja-Macaya et al., 2025) reported that although the “sup” mode provided higher per-base accuracy, the overall microbial diversity and taxonomic profiles were largely consistent with those obtained using the “hac” mode. Similarly, benchmarking analyses using the ZymoBIOMICS Gut Microbiome Standard (Veselovsky et al., 2025) demonstrated that both “hac” and “sup” basecallers yielded highly comparable read counts, precision, recall, and F1-scores, with negligible impact on taxonomic assignment accuracy. The primary distinction was the increased computational demand associated with sup mode.

Despite these observations, the detection of underrepresented bacterial species remains a persistent challenge, particularly when their relative abundance falls below 0.5%. To improve the identification of such taxa, several complementary strategies can be employed. Increasing sequencing depth remains the most direct approach, as it enhances the likelihood of capturing rare microbial signatures, though this must be balanced against cost and diminishing returns beyond a certain threshold. Another promising approach involves adaptive sampling, where real-time nanopore sequencing selectively enriches for underrepresented taxa or specific genomic regions. On the computational side, the use of error-correction tools,

such as Medaka or NanoPolish, combined with high-quality reference databases (e.g., SILVA, GTDB), can improve classification accuracy by reducing base-level noise. Additionally, hybrid assembly or co-binning strategies that integrate ONT long reads with Illumina short reads can increase confidence in low-abundance species identification and recover more complete genomes. Future research should also explore machine-learning-based taxonomic classifiers trained on error-aware nanopore data, which could better handle sequencing artifacts and improve the detection of rare taxa in complex communities. Collectively, these methodological and computational refinements are expected to advance the ability to profile low-abundance microbial members, enabling more accurate characterization of microbial communities, detection of potential biomarkers, and understanding of subtle ecological dynamics in clinical, environmental, and industrial microbiomes.

In our study, MS on Illumina proved to be the most comprehensive approach, accurately detecting all declared bacterial species in the ZymoBIOMICS Gut Microbiome Standard, including *V. rogosae*, which was misidentified as *V. parvula* in 16S rRNA gene amplicon sequencing data. In contrast, 16S rRNA gene amplicon sequencing is more suitable for qualitative analysis due to its biases in the quantitative determination of bacterial composition in the ZymoBIOMICS Gut Microbiome Standard. For instance, it has been previously shown that the V3–V4 region of 16S rRNA gene of *Bacillus cereus* is identical to that of *B. mobilis*, making these species indistinguishable through this approach. Similarly, the V3–V4 region of 16S rRNA gene of *Streptococcus pneumoniae* differs by only a single nucleotide from that of *S. infantis* (Gwak and Rho, 2020). These limitations underscore the need for improved algorithms capable of handling sequencing errors and distinguishing closely related species, particularly when sequencing the complete 16S rRNA gene.

We note that the DADA2 (Callahan et al., 2016) pipeline provides high accuracy for short-read Illumina data, but in many cases its performance is limited to genus-level resolution due to the short amplicon size, and it may misclassify closely related species. Importantly, several studies have demonstrated that DADA2 is not well-suited for ONT data because it was originally developed for the low error rates characteristic of Illumina sequencing. When applied to ONT reads, DADA2 performs poorly due to its stringent error model assumptions and inability to handle the higher indel and substitution error rates inherent to nanopore sequencing (Callahan et al., 2019; Santos et al., 2020; Chang et al., 2024). These reports consistently indicate that DADA2 tends to over-split true biological variants and produce unreliable ASV inference when used on ONT datasets.

In contrast, the Emu (Curry et al., 2022) pipeline is designed for full-length 16S rRNA gene amplicon reads generated by ONT, allowing for species-level classification; however, it exhibits reduced sensitivity for low-abundance taxa and is highly dependent on the completeness of the reference database. We also emphasize that both tools may contribute to discrepancies in the detection of rare or closely related species. Recent comparative studies further highlight these differences, showing that taxonomic resolution and sensitivity can vary substantially depending on the chosen pipeline (Strokhach et al., 2025; Veselovsky et al., 2025).

Challenges in analyzing 16S rRNA gene amplicon sequencing data on the ONT platform may stem from multiple factors. These include imperfections in primer complementarity (Frank et al., 2008), periodic taxonomic restructuring in reference databases (Vanhee et al., 2024), and the inherent detection limits of the ONT 16S rRNA gene amplicon sequencing platform (Lao et al., 2023). These factors collectively highlight the need for continuous improvement in both laboratory protocols and computational pipelines to achieve more accurate and reliable microbiome analyses. The ZymoBIOMICS Gut Microbiome Standard was used to quality control the microbiome studies. The taxonomic analysis results showed that current bioinformatics methods are insufficiently accurate in identifying taxa with a relative abundance of less than 0.5% (Veselovsky et al., 2025).

When analyzing the ZymoMS_Thermo sample using two major sequencing strategies – 16S rRNA gene amplicon sequencing and MS – we observed a comparable microbial composition, despite using different Illumina platforms for each. This indicates that MS offers robust performance in profiling well-defined microbial communities, even across distinct sequencing setups. However, for 16S rRNA-based taxonomic profiling, ONT outperformed Illumina in terms of species-level resolution, owing to its ability to sequence the full-length 16S rRNA gene amplicon. These findings support the view that the choice of sequencing platform and method should be aligned with specific research goals. Moreover, previous studies have shown that using a hybrid assembly approach combining Illumina and ONT reads can further improve the resolution and number of identified contigs (Overholt et al., 2020).

The variability in DNA extraction efficiency across methods further influenced the outcomes of sequencing. Protocols like ZymoMS_Thermo and the custom method demonstrated better performance in extracting nucleic acids from Gram-positive bacteria, underscoring the critical role of mechanical lysis, such as bead-beating, in recovering DNA from resilient cell types.

Conclusions

The accuracy of 16S rRNA gene amplicon sequencing can be enhanced by selecting appropriate DNA extraction methods. Using a combination of the PureLink™ Microbiome DNA Purification Kit (Thermo Fisher Scientific, USA) for DNA extraction and 16S rRNA gene amplicon sequencing on the ONT and Illumina platforms, we obtained results that closely matched the theoretical composition of the ZymoBIOMICS Gut Microbiome Standard. The Wizard® Genomic DNA Purification Kit (Promega) and the custom protocol proved most suitable for MS on the ONT platform, owing to their ability to recover HMW DNA. The PureLink™ kit and the custom protocol also demonstrated high efficiency in recovering DNA from Gram-positive bacteria.

Analysis of 16S rRNA gene amplicon sequencing data may lead to the misclassification of closely related species due to the high sequence homology of the 16S rRNA gene. This limitation can be addressed by MS, which provides greater taxonomic resolution. However, taxa with low-relative abundance (< 0.5%) remain difficult to detect with both 16S rRNA amplicon sequencing and MS approaches, primarily due to methodological constraints and limited sequencing depth. Despite these limitations, 16S rRNA gene amplicon sequencing remains a cost-effective tool for large-scale studies aimed at broad microbial community profiling. For studies requiring higher taxonomic resolution and comprehensive characterization, MS is preferable.

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.

Author contributions

KK: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. ZP: Data curation, Formal Analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. MM: Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. VV: Investigation, Methodology, Validation, Visualization, Writing – review & editing. SA: Formal Analysis, Investigation, Methodology, Writing – review & editing. VB: Formal Analysis, Methodology, Software, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. The financial support for this study was graciously 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>).

Conflict of interest

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

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Aja-Macaya, P., Conde-Pérez, K., Trigo-Tasende, N., Buetas, E., Nasser-Ali, M., Nión, P., et al. (2025). Nanopore full length 16S rRNA gene sequencing increases species resolution in bacterial biomarker discovery. *Sci. Rep.* 15, 1–12. doi: 10.1038/S41598-025-10999-8;SUBJMETA
- Arif, N., Do, T., Byun, R., Sheehy, E., Clark, D., Gilbert, S. C., et al. (2008). *Veillonella rogosae* sp. nov., an anaerobic, gram-negative coccus isolated from dental plaque. *Int. J. Syst. Evol. Microbiol.* 58, 581–584. doi: 10.1099/IJS.0.65093-0
- Bertolo, A., Valido, E., and Stoyanov, J. (2024). Optimized bacterial community characterization through full-length 16S rRNA gene sequencing utilizing MinION nanopore technology. *BMC Microbiol.* 24, 1–12. doi: 10.1186/S12866-024-03208-5
- Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., et al. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* 41, 1633–1644. doi: 10.1038/s41587-023-01688-w
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., et al. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* 47, e103–e103. doi: 10.1093/NAR/GKZ569
- 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 MinION and Illumina MiSeq sequencing. *Front. Microbiol.* 14. doi: 10.3389/FMICB.2023.1148466/BIBTEX
- Chang, J. J. M., Ip, Y. C. A., Neo, W. L., Mowe, M. A. D., Jaafar, Z., and Huang, D. (2024). Primed and ready: nanopore metabarcoding can now recover highly accurate consensus barcodes that are generally indel-free. *BMC Genomics* 25, 1–16. doi: 10.1186/S12864-024-10767-4/FIGURES/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
- Costea, P. I., Zeller, G., Sunagawa, S., Pelletier, E., Alberti, A., Levenez, F., et al. (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* 35, 1069–1076. doi: 10.1038/nbt.3960
- Curry, K. D., Wang, Q., Nute, M. G., Tyshayeva, 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
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi: 10.1093/GIGASCIENCE/GIAB008
- De Coster, W., and Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39. doi: 10.1093/BIOINFORMATICS/BTAD311
- Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Appl. Environ. Microbiol.* 74, 2461–2470. doi: 10.1128/AEM.02272-07/ASSET/D62E6744-DDF8-4FC4-B592-BA1C020711A3
- Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. *Annu. Rev. Pathol. Mech. Dis.* 14, 319–338. doi: 10.1146/ANNUREV-PATHMECHDIS-012418-012751
- Gulyás, G., Kakuk, B., Dörmő, Á., Járny, T., Prazsák, I., Csabai, Z., et al. (2024). Cross-comparison of gut metagenomic profiling strategies. *Commun. Biol.* 7, 1–22. doi: 10.1038/s42003-024-07158-6
- Gwak, H. J., and Rho, M. (2020). Data-driven modeling for species-level taxonomic assignment from 16S rRNA: application to human microbiomes. *Front. Microbiol.* 11. doi: 10.3389/FMICB.2020.570825
- Isokääntä, H., Tomnikov, N., Vanhatalo, S., Munukka, E., Huovinen, P., Hakanen, A. J., et al. (2024). High-throughput DNA extraction strategy for fecal microbiome studies. *Microbiol. Spectr.* 12:e0293223. doi: 10.1128/SPECTRUM.02932-23
- Karstens, L., Siddiqui, N. Y., Zaza, T., Barstad, A., Amundsen, C. L., and Sysoeva, T. A. (2021). Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *Sci. Rep.* 11, 1–9. doi: 10.1038/s41598-021-85482-1
- Kazantseva, J., Malv, E., Kaleda, A., Kallastu, A., and Meikas, A. (2021). Optimisation of sample storage and DNA extraction for human gut microbiota studies. *BMC Microbiol.* 21, 1–13. doi: 10.1186/S12866-021-02233-Y
- Khachatryan, L., de Leeuw, R. H., Kraakman, M. E. M., Pappas, N., te Raa, M., Mei, H., et al. (2020). Taxonomic classification and abundance estimation using 16S and WGS—A comparison using controlled reference samples. *Forensic Sci. Int. Genet.* 46, 102257. doi: 10.1016/j.fsigen.2020.102257
- 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. doi: 10.1038/S41587-019-0201-4
- Lao, H. Y., Wong, L. L. Y., Hui, Y., Ng, T. T. L., Chan, C. T. M., Lo, H. W. H., et al. (2023). The clinical utility of Nanopore 16S rRNA gene sequencing for direct bacterial identification in normally sterile body fluids. *Front. Microbiol.* 14. doi: 10.3389/FMICB.2023.1324494
- Larin, A. K., Klimina, K. M., Veselovsky, V. A., Olekhovich, 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 Genomic Data* 25, 1–7. doi: 10.1186/S12863-024-01192-3
- Leigh Greathouse, K., Sinha, R., and Vogtmann, E. (2019). DNA extraction for human microbiome studies: The issue of standardization. *Genome Biol.* 20, 1–4. doi: 10.1186/S13059-019-1843-8
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. doi: 10.48550/arXiv.1303.3997
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/BIOINFORMATICS/BTY191
- Lin, X., Waring, K., Ghezzi, H., Tropini, C., Tyson, J., and Ziels, R. M. (2024). High accuracy meets high throughput for near full-length 16S ribosomal RNA amplicon sequencing on the Nanopore platform. *PNAS Nexus* 3, pgae411. doi: 10.1093/PNASNEXUS/PGAE411
- Lourenço, C. F., Almeida, A. R., Soares, A. M. V. M., and Marques, C. R. (2024). Efficiency comparison of DNA extraction kits for analysing the cockle gut bacteriome. *Heliyon* 10, e38846. doi: 10.1016/J.HELİYON.2024.E38846

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/frmbi.2025.1688681/full#supplementary-material>

- Mori, H., Kato, T., Ozawa, H., Sakamoto, M., Murakami, T., Taylor, T. D., et al. (2023). Assessment of metagenomic workflows using a newly constructed human gut microbiome mock community. *DNA Res.* 30, 1–10. doi: 10.1093/DNARES/DSAD010
- Nielsen, J., Tillegreen, C. B., and Petranovic, D. (2022). Innovation trends in industrial biotechnology. *Trends Biotechnol.* 40, 1160–1172. doi: 10.1016/j.tibtech.2022.03.007
- Nouws, S., Bogaerts, B., Verhaegen, B., Denayer, S., Piérard, D., Marchal, K., et al. (2020). Impact of DNA extraction on whole genome sequencing analysis for characterization and relatedness of Shiga toxin-producing *Escherichia coli* isolates. *Sci. Rep.* 10, 1–16. doi: 10.1038/s41598-020-71207-3
- Overholt, W. A., Hölzer, M., Geesink, P., Diezel, C., Marz, M., and Küsel, K. (2020). Inclusion of Oxford Nanopore long reads improves all microbial and viral metagenome-assembled genomes from a complex aquifer system. *Environ. Microbiol.* 22, 4000–4013. doi: 10.1111/1462-2920.15186
- Pu, Y., Zhou, X., Cai, H., Lou, T., Liu, C., Kong, M., et al. (2025). Impact of DNA extraction methods on gut microbiome profiles: A comparative metagenomic study. *Phenomics* 5, 76–90. doi: 10.1007/S43657-025-00232-X
- 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
- Rajar, P., Dhariwal, A., Salvadori, G., Junges, R., Åmdal, H. A., Berild, D., et al. (2022). Microbial DNA extraction of high-host content and low biomass samples: Optimized protocol for nasopharynx metagenomic studies. *Front. Microbiol.* 13. doi: 10.3389/fmicb.2022.1038120
- 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
- Santos, A., van Aerle, R., Barrientos, L., and Martinez-Urtaza, J. (2020). Computational methods for 16S metabarcoding studies using Nanopore sequencing data. *Comput. Struct. Biotechnol. J.* 18, 296–305. doi: 10.1016/j.csbj.2020.01.005/ASSET/4197F224-886D-44AB-AB96-3528CC39ED48
- Strokach, A., Zoruk, P., Boldyreva, D., Morozov, M., Olekhovich, E., Veselovsky, V., et al. (2025). Comparative evaluation of sequencing technologies and primer sets for mouse gut microbiota profiling. *Front. Microbiol.* 16. doi: 10.3389/fmicb.2025.1584359
- Taş, N., de Jong, A. E., Li, Y., Trubl, G., Xue, Y., and Dove, N. C. (2021). Metagenomic tools in microbial ecology research. *Curr. Opin. Biotechnol.* 67, 184–191. doi: 10.1016/j.copbio.2021.01.019
- Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microb. Inform. Exp.* 2, 1–12. doi: 10.1186/2042-5783-2-3
- Vanhee, M., Floré, K., Vanthourenhout, S., Hellemans, J., Muyldermans, A., and Reyniers, M. (2024). Implementation of full-length 16S nanopore sequencing for bacterial identification in a clinical diagnostic setting. *Diagn. Microbiol. Infect. Dis.* 108, 116156. doi: 10.1016/j.diagmicrobio.2023.116156
- Veselovsky, V., Romanov, M., Zoruk, P., Larin, A., Babenko, V., Morozov, M., et al. (2025). Comparative evaluation of sequencing platforms: Pacific Biosciences, Oxford Nanopore Technologies, and Illumina for 16S rRNA-based soil microbiome profiling. *Front. Microbiol.* 16. doi: 10.3389/fmicb.2025.1633360
- Wright, M. L., Podnar, J., Longoria, K. D., Nguyen, T. C., Lim, S., Garcia, S., et al. (2023). Comparison of commercial DNA extraction kits for whole metagenome sequencing of human oral, vaginal, and rectal microbiome samples. *bioRxiv Prepr. Serv. Biol.* 4, 2023.02.01.526597. doi: 10.1101/2023.02.01.526597
- Xu, S., Zhan, L., Tang, W., Wang, Q., Dai, Z., Zhou, L., et al. (2023). MicrobiotaProcess: A comprehensive R package for deep mining microbiome. doi: 10.1016/j.xinn.2023.100388
- Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., and Forney, L. J. (2012). Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PloS One* 7, e33865. doi: 10.1371/JOURNAL.PONE.0033865
- Zhang, T., Li, H., Jiang, M., Hou, H., Gao, Y., Li, Y., et al. (2024). Nanopore sequencing: flourishing in its teenage years. *J. Genet. Genomics* 51, 1361–1374. doi: 10.1016/j.jgg.2024.09.007
- Zhang, T., Li, H., Ma, S., Cao, J., Liao, H., Huang, Q., et al. (2023). The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl. Environ. Microbiol.* 89, e0060523. doi: 10.1128/AEM.00605-23