

<https://doi.org/10.1038/s42003-024-07158-6>

Cross-comparison of gut metagenomic profiling strategies

Check for updates

Gábor Gulyás^{1,2,3}, Balázs Kakuk^{1,2,3}, Ákos Dörmő^{1,2,3}, Tamás Járay^{1,2}, István Prazsák^{1,2}, Zsolt Csabai^{1,2}, Miksa Máté Henkrich^{1,2}, Zsolt Boldogkői^{1,2}✉ & Dóra Tombácz^{1,2}✉

The rapid advancements in sequencing technologies and bioinformatics have enabled metagenomic research of complex microbial systems, but reliable results depend on consistent laboratory and bioinformatics approaches. Current efforts to identify best practices often focus on optimizing specific steps, making it challenging to understand the influence of each stage on microbial population analysis and compare data across studies. This study evaluated DNA extraction, library construction methodologies, sequencing platforms, and computational approaches using a dog stool sample, two synthetic microbial community mixtures, and various sequencing data sources. Our work, the most comprehensive evaluation of metagenomic methods to date. We developed a software tool, termed minitax, which provides consistent results across the range of platforms and methodologies. Our findings showed that the Zymo Research Quick-DNA HMW MagBead Kit, Illumina DNA Prep library preparation method, and the minitax bioinformatics tool were the most effective for high-quality microbial diversity analysis. However, the effectiveness of pipelines or method combinations is sample-specific, making it difficult to identify a universally optimal approach. Therefore, employing multiple approaches is crucial for obtaining reliable outcomes in microbial systems.

Metagenomics has recently experienced a boom due to rapid advancements in molecular biology, genomics, and bioinformatics^{1,2}. Recent studies have uncovered the vast diversity and complexity of the gut microbiome and its essential role in human health and disease^{3–8}. We are now able to conduct comprehensive taxonomic and functional profiling of microbial communities. The modern nucleic acid purification methods are highly efficient, the library preparation techniques yield high-quality samples, and current software tools adeptly handle metagenomics data. Although a plethora of metagenomics methods are available for both wet and dry lab procedures, there have been limited standardization initiatives to date^{9,10}. Although several publications address the impact of sequencing techniques and bioinformatics strategies on the variability of results^{2,11–13}, remarkably, no comprehensive study simultaneously assesses the complete laboratory processes and the subsequent data analysis protocols (Supplementary Data 1).

Choosing the best DNA purification method^{14–19}, sequencing platform, shotgun versus amplicon sequencing technique^{20,21}, optimal 16S rRNA hypervariable region^{22,23}, suitable data analysis software and database is crucial (Supplementary Data 1). An inappropriate choice of DNA extraction kit might result in inefficient lysis of the cell walls of Gram-positive (Gram+) bacteria, which have a more rigid cell wall structure, leading to

underrepresentation of these species^{24–27}. Indeed, technical variability among studies is frequently attributed to differences in DNA isolation methods^{28,29}. Similarly, the selection of library preparation protocols has been demonstrated to be important, given the notable differences in taxonomic accuracy produced by these methods^{12,30,31}. The influence of sequencing techniques² and bioinformatics strategies on result variability is rarely discussed in the scientific literature. However, choosing the most suitable bioinformatics techniques and databases is essential, as they significantly impact the results on community composition^{32,33}. While there is a wide range of dry-lab methods available, no single workflow has been shown to be ideal for all types of sequencing data. The two major sequencing methods are short-read sequencing (SRS) and long-read sequencing (LRS)³⁴. Traditionally, metagenomic whole genome sequencing (mWGS) is conducted on SRS platforms, accompanied by popular bioinformatic tools like Kraken²^{35–37}, sourmash^{38,39}, and MEGAN^{40,41}. A recent benchmark, using sequencing datasets of different mock microbial communities, showed that sourmash is the only tool that was able to produce excellent accuracy and precision on both SRS and LRS data³³. In another study, where the authors compared different programs and databases for SRS 16S-Seq⁴², a conclusion was that Kraken2, while originally developed for the analysis of

¹Department of Medical Biology, Faculty of Medicine, University of Szeged, Szeged, Hungary. ²MTA-SZTE Lendület GeMiNI Research Group, University of Szeged, Szeged, Hungary. ³These authors contributed equally: Gábor Gulyás, Balázs Kakuk, Ákos Dörmő. ✉ e-mail: boldogkoi.zsolt@med.u-szeged.hu; tombacz.dora@med.u-szeged.hu

WGS reads, proved to be applicable to a non-16S rRNA database (RefSeq). The emergence of LRS platforms from Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) has allowed the sequencing of the full 16S rRNA gene, potentially enabling higher resolution of classification¹². However, there are limited bioinformatic tools for these platforms, with Emu and EPI2ME¹¹ being significant exceptions.

Recent findings show a strong similarity between the canine and the human gut microbiomes⁴³. Given the genetic uniformity within dog breeds and the ability to control their diet, dogs stand out as an ideal model for microbiome research, with findings potentially applicable to human contexts. In our work, we examined the fecal samples from a Hungarian breed known as Pumi. Aiming to comprehensively evaluate metagenomic workflows for gut microbiome profiling, we performed a series of tests comparing several techniques, including DNA isolation, library preparation, DNA sequencing, and bioinformatics approaches. We created a new versatile program, named minitax, designed to reduce variability in bioinformatics workflows and provide uniform analysis across various sequencing platforms. In summary, the main objectives of our work are to evaluate existing wet-lab and dry-lab techniques, to attempt to identify the best practices for each stage of the process to ensure reliable and consistent gut microbiome profiling, and to develop a universally applicable bioinformatic tool.

Results

Study design

The objective of this study was to evaluate how various protocols influence the results regarding the composition of the gut microbiome. These protocols cover DNA isolation, library preparation, sequencing, and bioinformatics techniques (Figs. 1, 2). We evaluated the quantity, quality, and reproducibility of DNA extraction using four different isolation kits from Qiagen (Q), Macherey-Nagel (MN), Invitrogen (I), and Zymo Research (Z) (Fig. 3 and Supplementary Data 2a). We also characterized the microbial composition, examining predominant taxa and overall diversity. Libraries were prepared for mWGS using the DNA Prep Kit from Illumina (I WGS) and for amplicon sequencing of V-regions V1-V3 [PerkinElmer (PE V1-V3)], V1-V2, and V3-V4 [both from Zymo Research (Z V1-V2 and Z V3-V4, respectively)] of the 16S rRNA gene (Supplementary Data 2b). We also prepared V1-V9 libraries spanning the entire 16S rRNA region for sequencing on two LRS platforms: ONT MinION (ONT V1-V9) and PacBio Sequel IIe (PacBio V1-V9). Libraries were assessed for quality, volume, and consistency.

We employed the following bioinformatics approaches: DADA2⁴⁴ for amplicon-based SRS datasets, sourmash⁴⁵ for WGS samples, and Emu, a recently published, highly accurate software, optimized for LRS 16S rRNA-Seq⁴⁶. We also utilized the ONT's company-specific pipeline, EPI2ME⁴⁷ for processing the nanopore data. As part of this study, we developed a versatile, universally applicable program termed "minitax" to process various datasets, offering a robust tool for taxonomic assignment across all types of metagenomic data.

Following the initial alignment of sequencing reads to either a reference genome collection (by default) or to a 16S database using the minimap2 aligner⁴⁸, minitax identifies the best alignment and determines the most probable taxonomy for each read based on mapping qualities (MAPQ values) and CIGAR strings (Supplementary Data 2c).

To assess how library preparation methods, affect microbial composition accuracy, we used the same DNA isolation kit to standardize samples and eliminate potential inconsistencies. To compare and validate the results, different stages of the process were evaluated using a variety of samples and data types (Figs. 1, 2).

Comparison of DNA preparation techniques

Most DNA isolation kits employ affinity-based DNA purification, inhibitor removal buffers or columns, and lysis buffers, enzymes, or bead-beating for cell wall disruption. Although several studies advocate bead-beating for this purpose^{24,26}, many widely-used commercial kits do not include this step. We

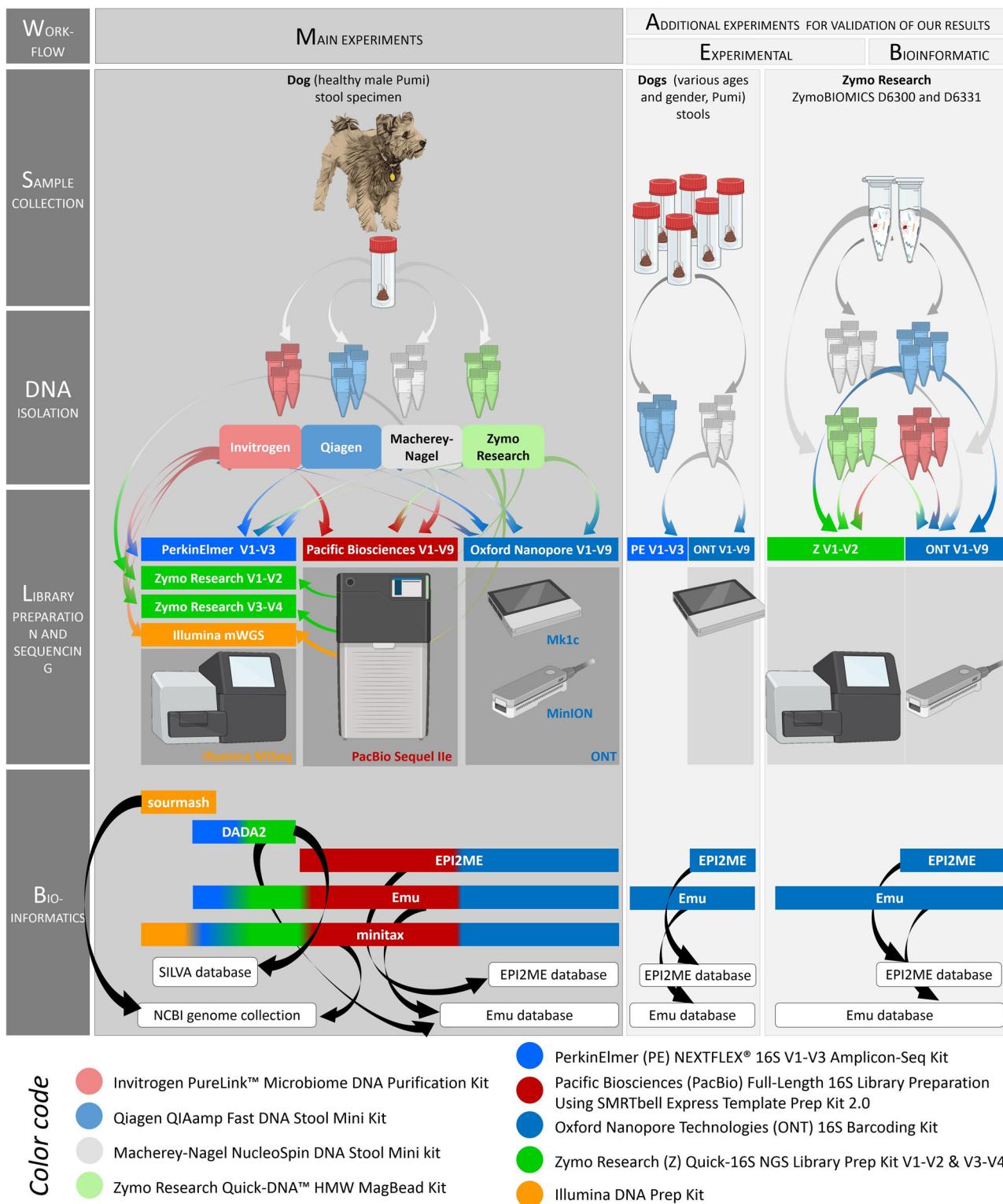
assessed four commercial DNA isolation kits differing in the aforementioned techniques (Supplementary Data 2a). Among the four methodologies scrutinized (Fig. 3), the Z approach necessitated the most extensive hands-on time, distinguishing it from the others in terms of labor intensity (Supplementary Fig. 1). The remaining three methods exhibited parity in this respect. The DNA extraction kits used in this study showed significant differences in both the quantity and quality of the extracted nucleic acid (Supplementary Data 3), as evidenced by variations in yield (Fig. 4a, b, Supplementary Fig. 2), microbe-to-host ratio (Fig. 4c, d), and reproducibility (Fig. 4e, f).

We observed substantial variation in the degradation rates across the different kits, with the Q kit yielding the most degraded DNA (Fig. 4a—lanes B1, E1). We utilized the Quick-DNA High Molecular Weight (HMW) MagBead Kit from Zymo Research to extract high-quality DNA samples necessary for sequencing the complete 16S rRNA gene using the ONT method (high fragment length of DNA is indicative of low degradation). However, our findings suggest that HMW DNA is not essential for sequencing the entire V1-V9 region. The DNA quality and quantity obtained using the MN (Fig. 4a—lanes F1, A2) and I (Fig. 4a - lanes B2, E2) kits were also suitable for LRS, as the variation in average DNA fragment length produced by these three kits did not significantly impact V1-V9 sequencing (Fig. 4a—lanes F1, E2 and Fig. 4b). However, it is important to note that for LRS-based WGS, maintaining HMW DNA integrity is essential. Based on the obtained DNA length, we found that the Z kit is a more suitable option for this purpose (Supplementary Data 3). We evaluated the DNA quality and the kit's efficiency based on the obtained fragment length and its ability to selectively extract bacterial DNA. Our results showed that, even when using the '*Isolation of DNA from Stool for Pathogen Detection*' protocol with the Q kit, the ratio of host DNA was significantly higher compared to the other three isolation kits (Fig. 4c, d).

Among the four DNA extraction kits we tested on canine stool samples, all but the Q kit produced adequate quantity and quality of DNA for sequencing (Fig. 4e, f and Supplementary Data 3a). Assessment of the Q kit using stool samples from six different dogs consistently resulted in below-average yields and quality (Supplementary Data 3b), whereas the other three kits reliably delivered very similar quality for each replicate. In our analysis, the Z kit demonstrated the most consistent results with minimal variation in concentrations among the replicates, whereas I exhibited the highest variance (Fig. 4e, f). This work identified substantial differences in DNA yields across the various kits (Supplementary Data 3). For canine fecal samples, the Q kit produced the lowest DNA yield. In contrast, the I kit yielded a moderate amount, while the MN kit provided the highest yield. Notably, despite starting with only half the initial sample volume, the Z kit achieved a relative high yield (Table 1 and Fig. 4e, f). Analysis of Variance (ANOVA) evaluation revealed significant differences among the kits, with an F-value of 511.63 and a p-value of less than 0.0001. Further comparisons showed that the I kit produced more DNA than the Q kit, while the MN kit outperformed both. The Z kit also yielded more DNA than the Q kit, but the MN kit still had the highest yield overall. Adjusting for sample volume, the Z kit's yield would nearly double, indicating that with equal sample sizes, the MN and Z kits would likely produce similar yields.

For canine fecal samples, the Z method generated the longest average peak fragment length, followed by the I method, which had a moderately long peak length. The MN kit yielded a slightly shorter peak length, while the Q kit resulted in the shortest average peak fragment length (Supplementary Data 3 and Table 1). A Tukey HSD test was performed to assess differences in DNA fragment lengths among various DNA extraction kits (Supplementary Data 3). The findings, depicted in Supplementary Fig. 3a, reveal significant differences between several kit pairs (Supplementary Data 3c). The distribution of fragment lengths across different kits for dog samples, depicted in Supplementary Fig. 3b, is shown through a violin plot. This plot uses color to represent varying concentrations and illustrates how fragment lengths differ by kit and correlate with concentration levels.

To test how significantly the DNA isolation kits influences the variation of the observed microbial composition, we used the same database (db)



and bioinformatic approach. Given that commonly used pipelines are applicable for particular sequencing methods, we used the minitax software, with a genome collection from NCBI as reference.

Previous studies demonstrated that α -diversity indices could successfully assess the efficiency of DNA extraction¹⁷. Here, we observed a significant and consistent reduction in both richness (as reflected by the Shannon index) and evenness (as reflected by the Simpson index) using the Q method compared to the other three methods (p -values < 0.000 for every comparison, using ANOVA) (Fig. 4g, h). Other pairwise richness

comparisons were not significant. However, the I kit showed lower evenness than the Z kit (Supplementary Data 4). We established a read number threshold for each sequencing approach where the α -diversity values remained stable and did not exhibit any significant reduction (Supplementary Fig. 4).

For a comprehensive evaluation of β -diversity, a sample-wise Bray-Curtis distance matrix was calculated. Subsequent analyses, including Non-Metric Multi-Dimensional Scaling (NMDS), Permutational Analysis of Dispersion (PERMDISP), Permutational Analysis of Variance

Fig. 1 | Workflow of the experimental part of the study. The figure provides a detailed representation of the workflows conducted in this study. The figure illustrates the workflows conducted in this study, which included: 1) evaluating the efficacy of various DNA isolation kits in terms of quality, quantity, and microbial representation from canine stool samples; 2) comparing library preparation techniques on SRS and LRS platforms for reproducibility; 3) introducing “minitax”, a tool designed to ensure consistent analysis across multiple sequencing platforms; 4) assessing the influence of different databases and tools on microbial profiling; and 5) comparing 16S V1-V9 sequencing on ONT and PacBio platforms to address literature gaps and emphasize bioinformatics workflows. Our goal was to identify reliable procedures for robust and reproducible gut microbiome profiling across both wet-lab and dry-lab methodologies. We performed additional experiments to

validate the most extreme experimental and bioinformatics results, particularly focusing on methods that yielded the most inconsistent outcomes in comparison to other techniques. For this purpose, we utilized samples from six additional dogs of various genders and ages. The workflow involved: 1) DNA isolation using the Q kit, 2) Library preparation with the PerkinElmer V1-V3 kit, and 3) Analysis of V1-V9 libraries using the EPI2ME software. Furthermore, we carried out experiments employing a Microbial Community Standard (MCS; Zymo Research D6300) as well as a Gut Microbiome Standard (GMS; Zymo Research D6331) to validate the effectiveness of the four DNA isolation kits used. This included: 1) DNA isolation using the kits applied for the dog samples, 2) Preparation of V1-V2 and V1-V9 libraries, and 3) Sequencing on the corresponding Illumina and ONT platforms based on the library. Created in BioRender. BioRender.com/k32q619.

(PERMANOVA), were conducted using these distance values. This analysis revealed distinct compositional differences attributed to the choice of DNA isolation method. The full-model PERMANOVA revealed that the DNA isolation had a substantial impact across various sequencing methods (WGS or 16S rRNA-Seq) and platforms, accounting for 28.2% of the overall variation observed in the microbial communities ($p < 0.001$).

An examination of sample dispersions for each DNA isolation method across all libraries, represented by the mean distance from the centroid in multivariate space (as estimated by PERMDISP2), revealed discernible variability. Such differences in dispersion can significantly influence the outcome of community composition analysis. These differences and similarities are graphically depicted in the Principal Coordinate Analyses (PCoA) for each isolation method (Fig. 4*i*). This analysis shows a large overlap between the I and MN methods, a partial overlap of the Z approach with the aforementioned two, and that the Q method clusters farther from all three.

Our dispersion analyses, conducted separately for each library, disclosed that the I approach consistently exhibits both the lowest average dispersions and the smallest standard deviations, with a mean distance to the centroid of 0.0626 and a standard deviation of 0.0586 (Supplementary Fig. 5a). The MN kit also yields low dispersions across almost every library, except the PE V1-V3. Compared to these two methods, the Z kit displayed results with increased dispersion in the LRS libraries and comparable dispersions in the SRS-based methods. The variability of the results obtained using Q kit was comparable to other methods in the two libraries suitable for in-depth analysis (PE V1-V3 and I WGS). Subsequent PERMANOVA evaluations for each library preparation method highlighted the influence of the DNA isolation techniques across the different V-regions and also in WGS. The results showed that the MN and I DNA isolation methods consistently produced similar microbial compositions across various library preparation protocols. In contrast, the Z and Q methods yielded compositions that significantly diverged from those of the MN and I methods, and often from each other as well. These are graphically depicted in the PCoA for each library calculated from the library-wise PERMDISP results (Supplementary Fig. 5b). The pairwise comparisons further demonstrated these differences, with significant variance across the entire V-regions. Notably, the only exception to this trend was the I/M pair.

Assessment of library construction and sequencing methods

In this part of the study, we aimed to elucidate the consistency in terms of variability among the six library preparation methods. To achieve this, we compared the results of each DNA isolation technique (I, MN, Q, and Z), for each sequencing library. Due to the quality and quantity issues we encountered with the Q kit, only the WGS, PerkinElmer (PE) V1-V3, and ONT V1-V9 libraries were prepared from these samples. Since the commonly used programs are typically designed for either amplicon or WGS methods, we chose to use the same program (minitax) for all sequencing methods, utilizing a genome collection from NCBI as a reference database.

Quality, yield, and reproducibility. The ONT 16S rRNA amplicon library preparation demands the least experimental effort, whereas the Illumina WGS and PacBio 16S rRNA amplicon library preparation necessitate the most hands-on time (Supplementary Fig. 6). Apart from

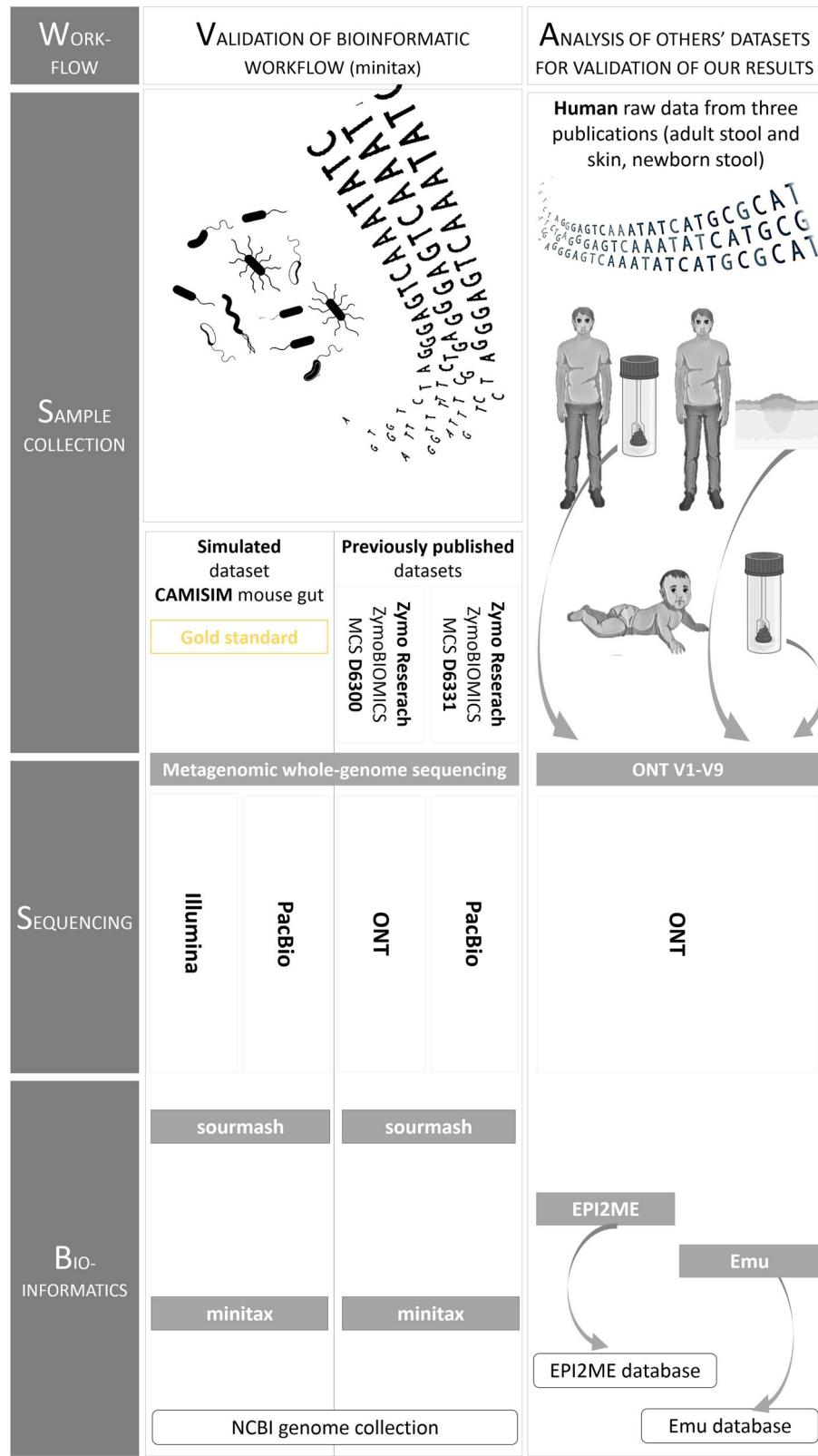
the PE method, which often resulted in two nonspecific DNA fragments (400 and 800 bps), the other kits demonstrated excellent performance in terms of quality, yield, and consistency for dog samples (Supplementary Fig. 7). Yet, Nanopore technology falls somewhat behind in terms of read quality (Supplementary Data 5).

Microbial composition, diversity and dispersion. The α -diversity analysis showed that in the SRS libraries, both richness and evenness measures were comparable among the three kits, with the exception of Q. In the LRS libraries, I method showed somewhat lower diversity, Z method the highest, while MN approach also exhibited similarly high values. We conducted β -diversity analyses using sample-wise Bray-Curtis distances. The NMDS plot depicted in Fig. 5a illustrates that the samples were more inclined to cluster according to their library preparation protocols than the DNA isolation methods employed. There is noticeable spread in samples extracted using the I and MN DNA isolation methods, indicating variability in community compositions according to the library preparation protocol. Interestingly, the V3-V4 group clustered far from the V1-V3 group, which was evidently more similar to the V1-V2 group. The Z method showed somewhat closer grouping, while the WGS and V1-V3 samples from the Q kit cluster together closely. Overall, the combined influence of both DNA isolation methods and sequencing protocols on community profiles is clear, with sequencing protocols appearing to be the more influential factor. Further evaluation using a full-model PERMANOVA, highlighted that indeed library selection was as the dominant factor in determining microbial community structures, accounting for 58.8% of the total observed variation. In addition to PERMANOVA, we also carried out PERMDISP analysis for each group (Fig. 5b). The PCoA visualization indicated that no DNA isolation method produced the same microbial compositions across the various libraries. In fact, only the V1-V2 and V1-V3 libraries showed considerable overlaps in the case of I and MN approach (Fig. 5b-d). Additionally, the LRS and WGS libraries produced from Z DNA displayed results similar to those generated by I and MN methods. Consistency within library preparation varies, as indicated by the spread of samples: the tightest clustering was shown by the V3-V4 method of the I samples, suggesting more consistent results.

A heatmap from the β -diversity analysis depicted in Fig. 5c shows the pairwise comparisons for each combination of DNA isolation and library preparation method. Irrespective of the isolation kit used, the PacBio results were found to be similar for both V1–V2 and V3–V4 regions (with the latter two being most alike). For the V3–V4 and V1–V2 regions, as well as for PacBio, it became apparent that the I and MN methods are more similar to each other than to the Z method. When comparing the libraries against themselves, it is evident that V1–V3 performs the least favorably.

Since the overlaps in Fig. 5b were only partial and a noticeable difference in variability was observed across the different libraries, we assessed the distances of each sample from the centroid of its respective group. We observed significant variability in the samples prepared using each DNA extraction method in the V1–V3 libraries. With the exception of the V1–V3 and LRS libraries prepared using Z DNA, all other library preparation methods produced results with relatively low variability (Supplementary Fig. 5c).

Fig. 2 | Workflow of the in silico analyses. Additionally, we conducted an in silico analysis to validate our bioinformatic tool, minitax. We compared the performance of minitax and sourmash by utilizing them for the analysis of: 1) simulated PacBio and Illumina data⁴⁴, and 2) previously published datasets³³. For further validation of our results, we also used previously published metagenomics data from human sources encompassing skin⁵⁷ as well as fecal samples from newborns⁵⁴ and adults⁵⁶. Created in BioRender. BioRender.com/e57v119.



Pairwise PERMANOVA, with library preparation technique as the sole variable, unambiguously demonstrated that all libraries showed significant differences from each other, even after correcting the p-values for multiple testing. The overlaps observed in the PCoA plot (Supplementary Fig. 5b) were determined to be non-significant according to the PERMANOVA, which assesses differences between group centroids. In the case of the V1-V3

library, however, detected differences are not solely attributed to variations between centroids, as they are also influenced by differences in group dispersions.

The microbial community composition analysis for the I samples showed that the V1-V3 library had the highest proportion of *Bacillota* (formerly *Firmicutes*), while *Bacteroidota* (formerly *Bacteroidetes*) and

Fig. 3 | The kits and sequencers utilized in this project. This figure lists the DNA extraction and library preparation kits, along with sequencing devices which were used in this experiment.



Fusobacteriota (formerly *Fusobacteria*) were present in much smaller proportions. In contrast, the V1-V2 and V3-V4 libraries displayed a more balanced distribution with *Bacteroidota* and *Fusobacteriota* (Fig. 5d). We compared our results and data with those of other studies^{43,49–53} (Supplementary Data 6 and Fig. 6), all of which focus on a single aspect of the canine gut microbiome. A unanimous finding across all these publications, including our own dataset, suggests the predominant presence of five phyla. However, their abundance varies significantly across approaches (Fig. 6), underscoring the impact the methods have on the obtained results.

Difference in ratio of Gram-positive and Gram-negative bacteria. To test whether certain DNA extraction method favor Gram-negative (Gram-) bacteria over Gram-positives, we aggregated species abundances based on their cell-wall staining characteristics and conducted a similar analysis. We found that, indeed, the observed differences among DNA isolation methods could be significantly attributed to the varying resistance of cell walls to treatments. Specifically, only the I and MN methods produced microbial profiles with near-identical ratios of Gram+ to Gram- bacteria. In stark contrast, all other pairwise comparisons highlighted pronounced disparities in these ratios. Concerning the differences among libraries, we observed significant disparities in multiple pairwise comparisons, with the V3-V4 region often being particularly notable (Supplementary Fig. 8).

Comparison of DNA isolation methods and sequencing libraries on synthetic microbial community standards

To evaluate the accuracy of our DNA isolation methods and library preparation protocols, we employed two microbial community mixtures [ZymoBIOMICS Microbial Community Standard D6300 (MCS) and ZymoBIOMICS Gut Microbiome Standard D6331 (GMS), both from Zymo Research] with defined compositions as reference standards (Fig. 7, Supplementary Fig. 9a, and Supplementary Data 7). We generated Illumina V1-2 and ONT V1-9 libraries from the MCS community, along with ONT V1-9 libraries from the GMS community, using DNA samples isolated by each of the four methods. The MCS is a mix of eight *Bacteria*, of which five belong to the *Firmicutes* phylum and three are *Proteobacteria*. The GMS, comprising 18 *Bacteria* and one *Archaea*, offers a more diverse mix than the MCS, making it a more accurate representation of the human gut microbiome. Although it was shown by Matsuo et al. that the V1-V9 primers had a bias against *Bifidobacterium*⁵⁴, we decided to calculate the statistics without removing this genus, as neither the manufacturer of the community mixture (Zymo), neither that of the primers (ONT) advise the user to do so. In

contrast to the fecal DNA sample analysis, where we assessed the uniformity of results within each group of methods (extraction, library preparation, and bioinformatics) and their similarity to one another, our main objective here was to evaluate each methods' ability to reproduce the expected composition of the microbial community. We used the minitax and the Emu software and with the NCBI genome collection and the Emu database to compare the observed microbial compositions with the expected ones.

Quality and yield of DNA. In our experiment, the DNA yields varied significantly across the different techniques for both MCS and GMS samples (Table 1). For MCS, the Z kit achieved the highest average yield, surpassing all others, including the MN kit, which also produced a notably higher yield than the Q and I kits. The ANOVA results confirmed these differences, with an F-value of 124.75 and a p-value of less than 0.0001. Post-hoc analysis further revealed that the Z approach yielded significantly more DNA than the Q and I methods, and it outperformed the MN kit as well. For GMS samples, the MN method delivered the highest average yield, although the Z technique provided a substantial yield that was only slightly lower. The Q kit produced the highest yield among the others, while the I kit yielded the least. ANOVA results for GMS samples also indicated significant differences among the techniques, with an F-value of 147.92 and a p-value of less than 0.0001. The HSD analysis confirmed that the MN method produced significantly more DNA than the I and Q approaches, though the Z kit's yield was only slightly less than that of the MN kit. Overall, we conclude that both the Z and MN kits demonstrated superior performance across both sample types, especially in comparison to the Q and I kits (Table 1).

Our fragment length analysis revealed significant differences in the degradation rates produced by the various DNA isolation techniques. In MCS samples, the Z, I, and Q methods produced the longest average fragment lengths, each exceeding 60,000 bps, while the MN technique had a slightly shorter average peak length of 54,740 bps (Table 1). For GMS samples, the Z and Q techniques again delivered the longest average fragment lengths, both over 60,000 bps. The I kit followed closely with an average peak length of 58,897 bps, and the MN kit had an average peak length of 55,538 bps (Table 1). Our findings demonstrate that the applied extraction method and the sample type significantly impact the quality of the isolated DNA (Supplementary Data 7, Supplementary Fig. 9b).

The Tukey HSD test showed significant differences in longest fragment lengths between kit pairs MN/I, MN/Q, and MN/Z. These results highlight the impact of DNA extraction methods on fragment length quality from MCS, as detailed in Supplementary Fig. 9b.

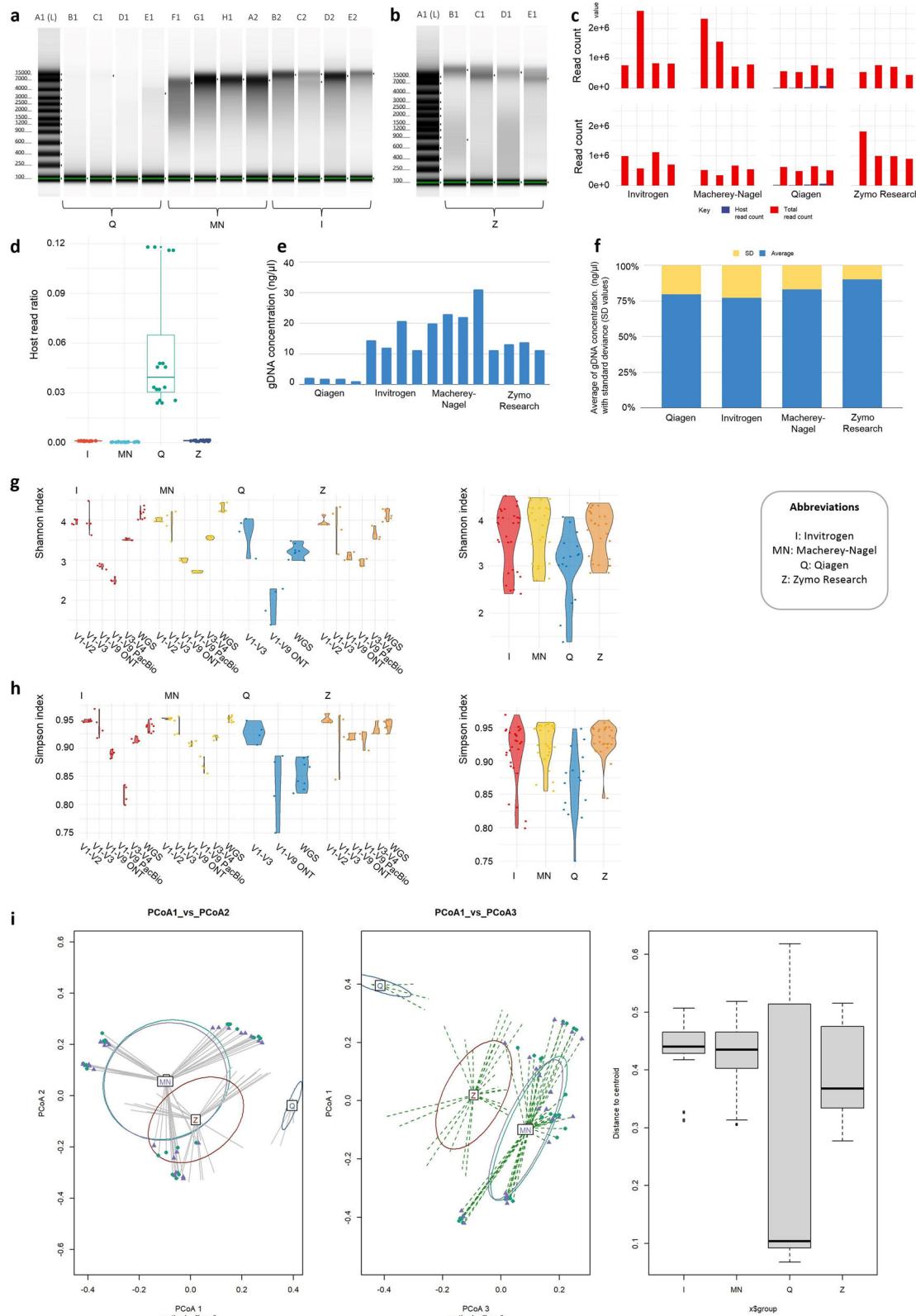


Fig. 4 | Impact of DNA isolation kit on yield, quality, and bacterial composition analysis. Quality of the Isolated DNA; molecular weight of DNA (a: lanes B1,E1, F1,A2 and B2,D2: four replicates of DNA samples were isolated using Q, MN and I kit, respectively. b: lanes B1,E1: DNA samples extracted by Z kit. c, d ratio of host DNA. e, f Yield and reproducibility of the isolated DNA. g-i Microbial composition,

diversity and dispersion; Alpha-diversities in the different library preparation methods, according to each DNA isolation method: g Shannon index; h Simpson index; Beta-diversity analysis of each DNA isolation method: i PCoA plots and distances to group centroids.

Table 1 | Average concentrations and lengths of obtained genomic DNA samples

The average of DNA concentration isolated from dog stool	Average gDNA cc. (ng/μL)
Qiagen	1.815 ± 0.464
Invitrogen	14.6 ± 4.35
Macherey-Nagel	24.0 ± 4.83
Zymo Research	12.4 ± 1.36 ng/μL
The average of DNA concentration isolated from the MCS sample	Average gDNA cc. (ng/μL)
Qiagen	0.364 ± 0.12
Invitrogen	0.287 ± 0.048
Macherey-Nagel	10.564 ± 0.91
Zymo Research	38.72 ± 3.33
The average of DNA concentration isolated from the GMS sample	Average gDNA cc. (ng/μL)
Qiagen	1.146 ± 0.157
Invitrogen	0.0683 ± 0.0076
Macherey-Nagel	12.25 ± 0.636
Zymo Research	4.087 ± 0.31
The average of the longest DNA fragments isolated from dog stool	Average length of DNA (bp)
Qiagen	9361
Invitrogen	26,341
Macherey-Nagel	21,325
Zymo Research	33,867
The average of the longest DNA fragments isolated from the MCS sample	Average length of DNA (bp)
Qiagen	>60,000
Invitrogen	>60,000
Macherey-Nagel	54,740
Zymo Research	>60000
The average of the longest DNA fragments isolated from the GMS sample	Average length of DNA (bp)
Qiagen	>60,000
Invitrogen	58,897
Macherey-Nagel	55,538
Zymo Research	>60,000

The first sections present the average concentrations of genomic DNA (gDNA) obtained from dog stool, MCS, and GMS samples, measured in ng/μL, respectively. Values are reported as means with standard deviations. The last blocks of tables report the average length of the longest DNA fragments (in base pairs, bp) obtained from dog stool, MCS, and GMS samples, respectively.

Microbial composition, diversity, and dispersion

MCS. The dispersion analysis showed that for the SRS library, in alignment with our observations from fecal DNA samples, the I method exhibited the lowest dispersion. It is noteworthy that in the LRS library, the Q method demonstrated similarly low variability. Upon examining the observed microbial compositions, we found that unlike the strong similarity seen in vivo, the MN and I methods showed significant divergence when evaluating the MCS. Although the Q method remained the most distinct from I approach, Z and MN techniques exhibited a closer association with each other and a notable separation from I method. This trend was evident across both library types, and also across the two programs and two databases utilized (minitax with Emu db, Emu with Emu db, and minitax with NCBI genome collection, Fig. 7a, b and Supplementary Figs. 10a, b and 11a, b, respectively) underscoring the intrinsic differences between in vivo and in vitro sample evaluations. The data obtained using the I kit showed the highest concordance with the theoretical values: no bacterial species were

found to be significantly different using either bioinformatic tools or databases (with the exception of *Pseudomonas aeruginosa* using Emu with the EMU db). Regarding the other extraction methods, they showed varying number of significant differences ranging from 1–5, depending on wetlab and bioinformatic factors.

GMS. A different pattern emerged when examining the more complex microbial mixture. The Z method identified the fewest significant deviations from the theoretical composition (7–8 species), while the I method identified the highest deviation (11 species). The other two methods fell in between, identifying 8–10 species whose ratios significantly differed from the defined composition. This trend was consistent across both bioinformatics pipelines and databases evaluated. The results obtained using the Z and MN approaches were more similar to each other than to those obtained by I, while the results from Q differed the most from the other three. This was evident in the results from minitax with the Emu db, Emu with the Emu db, and minitax with the NCBI genome collection (Fig. 7c, d and Supplementary Fig. 10c, d and 11c, d, respectively).

The I method, although performing well with the MCS, showed different performance with the more complex GMS, indicating that sample-specific factors and method robustness play significant roles. Intriguingly, I and MN performed similarly with the complex in vivo sample (dog feces), suggesting context-specific performance. Additionally, DNA concentration was very low for the GMS I method (Supplementary Data 7b) and for MCS using both Q and I, whereas only the Q method yielded suboptimal quantities in the in vivo sample. Therefore, DNA concentration results should be carefully considered, as low yields can affect downstream applications and analyses. Methods yielding higher DNA concentrations might be more suitable for samples with higher complexity. Overall, neither kit consistently correlated strongly with theoretical values across different sample complexities. Portik and colleagues also concluded that MCS (D6300) outperforms GMS (D6331)³³, suggesting that these microbial communities may not always be reliable indicators for selecting the most suitable DNA isolation kit or sequencing/library preparation method for each in vivo sample. Thorough testing is advisable before embarking on a large-scale study, especially if the sample source is less well-known.

Difference in ratio of Gram(+) and Gram(-) bacteria. In this part of the study, we obtained that the Q method tended to overrepresent Gram-bacteria in MCS samples and produced higher ratios for these bacteria in in vivo samples. Intriguingly, while the MN method displayed a similar overrepresentation in fecal samples, this trend was not observed in the MCS. For both kits, the main underrepresented Gram+ bacterial orders were *Lactobacillus/Limosilactobacillus*, whereas the majority of the overrepresented Gram- orders were *Escherichia* and *Salmonella*. We suggest that the MN method favors extracting DNA from Gram- bacteria, as evident in the controlled setting of MCS. This bias could be obscured in fecal samples due to their complexity, affecting the lysis of both Gram+ and Gram- bacteria. In the GMS samples, the I method underrepresented the Gram- bacteria, while MN, Q, and Z methods yielded a similar abundance, close to the theoretical values.

Laboratory work: key findings

Our analysis of various DNA extraction kits and library preparation methods revealed several key insights (Table 2 and Table 3). The Z kit provided a very high DNA yield and the longest fragment lengths, making it generally superior for high-quality microbial diversity analysis overall. The MN method demonstrated high overall DNA yield and good alignment with theoretical compositions but had the shortest fragment lengths. The I protocol offered good quality and consistency but showed higher variability in DNA yield and fragment length compared to Z and MN. The Q kit, while suitable for specific applications, had lower yields and higher contamination, affecting microbial diversity representation (Supplementary Fig. 3b). In terms of library preparation, Illumina MiSeq with Illumina DNA Prep (WGS) was recognized for its high-quality data and accuracy, despite being

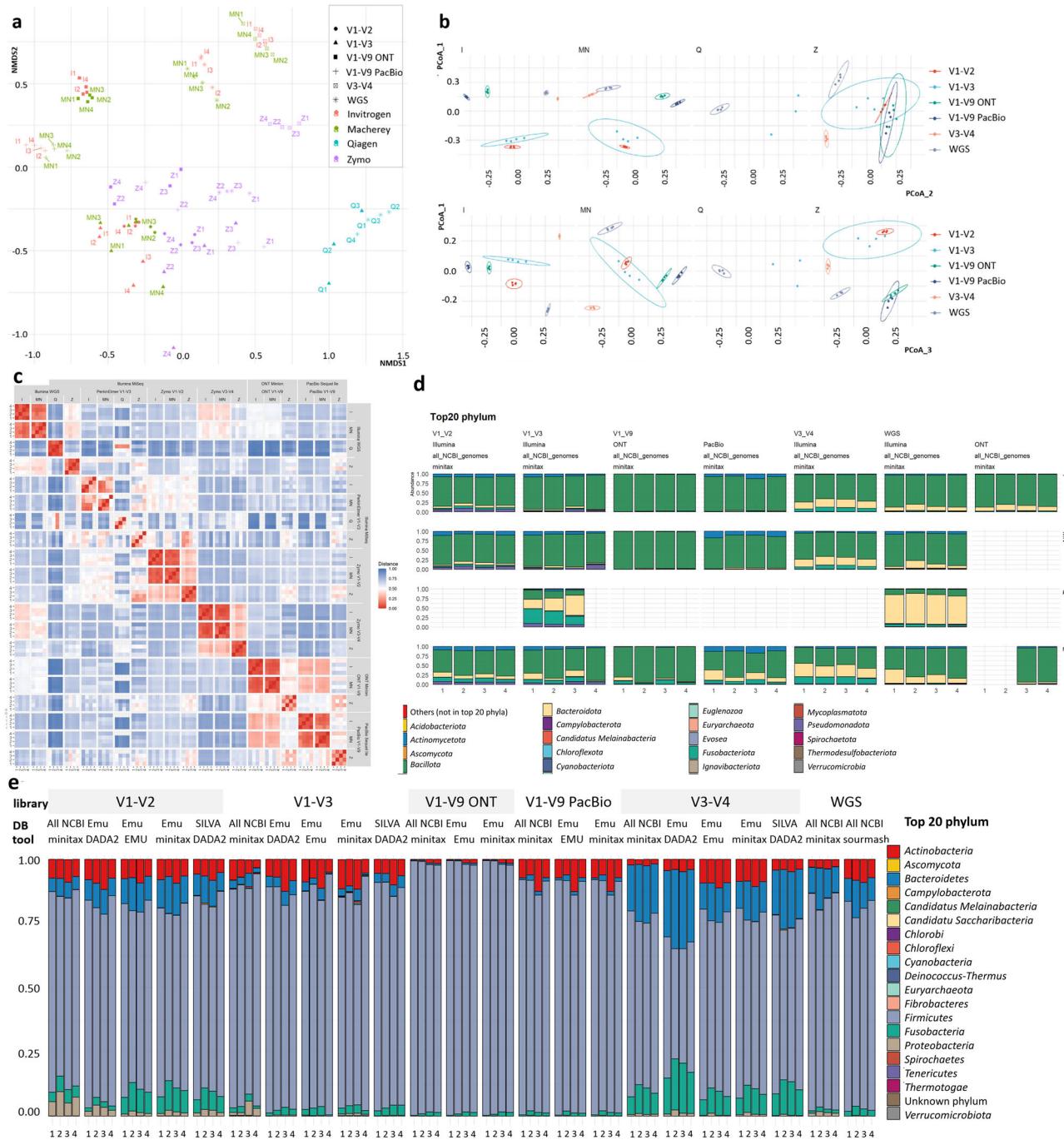


Fig. 5 | Comparison of library preparation protocols. **a** β -diversity analysis – NMDS. The sample-wise Bray-Curtis distances between the samples were calculated and plotted using NMDS. Colors show the DNA isolation methods, while shapes indicate the library preparation protocols. **b** β -diversity analysis—PERMDISP. PCoA visualization of the dispersion values from the PERMDISP results. The plot is faceted according to the DNA isolation methods and colors indicate the library preparation protocols. **c** β -diversity analysis—Heatmap. The sample-wise Bray-Curtis distances between the samples were calculated and visualized on a heatmap. The WGS data composition for the I and MN isolation kits is most similar to the bacterial composition from the Z V3-V4 library, followed by the ONT V1-V9 library. Notably, for both I and MN DNA isolation kits, the WGS and ONT V1-V9 libraries exhibit fairly similar compositional profiles. The Z DNA isolates produce relatively consistent microbial community profiles across different library preparation techniques. The PE V1-V3 library for Q DNA shows some similarity to WGS. The I and MN DNA samples exhibit significant overlap with the V1-V2

libraries prepared from I, MN, and Z DNA, with Z DNA showing a particularly high degree of overlap with itself. Additionally, Z DNA shows good overlap with the V1-V3 library compared to other libraries. The V3-V4 library consistently shows the least similarity to other libraries across all DNA isolation methods. The PacBio results are most similar to ONT, regardless of the DNA isolation kit used. When using Z DNA isolation method, PacBio shows significant similarity in microbial composition with other SRS amplicon-based sequencing results, particularly with V1-V2, V1-V3, and then V3-V4. WGS is significantly different. The Z-isolated DNA ONT V1-V9 libraries also show similar compositional similarities with these, with PacBio possibly showing a closer match. **d** Relative abundance of the top 20 phyla in each sample with. Rows indicate the DNA isolation methods, while columns indicate the library preparation protocols. **e** Barplot showing the top 20 phyla in the MCS samples, sequenced using Invitrogen DNA isolation kit, according to library preparation protocols, analyzed with minitax, DADA2, and Emu programs using several different databases.

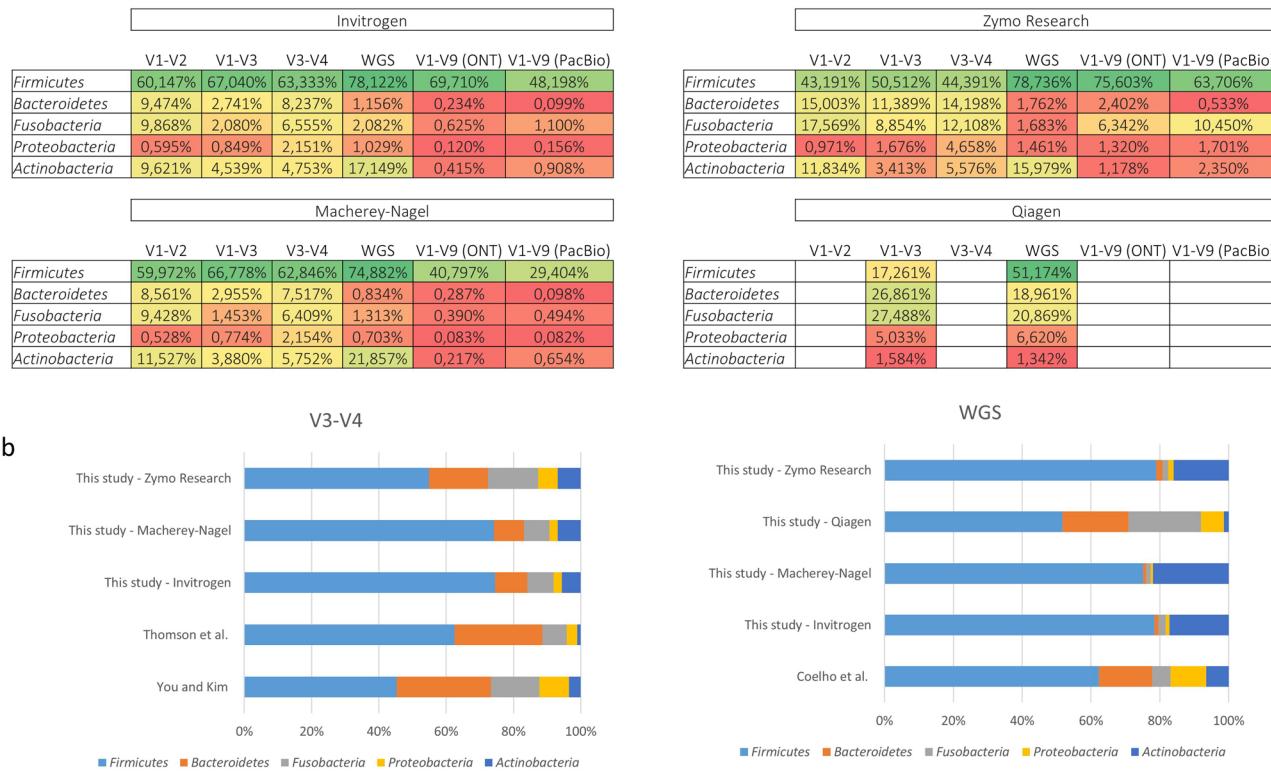
a

Fig. 6 | Comparison of the bacterial composition of various samples at the phylum level. This figure presents a comparison of bacterial composition at the Phylum level in various samples. **a** Tables show the proportions of the most abundant Phyla in samples obtained using different DNA extraction and library preparation

methods. **b** Bar charts highlight the discrepancies between the proportions of the most abundant bacterial Phyla from other groups' data and our datasets. Taxa not included in the defined composition are grouped together under the label "other".

labor-intensive and costly. Illumina MiSeq with PerkinElmer V1-V3 showed less consistency and produced erroneous fragments. ONT MinION with ONT V1-V9 offered good performance with less experimental work but had lower read quality compared to Illumina and PacBio. PacBio Sequel IIe with PacBio V1-V9 provided high-quality reads but was costly and labor-intensive.

Comparison of bioinformatics techniques

To eliminate the variability in canine microbiome composition caused by the different DNA extraction kits, we exclusively used the I DNA isolation method for comparing bioinformatic approaches. We chose the NCBI genome collection as a reference to reduce disparities arising from database selection (Fig. 5e).

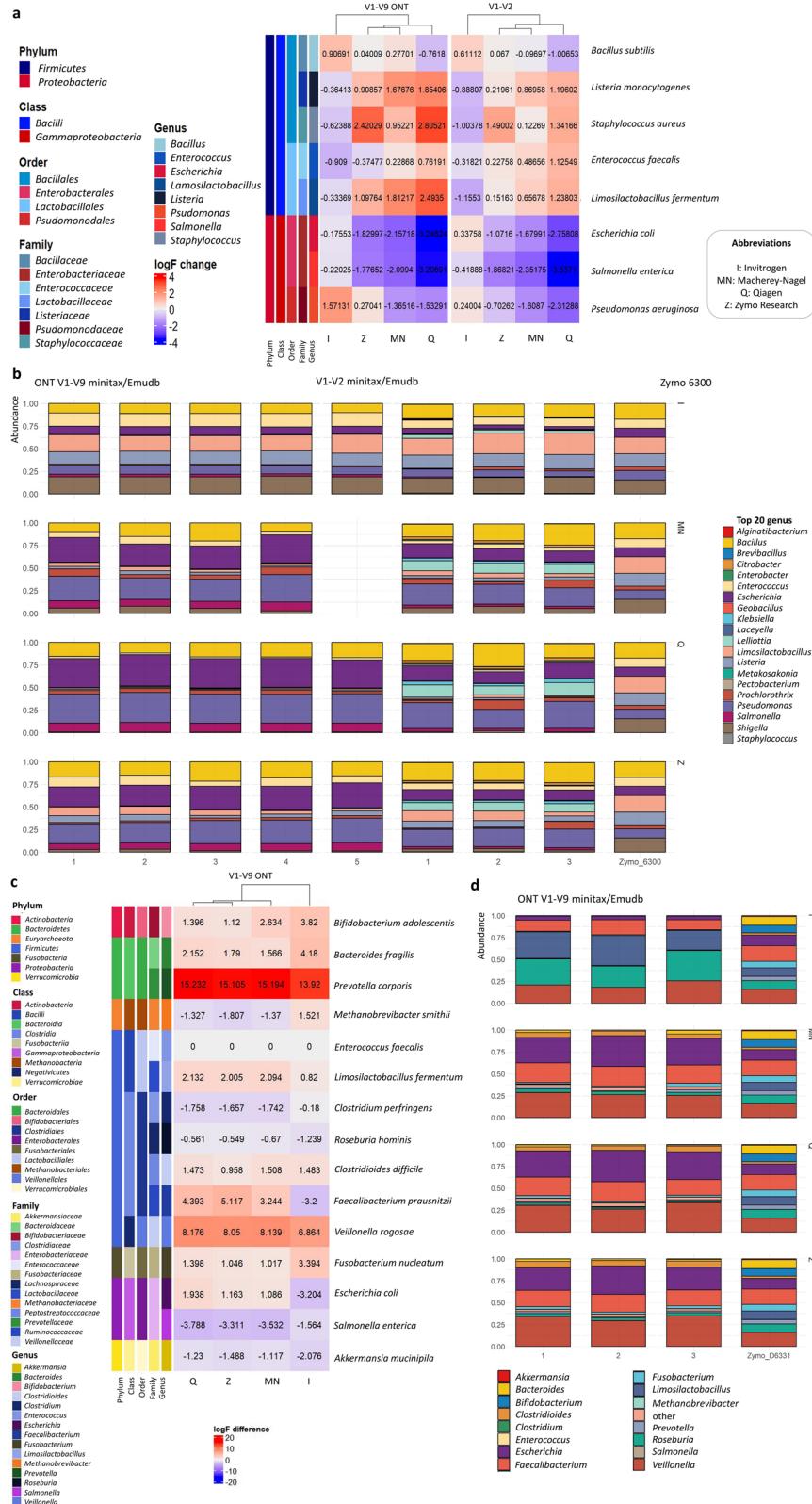
In this section of the study, we analyzed the canine microbiome compositions at the phylum, order, genus, and species levels, comparing the effects of different databases (Emu, NCBI genomes, and SILVA) and bioinformatic programs (Emu, minitax, DADA2, and sourmash). Using a full-model PERMANOVA, we found that bioinformatics approaches and databases significantly influence the outcome, accounting for 59.4% of the variation in microbial community composition, while the 'library' factor accounted for 20.1%, leaving 20.4% of the variation unexplained. Subsequent pairwise comparisons revealed that DADA2 showed notably significant differences when used with the Emu database compared to the other two programs, while minitax closely resembled Emu. Using DADA2 with the SILVA database highlighted even starker differences (Fig. 5e). While several significant differences (p -value < 0.05) were observed across different database and program combinations, the specific comparisons involving minitax with the NCBI genome collection showed significant differences only with minitax using the Emu database for the ONT V1-V9 library, and with sourmash using the NCBI genome collection for WGS. However, consistent results were seen across all library types, regardless of the

software/database pairing. Therefore, the minitax/NCBI genomes combination proved apt for evaluating DNA isolation kits and library types (Fig. 5e).

We also analyzed our MinION-sequenced samples using the ONT 16S Barcoding Kit along with the 16S module of the EPI2ME Labs software (EPI2ME Desktop, version 2021.09.09), which streamlines the analysis process and quickly generates a Sankey tree diagram on its online platform. We initially observed the *Blautia* genus to be the predominant component (Fig. 8a, b), with an average occurrence of about 80%. This result was consistent across all DNA isolation kits but contradicts both literature data and sequencing results from other methods (Illumina WGS, V1-V2, V3-V4, V1-V3, PacBio V1-V9). When we analyzed these samples using other programs (Emu and minitax), we found that the *Blautia* ratio aligns with the ~15% reported in the literature⁵⁰, thus EPI2ME significantly enriched *Blautia*. However, contrary to our initial assumptions, EPI2ME does not exaggerate the presence of *Blautia* but instead excludes some genera and underrepresents others. We analyzed the entirety of the ONT V1-V9 datasets available in the databases using EPI2ME and compared the outcomes with those derived from Emu. In our canine samples, the software failed to detect any reads corresponding to the *Peptacetobacter*, *Faecalimona*, and *Mediterraneibacter* genera, discarding 75% of the reads and skewing the actual compositional ratios. To verify the discrepancies in the genus composition in the EPI2ME results, and to determine if this distortion is conclusively due to bioinformatics, we extended our study beyond the original canine stool sample (Fig. 8a, b), incorporating fecal samples from six additional dogs (Fig. 8c), along with both neonatal⁵⁵ (Fig. 8d) and adult⁵⁴ (Fig. 8e) human stool specimens, adult human skin⁵⁶ (Fig. 8f), as well as MCS (Fig. 8g). We performed sequencing of ONT V1-V9 libraries and analyzed the data using both EPI2ME and Emu.

Next, we examined the six control dog samples and found that EPI2ME discards ~50% of the reads during processing (Fig. 8c). Similar to our

Fig. 7 | Comparative analysis of DNA isolation and library preparation protocols in MCS and GMS. **a** Heatmap of sample-wise differences in the MCS samples. The abundance values identified by the Emu application were compared to the theoretical values provided by Zymo and log₂ fold changes were estimated and are shown within the boxes. Deeper blue colors indicate lower experimental values compared to the theoretical, while more red colors indicate higher experimental values. **b** Barplots showing the top 20 phyla in MCS samples using the Illumina V1–V2 and ONT V1–V9 methods, according to the DNA isolation methods, analyzed with minitax using the Emu genome collection database. **c** Heatmap of sample-wise differences in the GMS. The abundance values identified by the Emu tool were compared to the theoretical values provided by Zymo Research, with log₂ fold changes calculated and displayed within the boxes. Darker blue colors represent lower experimental values compared to the theoretical, while dark red colors indicate higher experimental values. **d** Barplots show the top 20 phyla in GMS samples using the ONT V1–V9 methods, according to the DNA isolation methods, analyzed with minitax using the Emu database.



original dog sample, the genus *Peptacetobacter* was undetected, while the genera *Faecalimonas* and *Mediterraneibacter* were underrepresented. We used ONT V1–V9 data from European Nucleotide Archive (ENA) to compare EPI2ME and Emu. We found that for the fecal samples, EPI2ME did not analyze 15–25% of the data, and the genus *Escherichia* was entirely excluded (Fig. 8d, e). For skin samples, minimal variation was observed

between the two programs, with ~5% of the reads not analyzed by EPI2ME, mostly belonging to the genera *Cutibacterium* and *Staphylococcus* (Fig. 8f).

In the analysis of the microbial mix sourced from Zymo Research, we observed that EPI2ME did not include around 15% of the reads and failed to identify the genus *Escherichia*. Conversely, EPI2ME identified a genus (*Shigella*) that is not present in the mixture (Fig. 8g).

Table 2 | Summary of DNA isolation methods: yield, fragment quality, and suitability for various sample types

Kit	Advantages	Disadvantages
Qiagen	<ul style="list-style-type: none"> Suitable for certain applications 	<ul style="list-style-type: none"> Lowest DNA yield among the kits Higher levels of host DNA contamination Less accurate in representing microbial diversity
Invitrogen	<ul style="list-style-type: none"> Good overall quality High amount of DNA from stool samples Consistent results in various libraries and sequencing methods Closest match to theoretical composition in MCS samples 	<ul style="list-style-type: none"> Higher variability in yield and fragment length compared to Z and MN kits Less optimal performance with GMS samples Less effective in certain sample types (e.g., MCS and GMS) compared to Zymo and MN
Macherey-Nagel	<ul style="list-style-type: none"> High overall DNA yield Better performance in GMS samples Good alignment with theoretical microbial composition in MCS samples Comparable yield to Z kit when adjusted for initial volume 	<ul style="list-style-type: none"> Shorter average DNA fragment length
Zymo Research	<ul style="list-style-type: none"> High DNA yield in canine fecal and MCS samples, highest in GMS sample Provides high-quality DNA Longest average DNA fragment length Superior performance in MCS samples Closest match to theoretical composition in GMS samples Consistent results and performance for LRS and WGS libraries Stable performance with low relative dispersion 	<ul style="list-style-type: none"> Less accurate representation in MCS

This table details the advantages and disadvantages of various DNA isolation kits, including Zymo Research, Qiagen, Macherey-Nagel, and Invitrogen. Key factors include DNA yield, fragment length, contamination levels, and overall performance across different sample types.

Table 3 | Comparison of library preparation methods: performance metrics and practical considerations

Sequencing platform / library prep kit	Advantages	Disadvantages
Illumina MiSeq/Illumina DNA Prep (WGS)	<ul style="list-style-type: none"> Known for high-quality data and accuracy in determining true microbial compositions Shows high consistency across libraries High-quality libraries 	<ul style="list-style-type: none"> Labor-intensive and complex library preparation Highest hands-on time Generally more costly
Illumina MiSeq/PerkinElmer V1-V3	<ul style="list-style-type: none"> Widely used V-region 	<ul style="list-style-type: none"> The applied kit often produces nonspecific or erroneous fragments Shows the least similarity to other libraries Performance can be inconsistent across different samples
ONT MinION/ONT V1-V9	<ul style="list-style-type: none"> Requires less experimental work compared to other methods High-quality, high-consistency libraries Shows high similarity with PacBio V1-V9 results and moderate similarity to Illumina WGS data 	<ul style="list-style-type: none"> Generally lower read quality compared to Illumina and PacBio May not consistently match well with other libraries, particularly short-read amplicon libraries
PacBio Sequel IIe/PacBio V1-V9	<ul style="list-style-type: none"> Provides high-quality reads High-quality, high-consistency libraries Shows strong similarity with ONT results Shows good moderate similarity with V1-V2 and V1-V3 libraries 	<ul style="list-style-type: none"> Labor-intensive preparation Typically more costly
Illumina MiSeq/Zymo Research V1-V2	<ul style="list-style-type: none"> Similarity with PE V1-V3 data High-quality libraries High consistency, reliable data quality 	<ul style="list-style-type: none"> Moderate costs and labor intensity
Illumina MiSeq/Zymo Research V3-V4	<ul style="list-style-type: none"> Strong similarity with Illumina WGS (with I and MN DNA samples) High-quality libraries High consistency, reliable data quality 	<ul style="list-style-type: none"> Moderate costs and labor intensity

This table provides an overview of library preparation methods, including Illumina DNA Prep, PerkinElmer V1-V3, Zymo Research V1-V2 and V3-V4, ONT V1-V9, and PacBio V1-V9. It highlights the benefits and drawbacks related to sequencing quality, consistency, labor intensity, and cost for each preparation technique.

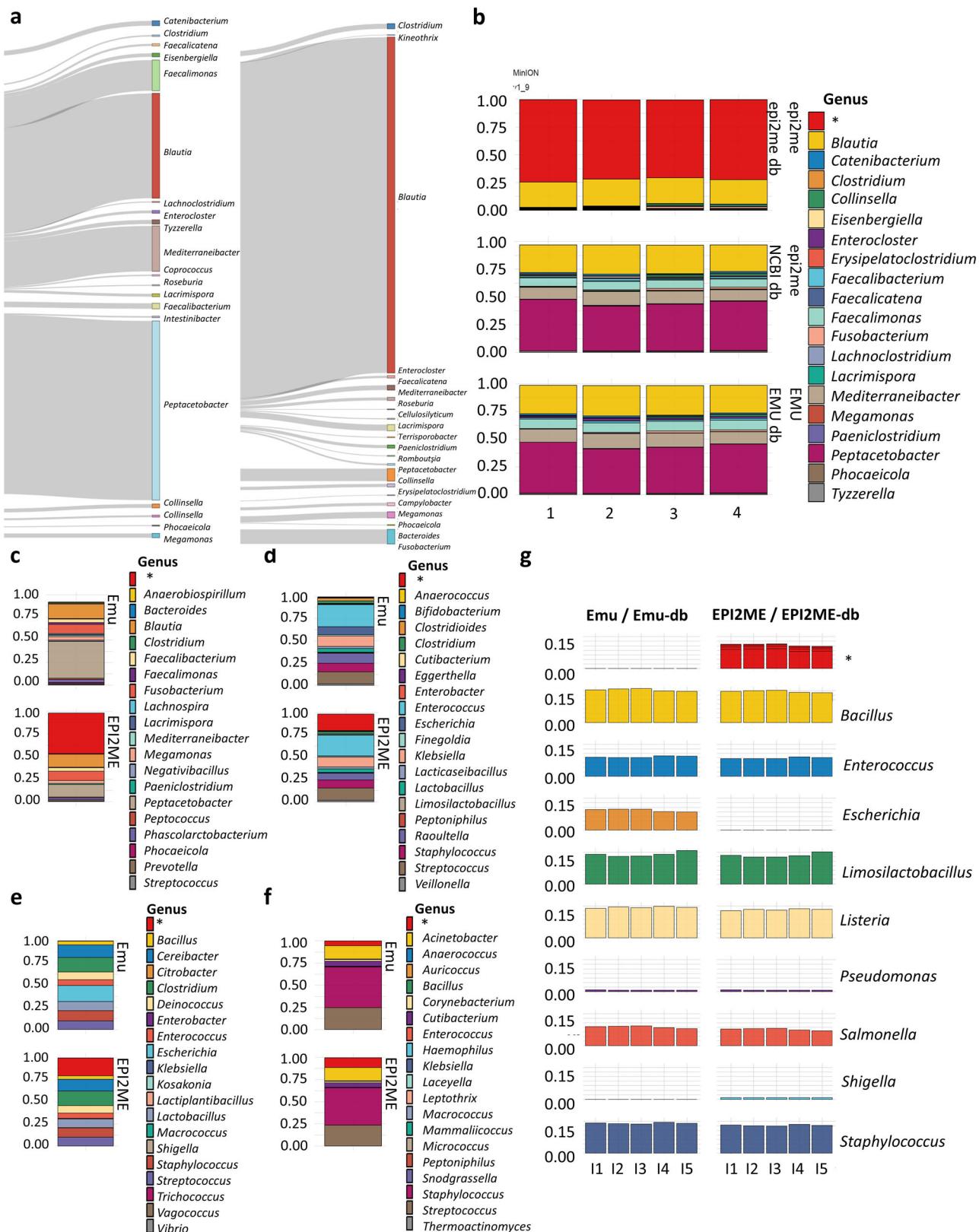


Fig. 8 | The EPI2ME analysis platform yields inaccurate genus-level data. Results from the 16S module of the EPI2ME Labs program by ONT employed on canine stool sample (**a,b**), fecal samples from six additional dogs (**c**), neonatal⁵⁴ (**d**) and adult⁵⁶ (**e**) human stool specimens, adult human skin specimen⁵⁷ (**f**), and MCS (**g**). In our canine samples, the software failed to yield abundances to the *Peptacetobacter*,

Faecalimonas, and *Mediterraneibacter* genera, discarding 75% of the reads and skewing the actual compositional ratios. * In (**b, c, d, e, f**, and **g**) of this figure the blocks highlighted in red represent the reads that the EPI2ME program filters out and excludes from the analysis due to the LCA tag.

We found that the EPI2ME software excluded a substantial fraction of data during its processing due to the application of Lowest Common Ancestor (LCA) tags. Specifically, within the BLAST analysis module, the LCA tag was designed to filter out reads when the top three predicted genera do not match the tag. We reprocessed the EPI2ME data and used the updated information to assign the reads to the NCBI database based on their taxon IDs. This revealed a similarity in the results produced by EPI2ME and Emu. Therefore, we can conclude that employing the LCA tag to filter out data in EPI2ME is unnecessary and, in fact, diminishes the accuracy of the program (Fig. 8b).

Evaluation of minitax: benchmarking across various sequencing methods and data types

To ensure consistent comparisons across varied sequencing data, we developed minitax, a tool designed for metagenome sequencing. We rigorously tested it across diverse platforms and datasets, comparing it to several bioinformatic programs based on their ability to accurately reconstruct the reference microbial composition, as indicated by the strength of the correlation (r^2 values) between the observed and theoretical compositions. Additionally, we used Chi-square statistics to assess whether there was a statistically significant difference between the observed and theoretical distributions.

Comparing minitax with Emu using ONT V1-V9 sequencing of MCS and GMS. We contrasted the performance of minitax with Emu, which, like minitax, utilizes minimap2 for initial read alignment but employs an expectation-maximization approach subsequently. When comparing the workflows on the MCS, minitax and Emu delivered similar results when utilizing the Emu db, indicating the robustness of minitax (Fig. 9a and Supplementary Fig. 12a). Notably, even though complete genome databases are not typically used for 16S rRNA-Seq, using minitax with the genome collection from NCBI, provided solid reconstructions up to the genus level. This adaptability is beneficial for those wanting a consistent database strategy for both WGS and 16S rRNA sequencing. However, species-level precision using the NCBI db shows a substantial decline, underscoring the importance of choosing a database that aligns with the desired level of resolution. This was also evident in the GMS results (Fig. 9a and Supplementary Fig. S12b). Since the programs performed similarly to each other across both synthetic microbial communities, the observed differences in microbial composition reconstruction efficiency between the MCS and GMS can likely be attributed to DNA isolation methods and other factors, rather than bioinformatics tools or database selection. The Chi-square test results indicated that only in the case of the I method for the MCS the microbial composition reconstructed using either Emu or minitax (with the Emu db) did not significantly differ from the theoretical composition (Supplementary Fig. S12a, b).

Comparing minitax with Emu and DADA2 using Illumina V1-V2 sequencing of MCS. We utilized an Illumina-sequenced V1-V2 dataset derived from the MCS DNA samples. In terms of r^2 values, both Emu and minitax workflows markedly outperform DADA2, both at the genus and the species-levels, even when using the standard SILVA db, which is commonly utilized by DADA2 (Fig. 9b). It is particularly noteworthy that while Emu and DADA2 are optimized for amplicon sequencing, minitax remains competitive, especially given its versatility in managing both 16S rRNA and WGS datasets. Emu provides slightly higher r^2 values (when using the Emu db) only at the species level, compared to minitax. In the MCS samples sequenced using the V1-V2 region the Chi-square tests indicated that both Emu and minitax reconstructed a microbial composition that was not significantly different from the theoretical one (Supplementary Fig. 12b), similarly to the ONT V1-V9 sequencing results.

Comparing minitax with sourmash using MCS data of PacBio HiFi WGS. We compared minitax to sourmash, the latter being

acknowledged as one of the leading software choices for the analysis of both LRS and SRS mWGS data¹⁰, using the PacBio HiFi dataset (NCBI accession: SRX9569057) from GMS. We carried out the comparison of r^2 values with and without excluding unclassified reads. Minitax outperformed Sourmash when unclassified reads were included in the abundance estimation, yet fell marginally short at the species level when these reads were excluded (Fig. 9c), as this altered the relative abundances of the identified taxa. Chi-square tests conducted using three different species detection threshold limits showed that, with minitax, the observed and theoretical distributions were not significantly different when using 0.1% and 0.01% detection thresholds (Supplementary Fig. 12c).

CAMISIM: simulated mouse gut datasets. Expanding our validation range, we examined simulated datasets from the CAMISIM mouse gut project, consisting of 10 samples each from PacBio and Illumina. For both data types, minitax exhibited reliable performance, reaching an $r^2 = 0.96$ at the phylum level, which decreased to a mean $r^2 = 0.46$ and 0.55 at the species level with Illumina and PacBio reads, respectively (Fig. 9d).

Discussion

In this comprehensive study, we compared metagenomic methods using canine fecal samples and two synthetic microbial community standards. Our primary objective was to evaluate the effectiveness of various methodological approaches including DNA extraction, library preparation, and bioinformatics analysis. To facilitate this, we developed “minitax”, a versatile bioinformatics tool designed for use with diverse metagenomic laboratory protocols.

Choosing the proper DNA isolation method is vital for maximizing yield and reducing fragmentation, thus preserving the integrity of further analyses¹⁷. The ideal extraction kit delivers high-quality, high-yield DNA to reduce the risk of false negatives⁵⁷, removes PCR inhibitors found in fecal samples^{58–62}, provides consistent results, and effectively breaks down Gram+ bacterial cell walls for a true reflection of community composition^{9,24–27}.

Our assessment of four DNA isolation kits - Zymo (Z), Qiagen (Q), Macherey-Nagel (MN), and Invitrogen (I)—revealed notable differences in DNA yield, fragment length, and overall effectiveness. For canine fecal samples, the Z method produced the highest yield and longest average DNA fragment length (Table 1, Fig. 4a–f), with minimal relative variation, indicating superior stability. In contrast, the Q method yielded the lowest amount of DNA with relatively high levels of host DNA contamination. The MN approach had the highest overall yield but shorter fragment lengths, while the I technique provided adequate quality with higher variability in yield and fragment length. For MCS samples, the Z kit yielded significantly more DNA with longer fragments, outperforming the MN kit, which had a shorter average peak length (Table 1). For GMS samples, the MN kit slightly edged out the Z kit in yield, though both performed well (Table 1). Overall, the Z and MN kits demonstrated superior DNA yields, with the Z kit excelling in MCS samples and the MN Kit in GMS samples.

In terms of microbial diversity in canine fecal samples, the I and MN kits consistently provided similar results across various libraries, sequencing methods, and bioinformatic analyses. The Z kit showed some divergence from these two, while the Q kit performed the worst in accurately representing microbial diversity (Fig. 4g–i).

For MCS samples, the I kit demonstrated the best alignment with the theoretical microbial composition, while the Z and MN kits were less accurate. The Q kit, however, was notably less effective, showing significant deviations from the theoretical composition. In GMS samples, the MN and Z kits were closest to the theoretical composition, though neither was optimal. The I kit, while better for MCS samples, did not perform as well with GMS samples (Fig. 7).

Our findings suggest that the differences in kit performance are influenced by the sample type and DNA purification method. For instance, the shield used for MCS and GMS samples seems to be suboptimal for all kits, potentially impacting extraction efficiency. Additionally, the very low

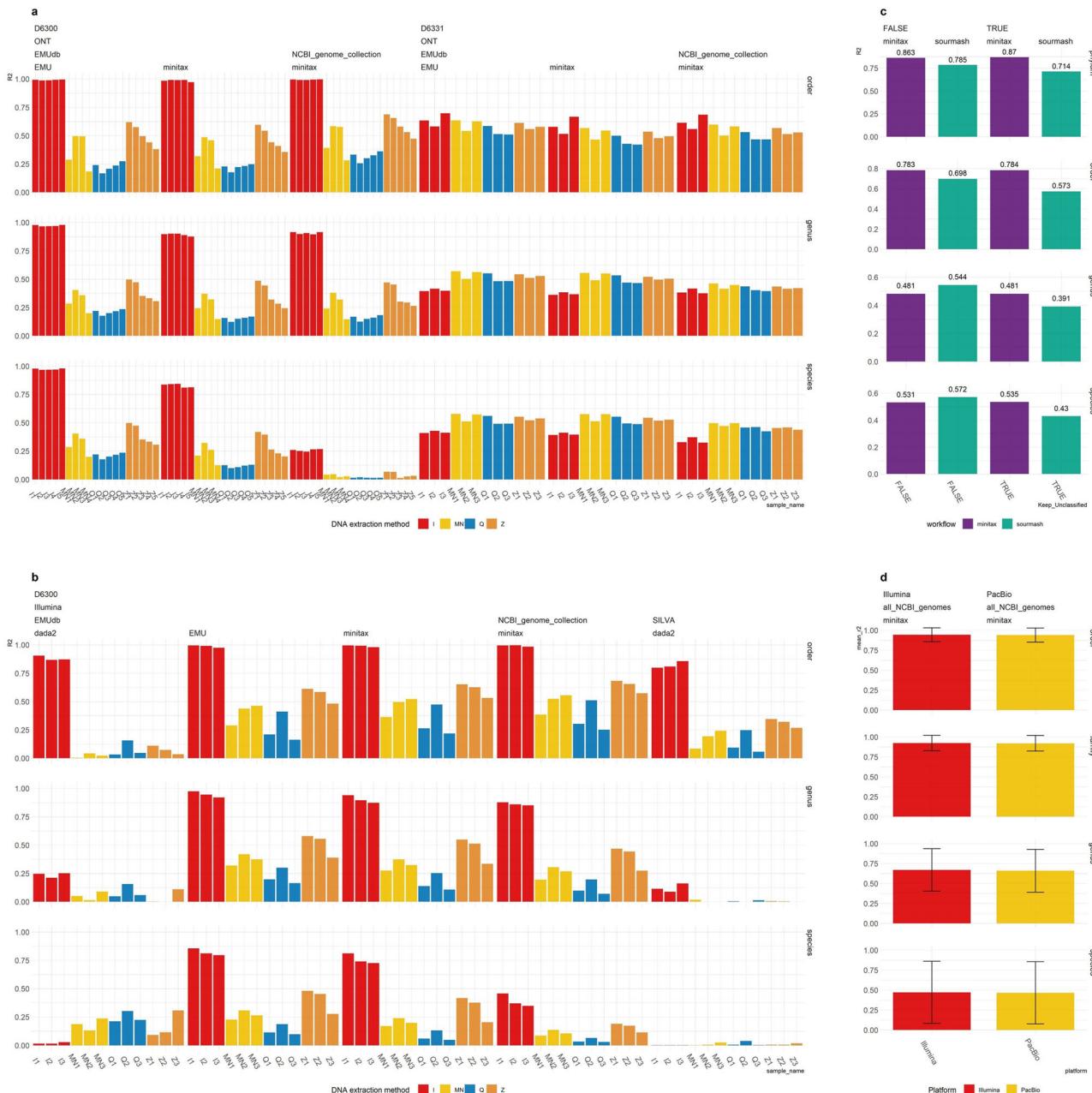


Fig. 9 | Comparative analysis of minitax performance across different sequencing platforms and datasets. The comparison between the theoretical and the observed microbial compositions, was generated using different software and databases. r^2 values were calculated across four taxonomic levels (phylum, order, genus, species) and plotted. **a** ONT V1-V9 Sequencing: performance comparison of minitax against Emu for the Zymo D6300 microbial community reference. **b** Illumina V1-V2 sequencing.

sequencing: minitax benchmarking against Emu and DADA2 for Zymo D6300 MCS. **c** PacBio HiFi WGS: minitax vs. sourmash performance comparison on Zymo D6331 MCS dataset. **d** CAMISIM mouse gut dataset: comparative analysis of minitax performance on Illumina and PacBio platforms. Ten samples were utilized from the CAMISIM database. Y-axis represent the mean r^2 -value for each group, error bars show standard deviation ($n = 10$).

DNA yields obtained with some kits might contribute to discrepancies in microbial community representation. Consequently, microbial standards appear not to be suitable for selecting the optimal DNA isolation kit.

Based on our analysis, the Z kit is recommended for most applications due to its high yield, stable performance, and ability to produce high-quality DNA, making it particularly beneficial for long-read sequencing and accurate microbial diversity analysis. The MN and I kits also perform well, but while they each have certain strengths, they also have specific weaknesses depending on the sample type and application (Table 2). The Q kit proved to be the least suitable, as it produces low-quality DNA and provides an inaccurate representation of the microbial community composition.

Overall, the choice of DNA isolation kit should be based on the specific needs of the sequencing application and the nature of the samples being analyzed.

Furthermore, we observed that ONT 16S rRNA amplicon libraries required the least experimental effort but demonstrated lower read quality compared to Illumina and PacBio methods. Illumina WGS and PacBio amplicon libraries, while more labor-intensive, consistently provided high-quality results. Notably, the PerkinElmer (PE) V1–V3 method often produced nonspecific DNA fragments, which adversely affected its performance. This contrasts with the other methods, which showed reliable quality and consistency.

Our study, in alignment with recent studies^{12,20}, indicates that WGS offers greater taxonomic diversity than 16S-Seq, however, unlike SRS platforms targeting one or a few variable regions, full 16S gene sequencing significantly improves bacterial community identification. Therefore, our analysis encompassed both SRS and LRS platforms, as well as 16S rRNA gene amplicon sequencing and mWGS strategies.

We found that the WGS data composition obtained using the I and MN DNA isolation methods closely resembles the bacterial composition from the Z V3-V4 library, with the ONT V1-V9 library also showing similar compositional profiles for both I and MN kits. Libraries prepared from Z DNA samples generally exhibit strong compositional similarity with those prepared using other methods, indicating that Z DNA isolates yield relatively consistent microbial community profiles across various library preparation techniques.

We observed that the V1-V3 libraries showed significant variability across all DNA extraction techniques. Notably, this variability was observed irrespective of the DNA extraction kit employed, reaffirming the necessity for careful selection of library preparation methods to ensure accurate microbial community profiling. Notably, an in-depth assessment of 16S V1-V9 sequencing using both ONT and PacBio platforms revealed that only the Z method produced consistent results across both libraries. Additionally, we assessed both a V1-V2 and a V1-V9 library on MCS. We obtained that the I method yielded the closest match between the experimental and theoretical compositions in both libraries. The MN method was more similar to the approach of Z approach and significantly differed from I method in both SRS and LRS libraries. These findings underscore the varying performances between *in vitro* and *in vivo* sample analyses, thereby highlighting the inapplicability of the MCS for validation purposes in other systems.

The α -diversity analysis showed that in SRS libraries, diversity metrics were generally consistent across most methods, except for variations linked to different library preparation techniques. Specifically, PacBio and Illumina methods exhibited distinct diversity profiles, significantly affecting microbial composition. The β -diversity analysis further demonstrated that library preparation methods had a greater influence on sample clustering than DNA isolation methods. The clear separation between V1-V2 and V3-V4 regions underscores the significant impact of library preparation protocols on microbial community profiles. The observed differences in microbial diversity and composition between V1-V3 and V3-V4 libraries suggest that sequencing methods play a crucial role in shaping the perceived microbial community. The consistent clustering of samples based on library preparation techniques, rather than DNA isolation methods, emphasizes the importance of selecting appropriate library preparation methods for accurate microbial analysis.

The significance of the applied bioinformatic tool in metagenomic analysis has been discussed by others³³. We also addressed this issue and found that approximately 60% of the variation in microbial community profiles is attributed to computational factors. Both Emu and minitax consistently yielded results that were closely aligned, and for MCS samples, they correlated highly with the theoretical compositions of each amplicon library. A similar pattern was evident between sourmash and minitax for the mWGS library, underlining the versatility of the minitax tool. In contrast, the results obtained by DADA2 pipeline varied considerably from those of minitax and Emu in both *in vivo* and *in vitro* samples.

When evaluating the best tools for specific types of sequencing data, Emu stood out as a reliable choice for amplicon sequencing, particularly across various V-regions. For mWGS, although this study did not focus extensively on these programs, sourmash—identified by Portik and colleagues³³ as one of the most precise and accurate tools—performed well in our analysis. These results suggest that Emu is optimal for amplicon data, while sourmash is highly effective for metagenomic WGS data. Minitax, however, proved to be a versatile and robust bioinformatics tool, performing consistently across various sequencing platforms and library types. While it did not always produce the highest correlation values, minitax often outperformed other methods by delivering reliable results. It demonstrated strong capabilities in handling both amplicon and metagenomic WGS data,

achieving genus-level resolution for amplicon data even when using complete genome databases. This versatility makes minitax a viable alternative, especially for studies requiring reliable comparisons across different sequencing approaches.

In summary, our research offers an in-depth assessment of the gut microbiome, underscoring that methodological optimization and uniformity are important for ensuring accurate representation and reproducibility. However, it is important to note that expanding the range of methodologies can also be advantageous as it helps to surmount the intrinsic limitations of a given approach, thereby enabling a more thorough understanding of the subject under investigation².

We recommend the following wet-lab pipelines for gut microbiome profiling:

1. For a cost-effective study, use **Z DNA isolation with ONT V1-V9 library preparation**. This combination offers an affordable solution, suitable for longitudinal studies where precise microbial composition is less critical.
2. For a balanced approach, **combine I DNA isolation with V3-V4 library preparation on Illumina**. This setup strikes a good balance between cost and accuracy, providing reliable microbial profiling at a moderate expense.
3. For projects requiring the highest precision, employ **MN DNA isolation with Illumina DNA Prep (WGS) library preparation**. Although this option is more expensive, it delivers the highest precision in microbial profiling, making it ideal for studies requiring the most detailed analysis.

For optimal bioinformatics analysis across different sequencing approaches, we recommend:

4. For studies requiring reliable comparisons across different sequencing approaches, consider using **minitax** with the NCBI genome collection, as it demonstrated versatile and robust performance across both amplicon and metagenomic WGS data.
5. **Emu** is recommended for optimal results in amplicon sequencing.
6. **Sourmash** is highly effective for metagenomic WGS, as identified in previous studies and supported by our analysis.

We did not fully process the dataset in all aspects, which makes it potentially valuable to the research community for further exploration and analysis.

Methods

Three distinct sample types were used to evaluate the performance of four different DNA isolation kits (Fig. 3). These samples included canine feces and two different microbial community mixtures, MCS and GMS, which comprised eight and eighteen bacterial strains, respectively.

Sample collection

A fecal sample from a 13.5-year-old healthy male Pumi (a Hungarian purebred dog) was collected within 1 min of defecation and was immediately frozen and stored at -80°C . Stool samples from six other healthy Pumi dogs (four puppies, a 7-year-old female, and a 6.5-year-old sterilized male) were used as controls (Table 4). Throughout the project, a single stool sample of 2.5 g was used for all DNA isolations. This sample was divided into portions required by the different extraction kits, resulting in a total of 16 DNA isolations (4 replicates per extraction kit). To avoid bias, the portions were not allocated based on their position within the original sample; instead, division and assignment to the kits were performed randomly. Given the small sample size, homogenization was not applied. This decision is supported by findings from Liang et al. (2020), who concluded that while homogenization is critical for metabolomic analyses, it is not necessary for microbiome studies. As controls, random 200 mg samples were taken from six canine specimens from six dogs. Furthermore, the ZymoBIOMICS Microbial Community Standard (MCS; Zymo Research, D6300) and the ZymoBIOMICS Gut Microbiome Standard (GMS; Zymo Research, D6331) were used to validate the obtained data.

Table 4 | Overview table of the fundamental characteristics of the dogs included in this study

Gender (ID number)	Age	Registered name “call name”	Sampling time	Living place at sampling time
Male	13.5 years	Harsányi-Tekergő Bátor ‘Toto’ (Toti)	Nov. 20, 2021	Szeged, Hungary
Female (1)	7 years	Galla-Hegyi Suba ‘Sugó’	May 08, 2021	Baja, Hungary
Male (1) ^a	6.5 years	Vöröskői Kondacsipkédő Lidérca ‘Loki’	May 13, 2021	Baja, Hungary
Female (2)	6 months	Duna-menti Dumás <u>Boszi</u>	May 9, 2021	Baja, Hungary
Female (3)	6 months	Duna-menti Dumás <u>Borcsa</u>	May 4, 2021	Budapest, Hungary
Male (2)	6 months	Duna-menti Dumás <u>Blöff</u>	May 5, 2021	Budakalász, Hungary
Male (3)	6 months	Duna-menti Dumás <u>Briós</u>	May 4, 2021	La Tour-de-Peilz, Switzerland

^aSterilized; underlined names are used in data identifiers in ENA. To distinguish the individual dogs in both the table and figures, including the Supplementary Figs., we identified them by their sex and sequence number within that sex. The main subject of the publication, the 13.5-year-old dog, is always discussed independently and thus it is not assigned a number. The results from the other six dogs, whose data were used solely for the validation of certain findings, are always presented together in the figures, and therefore, they have been numbered accordingly. We used the underlined name as an identifier for the files uploaded to the data repository.

Table 5 | The list of Illumina Reagent Kits used in this study

DOG	V1-V2	V1-V3	V3-V4	WGS	MCS	V1-V2
v2 500 cycles				1		
v2 300 cycles				1		
v2 nano 500 cycles	1	3	3		2	
v2 nano 300 cycles				2		

DNA purification and library preparation

In our work, the following commercially available DNA isolation kits were tested using four technical replicates for each: 1. QIAGEN QIAamp Fast DNA Stool Mini Kit (Ref. # 51604; Lot. # 169025369); 2. Invitrogen Pure-Link™ Microbiome DNA Purification Kit (Cat. # 423482); 3. Macherey-Nagel NucleoSpin DNA Stool Mini kit (Ref. # 740472.50; Lot. # 2302-001); and 4. Zymo Research Quick-DNA™ HMW MagBead Kit (Cat. # D6060). We used the following kits for library preparation: Illumina DNA Prep (Doc. # 1000000025416 v09), Oxford Nanopore Technologies Rapid Sequencing 16S Barcoding Kit (SQK-RAB204), Pacific Biosciences Full-Length 16S Library Preparation Using SMRTbell Express Template Prep Kit 2.0 (PN 101-916-900), PerkinElmer NEXTFLEX® 16S V1-V3 Amplicon-Seq Kit for Illumina (Cat. # 4202-02) and Zymo Research Quick-16S™ NGS Library Prep Kit (Cat. # D6400).

The DNA isolation and library preparation were performed according to the protocols of the kits. The detailed protocols are provided in the Supplementary Methods to ensure that the versions we used are available in case the manufacturers make changes to the protocols in the future.

DNA quantification and quality assessment

The Qubit 4.0 Fluorometer was used for quantifying the DNA yield, whereas the TapeStation 4150 instrument was applied for evaluating the DNA quality, including fragment size distribution.

Statistical analysis. Data were analyzed using one-way ANOVA to compare the DNA yield across the four kits. ANOVA tests were conducted separately for each sample type (canine feces, MCS, and GMS). To assess the significance of differences between pairs of group means, Tukey’s Honestly Significant Difference (HSD)⁶³ test was used. This post-hoc test helps determine which specific pairs of kits significantly differ from each other.

Relative standard deviation (RSD). The RSD was calculated to assess the variability within each kit’s performance. The RSD is computed as follows: $RSD = \frac{\text{mean}}{\text{standard deviation}} \times 100\%$.

Sequencing

Altogether nine MiSeq Reagent Kit v2 and four MiSeq Reagent Kit v2 Nano were used for SRS (Table 5). V1–V9 sequencing was carried out on ONT MinION and on PacBio Sequel IIe instruments. ONT V1–V9 barcoded libraries from the canine samples were loaded onto three MinION flow cells and three additional flow cells were run for sequencing the MCS and GMS samples. PacBio V1–V9 libraries were barcoded and run on a Sequel SMRT Cell 8 M.

Bioinformatic analysis

Brief description of the minitax software. We developed minitax, a versatile taxonomic assignment tool designed to address the challenges posed by diverse sequencing data types. Minitax enables taxonomic profiling across various sequencing platforms (ONT, PacBio, Illumina) and library types, including mWGS and 16S rRNA gene sequencing. At its core, minitax uses minimap2 with platform-specific parameter settings (as per minimap2) for initial read alignment to the supplied database. Minitax utilizes Rsamtools⁶⁴ to import the alignment into R and merges it with the database information. The workflow relies on data.table⁶⁵ to handle large datasets efficiently. It then performs sophisticated post-alignment processing steps to select the best alignments for each read. The post-alignment processing includes several key steps:

1. Alignment filtering: minitax first applies an overall MAPQ-based filter (e.g., MAPQ values between 1 and 59). It then selects the alignment with the highest MAPQ for each read, followed by keeping the highest CIGAR-score alignment based on platform-specific scoring matrices (Supplementary Data 8).
2. Taxonomic assignment: After filtering, minitax determines the lowest common taxonomic rank (“tax.identity”) and the corresponding rank level (“tax.identity.level”) for each read.
3. Taxonomic Refinement: minitax offers several refinement options to resolve reads with multiple valid alignments:

BestAIn: This method refines the taxonomic assignment of each read by evaluating the proportion of alignments supporting different taxa at a given rank. For each taxonomic rank (e.g., order, family, genus), BestAIn calculates the ratio of alignments that map to each possible taxon. If a particular taxon is supported by a majority of alignments (or exceeding a threshold, default 60%), and this taxon is more specific than the current assignment, BestAIn refines the read’s identity to this taxon. This approach ensures that the final assignment reflects the most probable and specific taxonomic identity based on the available alignment data.

RandAln: This method randomly selects one alignment for each read from the pre-filtered subset, providing a straightforward resolution for ties.

SpeciesEstimate: This method uses all valid alignments for a read to estimate species-level abundances. It normalizes the read counts based on the number of alignments, providing a weighted estimate of each species' presence.

LCA (Lowest Common Ancestor): This method assigns each read to the lowest common ancestor of all aligned taxa, providing a conservative approach that ensures the read is assigned at the most specific taxonomic rank shared by all alignments.

4. **Taxonomic summary:** Finally, minitax summarizes read counts based on the selected taxonomic rank and writes the outputs of each step in .tsv format. The final output is provided as a phyloseq⁶⁶ object, facilitating downstream analysis. Minitax is available at: <https://github.com/Balays/minitax>.

The CIGAR scoring schemes, along with other parameters, are user-controllable. The results in this study were obtained using the *RandAln* method, though in many cases, the *BestAln* method may yield more precise results.

Analysis of the raw data

Illumina V1–V2, V3–V and V1–V3. *DADA2 pipeline* on Illumina amplicon reads: The raw reads were subjected to DADA2⁴⁴ (version 1.22) for quality control, filtering, trimming and subsequent dereplication (ASV generation), bimera detection/removal and lastly the taxonomic assignment. The latter was carried out using either the SILVA 16S database⁶⁷ (version: 138.1) or Emu's database (version: v3.4.4). The exact parameters and settings can be found in SUPPTAB_X (dada2_config.tsv) and the complete workflow on the GitHub page (dada2.WF.R)

Emu and minitax pipeline on Illumina reads: The Emu software⁴⁶ (version: v3.4.4) and the in-house developed minitax program (version 1.0) were used on the raw Illumina reads with the following parameters: --type sr and --N 10, otherwise default settings was applied. We used Emu's default database in both cases and additionally a genome collection from NCBI in the case of minitax. Both programs utilize minimap2⁴⁸ for initial read alignment.

ONT V1–V9. The raw voltage values obtained from the MinION sequencer platform were basecalled with the 6.1.5 version of Guppy (MinKNOW 20.05.8) at super-high accuracy. The reads were demultiplexed using the SQK-RAB204 barcode sequences. During basecalling, the minimum quality score was set to 8, thus fail and pass reads were obtained. For further analyses only the pass reads were used. The pass reads were then applied as an input for Emu (version: v3.4.4) and minitax (version 1.0) using the following parameters: --type map-ont and --N 10 otherwise defaults, and using Emu's default database for both programs and additionally a genome collection from NCBI in the case of minitax. In addition, ONT's EPI2ME pipeline (3.6.1) was also used.

PacBio V1–V9. Basecalling, demultiplexing and generation of HiFi reads were carried out by using the SMRT Link 10.2.0.133434 analysis software from PacBio. High quality CCS reads were filtered: --min-qv 20. The filtered reads were then used as an input for Emu (version: v3.4.4) and minitax (version 1.0) using the following parameters: --type map-ph and --N 10 otherwise defaults, and using Emu's default database in both cases and additionally a genome collection from NCBI in the case of minitax.

Illumina WGS. The raw reads were trimmed using Trim Galore and host reads filtered with (<https://github.com/FelixKrueger/TrimGalore>) BMTagger (<https://hpc.ilri.cgiar.org/bmtagger-software>) using the *Canis lupus familiaris* genome assembly (GCF_014441545.1). The quality controlled reads were subjected to either the fast and sensitive taxonomic

classifier program sourmash (version 4.8.2) using the provided Genbank genomes database (from March 2022) or the minitax software (version: 1.0) using default parameters and a genome collection from NCBI as a reference.

Benchmarking datasets. Besides the Zymo D6300 Microbial Community Mixture sequenced using ONT V1–V9 and Illumina V1–V2 methods described earlier, we evaluated the performance of minitax on two other datasets published elsewhere.

PacBio HiFi WGS of Zymo MCM D6331 Microbial Community: This dataset was used in Portik et al. 2022, for benchmarking of long-read mWGS software. The fastq files were downloaded from NCBI (NCBI accession: SRX9569057) and used as an input for either the sourmash program or minitax. In both the D6300 and D6331 workflows, reads mapped to any species within the genus *Veillonella* were attributed to *Veillonella rogosae*, and similarly, all reads from the *Roseburia* genus were attributed to the species *Roseburia hominis*.

CAMISIM Datasets: The CAMISIM Simulated Mouse Gut Project⁶⁸, provided simulated reads in fastq format, of which we randomly downloaded 10 samples each for PacBio and Illumina to use as input for the programs.

Databases. We utilized a comprehensive set of ~18,000 genomes spanning 13,000 species of *Bacteria*, *Archaea* and *Eukaryotes*, downloaded in February, 2022 from NCBI as a reference for both WGS and 16S gene sequencing data. Additionally, for the amplicon sequencing datasets, we employed either the Emu or the SILVA (version 138) 16S databases, depending on the software used.

Taxonomy version. When comparing different DNA isolation and library preparation techniques, we used the NCBI taxonomy version corresponding to the genomes included in the collection. However, in Sections 2 and 3 of the results, we transitioned to the most recent (July 2024) version of the NCBI taxonomy. This update was not applied to other sections because the databases used for comparison with minitax results were based on the older version of the taxonomy.

CAMISIM. The simulated mouse gut dataset was used for additional validation of the minitax workflow. Fifteen samples for both the short-read (Illumina) and long-read (PacBio) datasets were employed.

Statistics and downstream data analysis. The results of the different programs were combined into 'phyloseq' objects⁶⁵ and subsequently combined into a single, comprehensive phyloseq object encompassing all the metagenomic read count data. Subsequent data analysis was performed using this phyloseq object in the R environment, employing the following packages: data.table⁶⁵, tidyverse⁶⁹, FactoMineR⁷⁰, and vegan⁷¹. The scripts for the complete downstream data analysis and plots can be downloaded from GitHub repository (<https://github.com/Balays/Microbiome-Method-Comparison>).

PERMANOVA, PERMDISP2, PCoA, and NMDS analyses. We applied various multivariate techniques to assess the differences in microbial community composition based on DNA isolation method and library preparation, using the Bray-Curtis dissimilarity matrix (Fig. 5c). The following analyses were conducted:

PERMANOVA (Permutational Multivariate Analysis of Variance): To test for significant differences in microbial composition between groups, we conducted a full-model PERMANOVA using the following formula:

```
adonis2(otutab_t ~ DNA_isolation_method + library, data=sampdat,
method='bray')
```

This model tested the combined effects of DNA isolation method and library preparation on microbial community composition, using 999 permutations.

PERMDISP2 (Permutational Analysis of Multivariate Dispersion): We used betadisper to assess whether group dispersion (i.e., within-group variability) differed between groups, applying it to the Bray-Curtis dissimilarity matrix:

```
betadisper(distance_matrix, groups)
```

PCoA (Principal Coordinates Analysis): We performed PCoA on the Bray-Curtis dissimilarity matrix for each DNA isolation method separately, using the output of *betadisper*, to evaluate the effect of library preparation within each method. The distances to centroids were calculated, and results were merged across the isolation methods for visualization (Fig. 5b).

NMDS (Non-metric Multidimensional Scaling): To further explore the relationships between samples, we conducted NMDS using Bray-Curtis dissimilarity on the normalized abundance data:

```
ordnate(ps.prop, method="NMDS", distance='bray')
```

NMDS was used to visualize the similarity between samples, with the relative proximity of points reflecting the closeness of microbial compositions across samples (Fig. 5a).

Performance metrics and statistics. To evaluate minitax, and other tools in terms of their ability to reconstruct microbial abundances of the reference compositions, precision, recall, F1 and F0.5 scores were calculated by comparing the presence or absence of taxa between the theoretical and experimental datasets. These values were calculated from the True Positives (TP), False Positives (FP), and False Negatives (FN) against the theoretical compositions for 1%, 0.1%, and 0.01% species detection thresholds after Portik et al.³³.

We also computed the Chi-squared statistic for each taxon using the formula:

$$\chi^2 = \frac{(Observed - Expected)^2}{Expected},$$

which was then summed across all taxa to assess the overall goodness-of-fit between the theoretical and observed distributions. The Chi-squared test was used to assess whether the observed distribution significantly differed from the theoretical distribution. A Bonferroni correction was applied to account for multiple comparisons, and the null hypothesis was rejected if the corrected *p*-value was below the significance threshold (0.05).

In addition, we calculated Pearson's correlation coefficients (*r*² values) to evaluate the linear relationship between the theoretical and observed microbial abundances.

We also identified significantly different taxa (between the expected and observed compositions) using DESeq⁷² with the following commands:

```
de.seq <- phyloseq_to_deseq2(ps, ~DNA_isolation_method)
de.seq <- DESeq(de.seq, test = "Wald", fitType = "parametric", sfType =
"poscounts").
```

The taxa that showed a *p*-value of <0.05 and a log2Fold difference of $\geq 2 | \leq -2$ were considered significantly different.

Downsampling. The raw reads from the Illumina mWGS datasets were randomly subsampled to the following read counts: 2,500,000, 2,000,000, 1,500,000, 1,000,000, 750,000, 500,000, 250,000, 22,5000, 200,000, 175,000, 150,000, 125,000, 100,000, 75,000, 50,000, 20,000, 10,000 and 5000. Subsequently the Kaiju program (version 1.9.0) was run on the subsets using the progenomes (v2) database to assign taxonomic classifications.

Similarly, the raw reads from the non-mWGS datasets were randomly subsampled to the following read counts: 2,500,000; 2,000,000; 1,500,000; 1,000,000; 750,000; 500,000; 250,000; 225,000; 200,000; 175,000; 150,000; 125,000; 100,000; 75,000; 50,000; 20,000; 10,000; and 5000. The subsampled datasets were then processed using Emu (version: v3.4.4) with default settings.

Shannon index values were calculated from the obtained results using the phyloseq R package. Every obtained taxonomic abundance output was

then analyzed with the phyloseq⁶⁶ R package and visualized with ggplot2 (<https://github.com/gabor-gulyas/Technical-article-downsample>).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All our sequencing data have been submitted to the ENA under the accession PRJEB59610. The datasets from other sources were downloaded from the ENA with the following accession numbers: PRJNA783735, PRJNA678365, PRJNA871395. Supplementary Data files (1–21): figshare, <https://doi.org/10.6084/m9.figshare.27232626>⁷³. In addition, data used to generate figures can be found under the project's github repository: <https://github.com/Balays/Microbiome-Method-Comparison>, specifically:—Figure 4 g: Fig4.G-H_MCM_all_Adversity_estimates.tsv – Figure 4 i: Fig4_statistics_data.rds—Figure 5 a: Fig_5_A_BrayCurtis_distances.tsv—Figure 5 b: Fig_5B_PCoA_data.tsv—Figure 5 c: Fig_5C_Heatmap_data.tsv – Figure 5 d: ./Dog_Feces/Microbial_abundances for each taxon level—Figure 5 e: DogFeces_all_methods_PS.rds – Figure 7: sig.freq.tsv and l2F_diff.dt.tsv separately for each group—Figure 7: ./MCM/.../sig.freq.tsv and l2F_diff.dt.tsv separately for each group - Figure 9 a-b: ./MCM/.../Detection_statistics.tsv for each each group—Figure 9 c: ./Portik_etal_2022/ Detection_statistics.tsv - Figure 9 d: ./CAMISIM/Detection_statistics.tsv. In addition, phyloseq objects that can be imported into R and contain the microbial abundances for the respective samples are also available: - MCM_ZymoD6300_all_PS.rds - MCM_ZymoD6331_all_PS.rds - DogFeces_all_methods_PS.rds. The data used to generate Fig. 4e, f are available in Supplementary Data 3, the data for Fig. 6 can be found in Supplementary Data 6, while the data for generate Fig. 8 are in Supplementary Data 21.

Code availability

minitax: <https://github.com/Balays/minitaxStatistics>: <https://github.com/Balays/Microbiome-Method-Comparison> other in-house scripts - <https://github.com/Balays/Microbiome-Method-Comparison> - <https://github.com/gabor-gulyas/Technical-article-downsample> DADA2: <https://benjineb.github.io/dada2/> Emu: <https://gitlab.com/treangenlab/emu> Trim Galore: <https://github.com/FelixKrueger/TrimGalore> BMTagger: <https://hpc.ilri.cgiar.org/bmtagger-software>

List of abbreviations

I	Invitrogen
MN	Macherey-Nagel
Q	Qiagen
PE	PerkinElmer
Z	Zymo Research
I WGS	Illumina DNA Prep mWGS library
ONT	Oxford Nanopore Technologies V1-V9 amplicon library
V1-V9	Pacific Biosciences V1-V9 amplicon library
PacBio	Pacific Biosciences V1-V9 amplicon library
V1-V9	PerkinElmer V1-V3 amplicon library
V1-V3	Zymo Research V1-V2 amplicon library
V1-V2	Zymo Research V3-V4 amplicon library
V3-V4	Long-Read Sequencing
LRS	Metagenomic Whole Genome Sequencing
mWGS	Oxford Nanopore Technologies
ONT	Pacific Biosciences
PacBio	Short-Read Sequencing
SRS	Whole-Genome Sequencing
WGS	ZymoBIOMICS Gut Microbiome Standard D6331
GMS	ZymoBIOMICS Microbial Community Standard D6300
MCS	Analysis of Variance
ANOVA	Tukey's Honestly Significant Difference
HSD	

MAPQ	Mapping Quality
NMDS	Non-Metric Multi-Dimensional Scaling
PCoA	Principal Coordinate Analyses
PER-MANOVA	Permutational Analysis of Variance
PER-MIDISP	Permutational Analysis of Dispersion
db	database
Gram-positive	Gram+
Gram-negative	Gram-
HMW	High Molecular Weight
LCA	Lowest Common Ancestor

comparison of three DNA extraction methods and two bioinformatic pipelines. *mSystems* **5**, e00547–19 (2020).

19. Tournousse, D. M. et al. Validation and standardization of DNA extraction and library construction methods for metagenomics-based human fecal microbiome measurements. *Microbiome* **9**, 95 (2021).
20. Lewis, S. et al. Comparison of 16S and whole genome dog microbiomes using machine learning. *BioData Min.* **14**, 41 (2021).
21. Shaffer, J. P. et al. A comparison of six DNA extraction protocols for 16S, ITS and shotgun metagenomic sequencing of microbial communities. *Biotechniques* **73**, 34–46 (2022).
22. Bukin, Y. et al. The effect of 16S rRNA region choice on bacterial community metabarcoding results. *Sci. Data* **6**, 190007 (2019).
23. Kerrigan, Z., Kirkpatrick, J. B. & D'Hondt, S. Influence of 16S rRNA hypervariable region on estimates of bacterial diversity and community composition in seawater and marine sediment. *Front. Microbiol* **10**, 1640 (2019).
24. de Boer, R. et al. Improved detection of microbial DNA after bead-beating before DNA isolation. *J. Microbiol. Methods* **80**, 209–211 (2010).
25. Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z. & Forney, L. J. Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* **7**, e33865 (2012).
26. Knudsen, B. E. et al. Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. *mSystems* **1**, e00095–16 (2016).
27. Teng, F. et al. Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci. Rep.* **8**, 16321 (2018).
28. Sinha, R. et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat. Biotechnol.* **35**, 1077–1086 (2017).
29. Mirzayi, C. et al. Reporting guidelines for human microbiome research: the STORMS checklist. *Nat. Med.* **27**, 1885–1892 (2021).
30. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **69**, 330–339 (2007).
31. Mallott, E. K., Malhi, R. S. & Amato, K. R. Assessing the comparability of different DNA extraction and amplification methods in gut microbial community profiling. *Access Microbiol.* **1**, e000060 (2019).
32. Tanca, A. et al. The impact of sequence database choice on metaproteomic results in gut microbiota studies. *Microbiome* **4**, 51 (2016).
33. Portik, D. M., Brown, C. T. & Pierce-Ward, N. T. Evaluation of taxonomic classification and profiling methods for long-read shotgun metagenomic sequencing datasets. *BMC Bioinforma.* **23**, 541 (2022).
34. Boldogkői, Z., Moldován, N., Balázs, Z., Snyder, M. & Tombácz, D. Long-read sequencing—a powerful tool in viral transcriptome research. *Trends Microbiol.* **27**, 578–592 (2019).
35. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
36. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
37. Lu, J. & Salzberg, S. L. Ultrafast and accurate 16S rRNA microbial community analysis using Kraken 2. *Microbiome* **8**, 124 (2020).
38. Brown, C. T. & Irber, L. sourmash: a library for MinHash sketching of DNA. *J. Open Source Softw.* **1**, 27 (2016).
39. Pierce, N. T., Irber, L., Reiter, T., Brooks, P. & Brown, C. T. Large-scale sequence comparisons with sourmash. *F1000Research* **8**, 1006 (2019).
40. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).

Received: 2 April 2024; Accepted: 28 October 2024;

Published online: 06 November 2024

References

1. The Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).
2. Walker, A. W. & Hoyle, L. Human microbiome myths and misconceptions. *Nat. Microbiol.* **8**, 1392–1396 (2023).
3. Gilbert, J. A. et al. Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).
4. Schlomann, B. H. & Parthasarathy, R. Timescales of gut microbiome dynamics. *Curr. Opin. Microbiol.* **50**, 56–63 (2019).
5. Jansma, J. & El Aidy, S. Understanding the host-microbe interactions using metabolic modeling. *Microbiome* **9**, 16 (2021).
6. Hou, K. et al. Microbiota in health and diseases. *Sig. Transduct. Target Ther.* **7**, 135 (2022).
7. Sun, Z. et al. Interactions between human gut microbiome dynamics and sub-optimal health symptoms during seafaring expeditions. *Microbiol Spectr.* **10**, e0092521 (2022).
8. Natalini, J. G., Singh, S. & Segal, L. N. The dynamic lung microbiome in health and disease. *Nat. Rev. Microbiol.* **21**, 222–235 (2023).
9. Santiago, A. et al. Processing faecal samples: a step forward for standards in microbial community analysis. *BMC Microbiol.* **14**, 112 (2014).
10. Sinha, R. et al. The microbiome quality control project: baseline study design and future directions. *Genome Biol.* **16**, 276 (2015).
11. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35**, 833–844 (2017).
12. Johnson, J. S. et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).
13. Abellan-Schneyder, I. et al. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *mSphere* **6**, e01202–20 (2021).
14. Wesolowska-Andersen, A. et al. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* **2**, 19 (2014).
15. Jones, M. B. et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc. Natl. Acad. Sci. USA* **112**, 14024–14029 (2015).
16. Bag, S. et al. An improved method for high quality metagenomics DNA extraction from human and environmental samples. *Sci. Rep.* **6**, 26775 (2016).
17. Costea, P. I. et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat. Biotechnol.* **35**, 1069–1076 (2017).
18. Ducarmon, Q. R., Hornung, B. V. H., Geelen, A. R., Kuijper, E. J. & Zwittink, R. D. Toward standards in clinical microbiota studies:

41. Huson, D. H. et al. MEGAN community edition: interactive exploration and analysis of large-scale microbiome sequencing data. *PLOS Comput. Biol.* **12**, e1004957 (2016).
42. Odom, A. R. et al. Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. *Sci. Rep.* **13**, 13957 (2023).
43. Coelho, L. P. et al. Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* **6**, 72 (2018).
44. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
45. Menzel, P., Ng, K. L. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **7**, 11257 (2016).
46. Curry, K. D. et al. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* **19**, 845–853 (2022).
47. Kerkhof, L. J. Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol. Ecol.* **97**, fiab001 (2021).
48. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
49. You, I. & Kim, M. J. Comparison of gut microbiota of 96 healthy dogs by individual traits: breed, age, and body condition score. *Animals* **11**, 2432 (2021).
50. Söder, J. et al. Composition and short-term stability of gut microbiota in lean and spontaneously overweight healthy Labrador retriever dogs. *Acta Vet. Scand.* **64**, 8 (2022).
51. Thomson, P., Santibáñez, R., Rodríguez-Salas, C., Flores-Yáñez, C. & Garrido, D. Differences in the composition and predicted functions of the intestinal microbiome of obese and normal weight adult dogs. *PeerJ* **10**, e12695 (2022).
52. Li, Z. et al. Analysis and comparison of gut microbiome in young detection dogs. *Front. Microbiol.* **13**, 872230 (2022).
53. Xu, J. et al. The response of canine faecal microbiota to increased dietary protein is influenced by body condition. *BMC Vet. Res.* **13**, 374 (2017).
54. Matsuo, Y. et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. *BMC Microbiol.* **21**, 35 (2021).
55. Cha, T. Gut microbiome profiling of neonates using nanopore MinION and Illumina MiSeq sequencing. *Front. Microbiol.* **14**, 1148466 (2023).
56. Rozas, M., Brillet, F., Callewaert, C. & Paetzold, B. MinION™ nanopore sequencing of skin microbiome 16S and 16S-23S rRNA gene amplicons. *Front. Cell Infect. Microbiol.* **11**, 806476 (2022).
57. Salonen, A. et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* **81**, 127–134 (2010).
58. Monteiro, L., Cabrita, J. & Mégraud, F. Evaluation of performances of three DNA enzyme immunoassays for detection of Helicobacter pylori PCR products from biopsy specimens. *J. Clin. Microbiol.* **35**, 2931–2936 (1997).
59. Flekna, G., Schneeweiss, W., Smulders, F. J. M., Wagner, M. & Hein, I. Real-time PCR method with statistical analysis to compare the potential of DNA isolation methods to remove PCR inhibitors from samples for diagnostic PCR. *Mol. Cell Probes* **21**, 4 (2007).
60. Nechvatal, J. M. et al. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. *J. Microbiol. Methods* **72**, 124–132 (2008).
61. Li, X. et al. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. *J. Int. Med. Res.* **48**, 300060520925594 (2020).
62. Josefson, M. H., Andersen, S. C., Christensen, J. & Hoofar, J. Microbial food safety: potential of DNA extraction methods for use in diagnostic metagenomics. *J. Microbiol. Methods* **114**, 30–34 (2015).
63. Tukey, J. W. The problem of multiple comparisons. In: *Proc. 1953 Biennial Research Conference on the Teaching of Statistics* (pp. 15–20). Princeton University, 1953. Tukey, J. W. *Exploratory Data Analysis* Addison-Wesley (1977).
64. Morgan, M., Pagès, H., Obenchain, V. & Hayden, N. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. <https://doi.org/10.18129/B9.bioc.Rsamtools>, R package version 2.18.0, <https://bioconductor.org/packages/Rsamtools>. (2023).
65. Barrett, T. et al. data.table: Extension of ‘data.frame’. R package version 1.16.99, <https://Rdatatable.gitlab.io/data.table>, <https://github.com/Rdatatable/data.table>, <https://r-datable.com>
66. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
67. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
68. Fritz, A. et al. CAMISIM: simulating metagenomes and microbial communities. *Microbiome* **7**, 17 (2019).
69. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
70. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
71. Oksanen, J. et al. vegan: Community Ecology Package_. R package version 2.6-4. <https://github.com/vegandevels/vegan> (2022).
72. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
73. Tombácz, D. et al. Cross-comparison of gut metagenomic profiling strategies [Data set]. figshare <https://doi.org/10.6084/m9.figshare.27232626> (2024).

Acknowledgements

This work was supported by the Momentum program I of the Hungarian Academy of Sciences LP2020-8/2020 (D.T.), the National Research, Development and Innovation Office FK 142676 (D.T.) and K 142674 (Z.B.). Á.D. was a grantee of the New National Excellence Program (ÚNKP-22-3-SZTE-245). APC fee was covered by the University of Szeged Open Access Fund: 6927. The primary samples were collected from Harsányi-Tekergő Bátor ‘Toto’, thanks to the dog’s owners (G, P and M Tombácz from Szeged, Hungary). We also express our gratitude to Ildikó Abonyi, the owner of the Duna-menti Dumás Kennel (Baja, Hungary; <https://www.pumikennel.eu/>), for providing additional control dog samples (Sugó, Loki, and Boszi) and for facilitating the connection with new owners to ease the sample collection process. We would like to thank the owners of the dogs for their contributions: Eszter Dandé (Borcsa), Tünde Gyöngyös (Briós) and Júlia Lang (Blöff).

Author contributions

D.T. and Z.B. conceived the project and designed experiments. G.G., Á.D., T.J., Z.C., and D.T. performed experiments. G.G., K.B., D.T., Z.B., and I.P. conducted computational analyses. D.T. supervised the project. Z.B. and D.T. contributed key reagents. D.T., Z.B., B.K., G.G., and M. M.H. wrote the manuscript. All authors participated in the review and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at
<https://doi.org/10.1038/s42003-024-07158-6>.

Correspondence and requests for materials should be addressed to Zsolt Boldogkői or Dóra Tombácz.

Peer review information *Communications Biology* thanks Andrea Azcarate-Peril, Xiaowei Zhan, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: George Inglis and Christina Karlsson Rosenthal. A peer review file is available.

Reprints and permissions information is available at
<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024