



OPEN

Comparative analysis of illumina and oxford nanopore sequencing platforms for 16S rRNA profiling of respiratory microbial communities

Guillem Macip¹, Alba Soler-Comas¹, Andrea Palomeque^{1,2}, Ana Motos¹, Blanca Llonch¹, Joan Canseco-Ribas¹, Leticia Bueno-Freire¹, Davide Calabretta^{1,3}, Kasra Kiarostami¹, Roberto Cabrera¹, Miquel Ferrer^{1,2}, Antoni Torres^{1,2}✉ & Laia Fernandez-Barat¹✉

The respiratory microbiome plays a crucial role in health and disease, necessitating accurate characterization through high-throughput sequencing technologies. This study provides a comparative analysis of Illumina NextSeq and Oxford Nanopore Technologies (ONT) sequencing platforms for 16 S rRNA profiling of respiratory microbial communities. Illumina sequencing, known for its high accuracy and short-read lengths (~300 bp), is widely used for genus-level microbial classification but struggles with species-level resolution due to its limited read length. In contrast, ONT generates full-length 16 S rRNA reads (~1,500 bp), enabling higher taxonomic resolution but historically exhibiting higher error rates (5–15%). Analysis of alpha and beta diversity indicated that Illumina captured greater species richness, while community evenness remained comparable between platforms. Beta diversity differences were significant in pig samples but not in human samples, suggesting that sequencing platform effects are more pronounced in complex microbiomes. Taxonomic profiling revealed that Illumina detected a broader range of taxa, while ONT exhibited improved resolution for dominant bacterial species. ANCOM-BC2 differential abundance analysis highlighted platform-specific biases, with ONT overrepresenting certain taxa (e.g., *Enterococcus*, *Klebsiella*) while underrepresenting others (e.g., *Prevotella*, *Bacteroides*). These findings emphasize that platform selection should align with study Objective: Illumina is ideal for broad microbial surveys, whereas ONT excels in species-level resolution and real-time applications. Future research should explore hybrid sequencing approaches to leverage the strengths of both technologies, thereby improving microbiome characterization in both clinical and preclinical settings.

The human respiratory tract is a complex environment that hosts diverse microbial communities, collectively referred to as the **respiratory microbiome**. These communities play essential roles in maintaining respiratory health, modulating immune responses, and influencing susceptibility to various respiratory diseases, such as asthma, chronic obstructive pulmonary disease (COPD), and pneumonia. Dysbiosis—disruptions in the normal microbial composition—has been linked to increased inflammation, pathogen colonization, and worse clinical outcomes in respiratory diseases¹. Consequently, the accurate characterization of respiratory microbial communities is essential for understanding their roles in both health and disease.

Recent advancements in **next-generation sequencing (NGS)** technologies have revolutionized the study of the respiratory microbiome, allowing for high-throughput, culture-independent analysis of microbial communities. Among these technologies, **16 S ribosomal RNA (rRNA) gene sequencing** has emerged as the gold standard for bacterial community profiling due to its cost-effectiveness, broad taxonomic coverage, and applicability to low-biomass respiratory samples².

The 16 S rRNA gene contains both highly conserved regions, which serve as universal primer-binding sites, and hypervariable regions (V1–V9) that provide taxonomic specificity, making it an ideal target for amplifying and sequencing bacterial DNA from low-biomass respiratory samples. While the **V3–V4 regions**³ are commonly targeted for taxonomic classification, full-length sequencing of the 16 S rRNA gene offers higher taxonomic resolution, often enabling species-level identification⁴.

¹Ciberes-IDIBAPS, University of Barcelona, Barcelona, Spain. ²Pneumology Dept, Respiratory Institute, Hospital Clinic of Barcelona, Barcelona, Spain. ³Department of Anesthesia and Critical Care, ASST Ovest Milanese Ospedale Civile di Legnano, Milan, Italy. ✉email: atorres@recerca.clinic.cat; lfern1@recerca.clinic.cat

Despite its advantages, 16 S rRNA gene sequencing is not without limitations. Challenges such as **PCR amplification biases**, contamination from low-biomass samples, and limited functional insights necessitate careful experimental design and rigorous analytical approaches⁵. Research has demonstrated that the respiratory microbiome undergoes significant changes in response to various disease states, which often manifest as reduced microbial diversity and an increased abundance of pathogenic taxa¹. Additionally, the accuracy and resolution of 16 S rRNA gene sequencing are influenced by the choice of sequencing platform and bioinformatics pipeline, highlighting the importance of platform selection for respiratory microbiome studies⁶.

Among the available NGS platforms, the **Illumina NextSeq system** and **Oxford Nanopore Technologies (ONT)** platforms have emerged as some of the most widely used for 16 S rRNA gene sequencing in respiratory microbiome research. The Illumina systems has long been the benchmark for high-accuracy sequencing, producing millions of **short, paired-end reads (~300 bp)** with an error rate of less than 0.1%². These platforms typically target hypervariable regions of the 16 S rRNA gene, such as V3–V4, providing reliable genus-level classification. However, its short-read length limits its ability to resolve closely related bacterial species, which is a critical limitation in complex microbial environments like the lower respiratory tract⁷.

In contrast, **ONT platforms**, such as the MinION, utilize nanopore technology to generate **long reads** that can span the entire 16 S rRNA gene (~1,500 bp), enabling species-level and even strain-level resolution⁸. This long-read capability allows for a more comprehensive view of microbial diversity and enhances the resolution of closely related taxa. However, ONT platforms are traditionally associated with **higher error rates (5–15%)**, though recent advancements in base-calling algorithms and error-correction tools have significantly improved their accuracy⁹.

Both platforms have demonstrated utility in studying the lower respiratory microbiome. **Illumina NextSeq's high throughput** and accuracy make it well-suited for large-scale population studies and clinical applications where reproducibility and depth are critical. Conversely, **ONT's long-read capability** is particularly advantageous for studies that require species-level identification and rapid, field-based sequencing applications⁶. The selection of a sequencing platform for respiratory microbiome analysis is thus often dictated by the specific research goals, such as the need for species-level resolution, sample size, and available resources.

This study aims to provide a comprehensive comparison of the Illumina NextSeq and ONT platforms in the context of **16 S rRNA gene sequencing for lower respiratory microbiome analysis**. By critically assessing the strengths and limitations of each platform, this work seeks to offer insights into the optimal platform selection for future studies focused on unravelling the complex interactions between the respiratory microbiome and human health.

Materials and methods

Sample collection and DNA extraction

A total of 34 respiratory samples were collected for 16 S rRNA gene analysis. These included human specimens from ventilator-associated pneumonia (VAP) patients ($n=20$) from Hospital Clinic of Barcelona and samples from an experimental swine model of VAP ($n=14$). All samples were stored at $-80\text{ }^{\circ}\text{C}$ immediately upon collection and later processed in parallel using two sequencing platforms: Illumina NextSeq targeting the V3–V4 region and ONT MinION Mk1C covering the full-length 16 S rRNA gene.

Genomic DNA was extracted from approximately 1 mL of sample using the Sputum DNA Isolation Kit, SKU 46,200 (Norgen Biotek, Ontario, Canada), following the manufacturer's instructions with modifications to optimize DNA yield and purity. DNA quality and concentration were assessed using a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA) and a Qubit 4 fluorometer (Thermo Fisher Scientific, Massachusetts, USA).

Library Preparation and sequencing

Illumina sequencing

For Illumina sequencing, DNA libraries of V3–V4 hypervariable region of the 16 S rRNA gene were prepared using [QIAseq 16 S/ITS Region Panel (Qiagen, Hilden, Germany)]. The following amplification program was used: denaturation at $95\text{ }^{\circ}\text{C}$ for 5 min; 20 cycles of denaturation at $95\text{ }^{\circ}\text{C}$ for 30 s; primer annealing at $60\text{ }^{\circ}\text{C}$ for 30 s, extended to $72\text{ }^{\circ}\text{C}$ for 30 s and final elongation at $72\text{ }^{\circ}\text{C}$ for 5 min. Subsequently, another amplification was performed to attach the QIAseq 16 S/ITS Index barcodes (Qiagen, Hilden, Germany). QIAseq 16 S/ITS Smart Control (Qiagen, Hilden, Germany) is an unclassifiable synthetic DNA which was used as a positive control for library construction steps.

Following Qiagen's ISO-certified Quality Management System, each QIAseq 16 S/ITS Region Panel, QIAseq 16 S/ITS Index Kit and QIAseq 16 S/ITS Smart Control Kit was also tested to predetermined specifications to ensure consistent product quality. Finally, the pool of DNA products obtained during library preparation were sequenced on an NextSeq Sequencing Platform (Illumina, San Diego, CA, USA) to generate paired-end reads with a read length of $[2 \times 300\text{ bp}]$.

Following Qiagen's ISO-certified Quality Management System, each QIAseq 16 S/ITS Region Panel, QIAseq 16 S/ITS Index Kit and QIAseq 16 S/ITS Smart Control Kit was also tested to predetermined specifications to ensure consistent product quality. Finally, the pool of DNA products obtained during library preparation were sequenced on an NextSeq Sequencing Platform (Illumina, San Diego, CA, USA) to generate paired-end reads with a read length of $[2 \times 300\text{ bp}]$.

Nanopore sequencing

The extracted DNA was used to prepare sequencing libraries with the Oxford Nanopore Technologies (ONT) 16 S Barcoding Kit 24 V14 (SQK-16S114.24), following the manufacturer's protocol (Oxford Nanopore Technologies, UK). Barcoded libraries were pooled and loaded onto a MinION flow cell (R10.4.1). Sequencing was performed using MinKNOW software (v24.02.16) onboard the MinION Mk1C until end of life of the flow cell (72 h).

Data preprocessing and quality control

Illumina data

Illumina data was processed using **nf-core/ampliseq** version 2.11.0^{10,11}, part of the nf-core collection of workflows¹², utilizing reproducible environments from the Bioconda¹³ and Biocontainers projects¹⁴. Default parameters were applied unless specified otherwise. FastQC was used to evaluate sequence quality, and results were summarized with **MultiQC**¹⁵. Primer trimming was performed with **Cutadapt**¹⁶, discarding sequences lacking primers. Sequences not meeting quality thresholds were filtered out.

Sequences were processed using DADA2¹⁷ with default filtering parameters to remove PhiX contamination, trim reads, discard sequences exceeding expected error thresholds, correct errors, merge paired-end reads and remove PCR chimeras. Filtered amplicon sequencing variants (ASVs) were retained for taxonomic classification using the “Silva 138.1 prokaryotic SSU” database.

Nanopore data

Raw reads generated by MinKNOW were basecalled and demultiplexed using the Dorado basecaller (v7.3.11), integrated into MinKNOW v24.02.16. The High Accuracy (HAC) model was used with default parameters unless otherwise specified. Reads were automatically demultiplexed based on barcode sequences, while adapter removal and initial quality filtering were handled by MinKNOW. Filtered reads were exported in FASTQ format for further analysis.

Post-sequencing, reads were processed using the EPI2ME Labs 16 S Workflow¹⁸. This pipeline included additional quality control, read filtering, and taxonomic classification against the “Silva 138.1 prokaryotic SSU” database. The resulting dataset consisted of high-quality reads suitable for taxonomic classification and downstream analyses.

Downstream analysis

Data analysis

All downstream analyses were performed in **R v4.2.0** using custom scripts in **RMarkdown**, ensuring reproducibility with packages such as **phyloseq**¹⁹, **tidyverse**²⁰, **vegan**²¹, **ANCOMBC**²², and visualization tools like **ggplot2**²³ and **viridis**²⁴.

OTU abundance tables were processed into **phyloseq** objects by combining taxonomic and sample metadata. Taxonomy ranks were extracted and organized from **Kingdom** to **Genus**, ensuring structured metadata for downstream analyses. Additionally, for ONT data we also evaluated its performance for different experimental conditions, specifically duration of sequencing (12 h, 18 h, 24 h, 72 h). Missing values were replaced with zeros.

Diversity analysis

In order to better understand how each platform performed, we first evaluated for each metric, the complete datasets (Illumina vs. Nanopore at 72 h). Alpha diversity was assessed using both Shannon diversity and Observed features indices to evaluate species richness and evenness across sequencing platforms (Illumina and Nanopore) and sample groups (Human and Pig).

Alpha diversity metrics were calculated using **phyloseq**, focusing on **Observed OTUs** and **Shannon Diversity**. **Dunn's test** with Bonferroni correction and **Kruskal-Wallis tests** evaluated differences among groups. Spearman rank correlations were calculated between the diversity indices and total read counts. Diversity results were visualized using **ggplot2**, with boxplots comparing Observed OTUs and Shannon Diversity across experimental groups. Scatterplots showing diversity indices versus total read counts were also generated, with linear regression lines indicating overall trends.

Beta diversity was assessed using **Bray-Curtis** and **Jaccard** distances calculated with the **phyloseq** and **vegan** R packages. Principal Coordinates Analysis (PCoA) was performed, and beta diversity plots were generated using **ggplot2**. Permutational Multivariate Analysis of Variance (PERMANOVA) was conducted using **adonis2()** from **vegan** to evaluate group differences based on metadata variables. Pairwise Adonis tests (**pairwise.adonis()**) were performed to detect significant differences between specific sample groups.

Taxonomic comparison

Taxonomic composition was compared between Illumina and ONT datasets by identifying the top five most abundant taxa per sample and their taxonomic classification. Each **phyloseq** object was normalized. Results were stored in data frames for subsequent analysis. Stacked bar plots were generated using **ggplot2**, visualizing the top taxa distribution across samples. Pairwise comparisons between Illumina and ONT taxa were conducted using the Wilcoxon signed-rank test.

Differential abundance analysis was conducted using **ANCOM-BC2** (Analysis of Compositions of Microbiomes with Bias Correction 2) to identify taxa with significant compositional differences across groups. Custom contrast matrices were used to test global, pairwise, and trend-based differences. The analysis incorporated structural zero handling, a minimum prevalence threshold (**prv_cut** = 0.10), and Holm-Bonferroni adjustments for multiple testing correction. ANCOM-BC2 was applied using the **ANCOMBC** R package, and significance was determined using false discovery rate (FDR) correction at $q < 0.05$.

Results

Alpha diversity analysis

Shannon indices in the Human dataset sequenced with **Illumina** ranged from 0.19 to 2.81, with Observed OTUs counts varying from 70 to 121. In the Pig dataset, Shannon indices ranged between 0.59 and 2.83, with feature counts from 68 to 118, indicating relatively balanced microbial richness and evenness (Fig. 1).

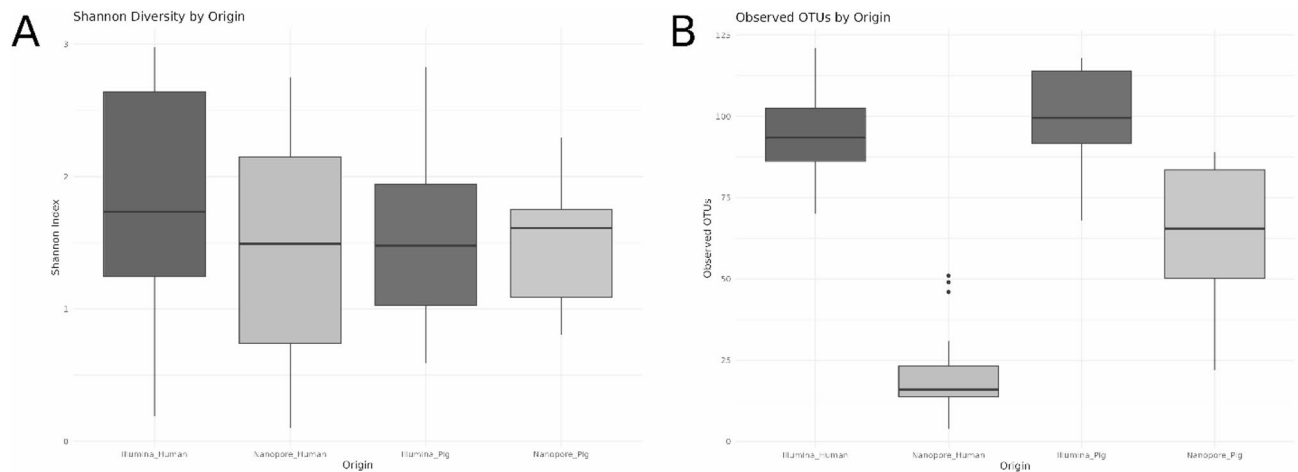


Fig. 1. Boxplots representing alpha diversity indices stratified by sequencing platform and sample group (Human or Pig). **A)** Boxplot of Shannon diversity. The mean diversity values for each group are 1.841, 1.504, 1.526, and 1.474, with a p-value of 0.4807, indicating no statistically significant differences between groups. **B)** Boxplot of observed OTUs. The mean observed OTU counts for each group are 93.40, 20.75, 100.04, and 64.29, with a p-value < 0.0001, suggesting statistically significant differences between groups.

Shannon indices in the Human dataset sequenced with **Nanopore** varied from 0.10 to 2.75, with Observed OTUs counts ranging from 4 to 51, reflecting a lower alpha diversity vs. Illumina. For the Pig dataset, Shannon indices ranged between 0.80 and 2.29, with feature counts from 22 to 89, showing moderately high richness in some samples (Fig. 1).

Platform comparisons revealed that Illumina sequencing consistently captured higher Shannon indices and Observed OTUs across both Human and Pig samples, suggesting greater sensitivity for alpha diversity metrics.

Comparisons between Human and Pig samples showed that Pig samples exhibited higher alpha diversity on average compared to Human samples, regardless of sequencing platform, indicating potentially greater microbial richness or evenness in the Pig microbiome.

Alpha diversity by timepoints

Alpha diversity metrics comparing Illumina and Nanopore sequencing methods at various time points (12, 18, 24, and 72 h) were evaluated for each dataset across groups.

In the **Human** dataset, significant differences in **Observed OTUs** were identified consistently across groups and time points (Supplementary Material Table 1). In contrast, the **Shannon diversity** metric showed no significant differences between sequencing platforms across groups, with p-values consistently above 0.13. This indicates that, while Illumina sequencing captured greater species richness, community evenness remained comparable between platforms. No differences were found between Nanopore timepoints.

In the **Pig** dataset, the **Observed OTU** metric similarly revealed significant differences, with Illumina sequencing showing greater richness (Supplementary Material Table 2). The **Shannon diversity** metric, as in the human dataset, did not show significant differences across groups or time points, with p-values consistently above 0.5. This indicates platform-specific variations in richness but not evenness, consistent with the Human results. As with the previous dataset, no differences were found between Nanopore timepoints.

Across both datasets, Illumina sequencing demonstrated significantly higher species richness as measured by **Observed OTU**, while **Shannon diversity** revealed no significant differences in community evenness between platforms.

Beta diversity analysis

We first constructed the similarity matrix for each complete dataset (Illumina vs. Nanopore at 72 h) using Bray-Curtis and Jaccard methods. With the PERMANOVA analysis we then evaluated the effects of host type and sequencing platform on microbial beta diversity on each created matrix (Fig. 2).

For both Bray-Curtis and Jaccard metrics, host type and sequencing platform collectively explained 32.7% of the variation in microbial beta diversity ($R^2 = 0.327$). The remaining 67.3% of variation ($R^2 = 0.673$) was attributed to within-group differences, likely arising from individual variability or noise. PERMANOVA results yielded an F-statistic of 10.37 and a p-value of 0.001 for both metrics.

Pairwise comparison results (illumina vs. nanopore for each host type)

For Bray-Curtis the microbial beta diversity between Illumina and Nanopore sequencing in **Human** samples shows no significant difference ($R^2=0.049$, $F=1.97$, $p=0.306$). Similarly, no significant difference is observed with the Jaccard metric ($R^2=0.035$, $F=1.39$, $p=0.774$) (Supplementary Material Fig. 1 A).

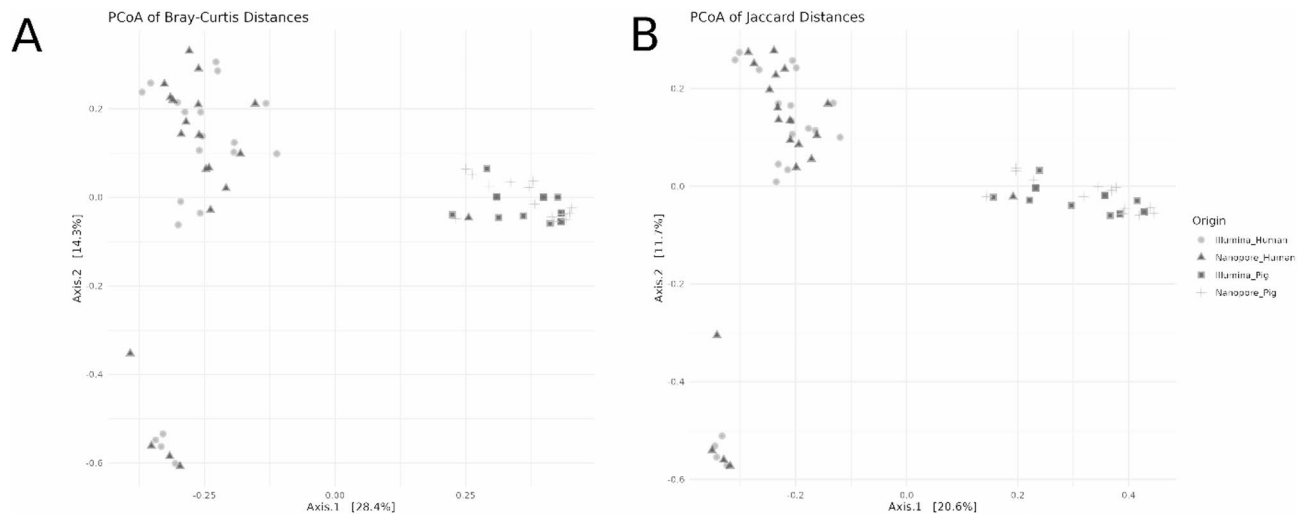


Fig. 2. Principal Coordinates Analysis (PCoA) of beta diversity distances stratified by sequencing platform and sample group (Human or Pig). **(A)** Bray-Curtis distances: PERMANOVA results indicate no effect of the model ($Df=3$, $SumOfSqs=8.060$, $R^2=0.327$, $F=10.373$, $p=0.001$), with residual variation accounting for $R^2=0.673$. **(B)** Jaccard distances: PERMANOVA also shows no significant model effect ($Df=3$, $SumOfSqs=8.060$, $R^2=0.327$, $F=10.373$, $p=0.001$), with residual variation accounting for $R^2=0.673$. Both analyses were conducted using 999 permutations.

For Bray-Curtis in **Pig** samples, there is a significant difference between Illumina and Nanopore sequencing ($R^2=0.214$, $F=7.07$, $p=0.006$). Similarly, the Jaccard metric also reveals significant differences ($R^2=0.165$, $F=5.15$, $p=0.006$) (Supplementary Material Fig. 1B).

The general PERMANOVA results confirm that host type and sequencing platform collectively explain a significant proportion of the variation in microbial beta diversity. However, when focusing specifically on the sequencing platform we find that in Human samples, the sequencing platform (Illumina vs. Nanopore) does not significantly impact microbial community composition, as indicated by both Bray-Curtis and Jaccard metrics. In Pig samples, the sequencing platform has a significant effect on microbial beta diversity, suggesting that the platform may influence the resolution or detection of microbial community differences in this host type.

Beta diversity by timepoints

Similarly to what happened with the complete dataset, the human samples, across all time points (12 h, 18 h, 24 h, 72 h), Bray-Curtis and Jaccard metrics showed low R^2 values (~ 0.028 – 0.049) and non-significant p -values ($p \geq 0.44$). Additionally, microbial beta diversity was stable within Nanopore samples over time ($R^2 < 0.01$, $p=1$) (Supplementary Material Table 3).

Regarding the pig samples, both Bray-Curtis and Jaccard metrics indicated significant differences between Illumina and Nanopore sequencing ($R^2 \sim 0.12$ – 0.21 , $p=0.01$) across all timepoints. Additionally, within Nanopore comparisons revealed stable beta diversity across time points ($R^2 < 0.01$, $p=1$) (Supplementary Material Table 4).

Top taxa comparison

Taxonomic profiles of the **Human** dataset revealed differences in relative abundances of dominant taxa between Illumina and Nanopore sequencing platforms (Fig. 3A). Across all samples, *Staphylococcus*, *Streptococcus*, *Prevotella*, and *Neisseria* were consistently identified as dominant taxa. Less abundant taxa, including *Cutibacterium*, and *Aggregatibacter*, were consistently detected by both platforms but showed platform-specific differences. The “Other” category, representing taxa outside the top five most abundant, was higher in Illumina across samples. This suggests a broader classification range for certain reads in Illumina; however, this broader range may also increase the risk of false-positive identifications due to the limited context provided by short reads. Studies have shown that short-read classifiers often misclassify low-abundance taxa and require heavy filtering to reduce these errors, which can compromise recall²⁵. Conversely, Nanopore consistently reported higher proportions of the “Unknown” taxa, which may reflect differences in database alignment or limitations in reference libraries.

Analysis of the **Pig** dataset similarly highlighted platform-specific differences in taxonomic profiles (Fig. 3B). *Pseudomonas*, *Mycoplasma*, and *Bacteroides* were identified as dominant taxa across samples, with relative abundances varying between platforms. *Mycoplasma* was consistently more abundant in Illumina across multiple samples. As in Human, the “Other” category showed higher proportions in Illumina. Less abundant taxa, like *Bacteroides*, *Veillonella*, and *Streptococcus* show higher relative abundances in Illumina data compared to Nanopore.

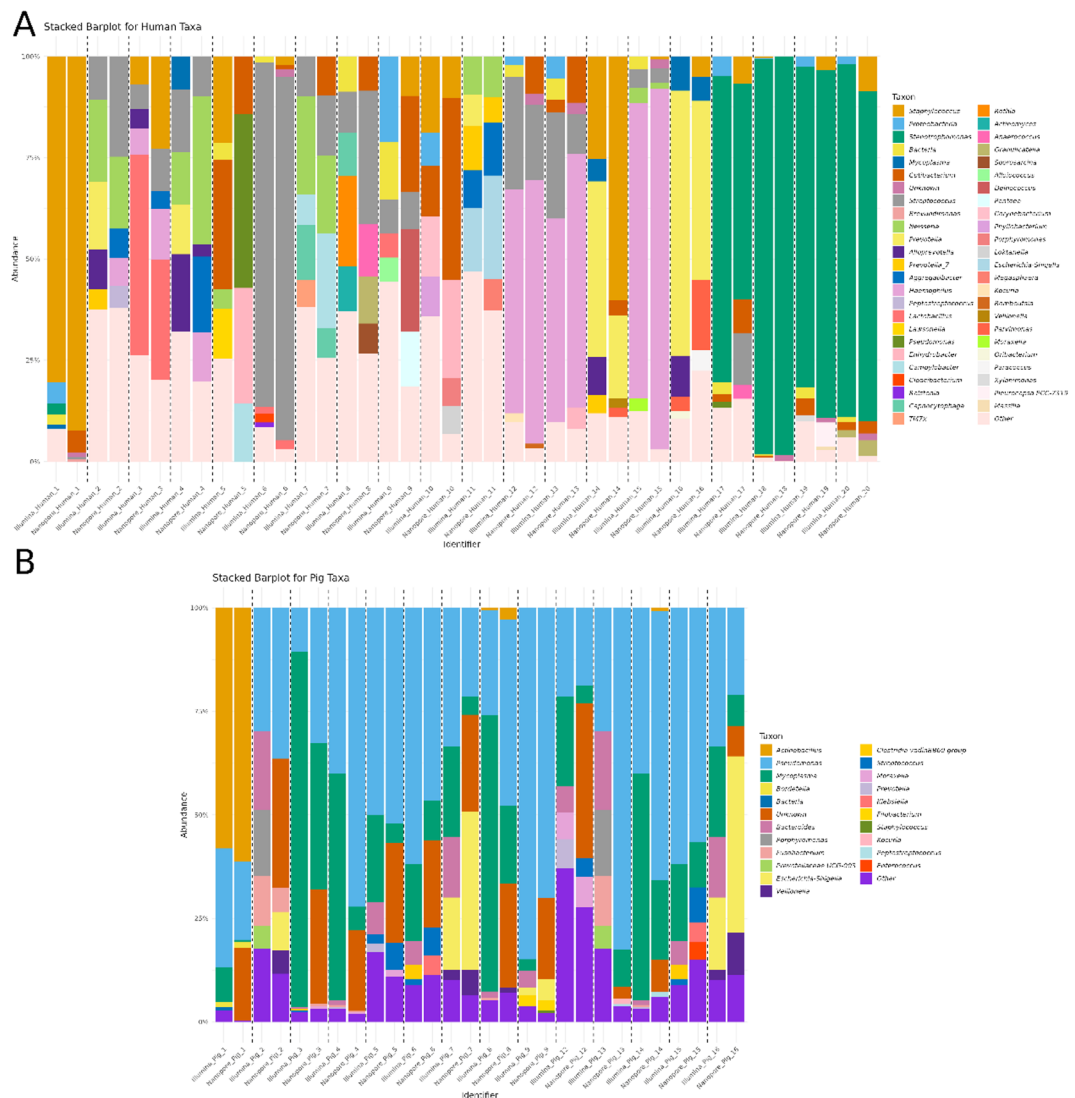


Fig. 3. Stacked barplots displaying the five most abundant taxa in each sample, resolved to the genus level when possible, prioritizing the most specific taxonomic classification. **(A)** This graph represents Human samples collected from ventilator-associated pneumonia (VAP) patients, with a diverse range of pathogens identified as the causative agents of infection. The taxa composition reflects the microbial communities associated with this heterogeneous infection. **(B)** This graph represents an experimental VAP model in swine, inoculated with *Pseudomonas aeruginosa*, a common pathogen associated with VAP. The barplot highlights the microbial community dynamics in a controlled setting where *Paeruginosa* is the primary pathogen.

ANCOM-BC2

In the **Human** dataset, ANCOM-BC2 revealed significant differences in relative taxonomic abundances between Illumina and Nanopore sequencing platforms. Around 10% of the total taxa present in the Human dataset were significant or present in enough samples to be included in the analysis. Many taxa exhibited negative log fold changes (lfc) across Nanopore time points (12, 18, 24, 72 h) compared to the Illumina baseline, indicating lower relative abundance in Nanopore samples (Fig. 4A). Several taxa exhibited significant changes in abundance ($q < 0.05$), including *Prevotella*, *Pseudomonas*, and *Cutibacterium*. These taxa displayed consistent declines across Nanopore time points, highlighting potential platform-specific biases or true biological effects (Supplementary Material Table 5).

Taxa flagged as failing sensitivity analysis (passed_ss_ = FALSE) require cautious interpretation. For instance, *Stenotrophomonas* exhibited a mild decrease (lfc = -0.76 at Nanopore12) but failed sensitivity testing, suggesting unreliable results. Similarly, *Veillonella* showed slight variations but was not robust under sensitivity analysis.

In the **Pig** dataset, significant changes in relative abundances were observed between Illumina and Nanopore platforms for Pig samples. Around 15% of the total taxa present in the Pig dataset were significant or present in enough samples to be included in the analysis. Many taxa exhibited large negative lfc values for Nanopore samples compared to the Illumina baseline (Fig. 4B). Declines in taxa such as *Bacteroides*, *Prevotella*, and *Alloprevotella* were consistent across all Nanopore time points. In contrast, taxa like *Clostridium sensu stricto* 1

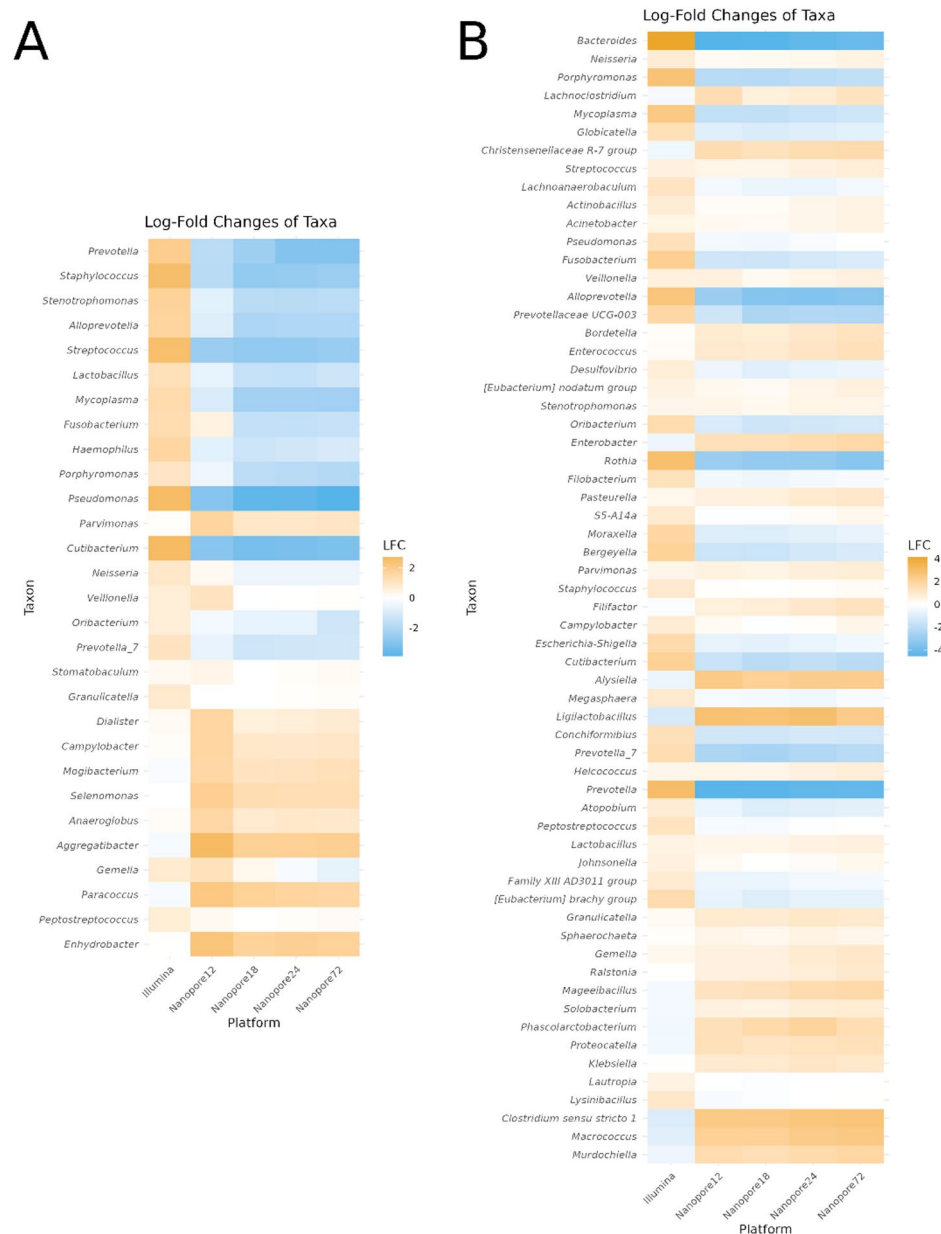


Fig. 4. Log-Fold Changes of Taxa Across Platform and Timepoints. Differential abundance of bacterial taxa is shown with log-fold change (LFC) values displayed along the y-axis for each taxon. The x-axis represents sequencing methods, with Illumina, and Nanopore at time points 12, 18, 24, and 72 h. LFC is represented using a color gradation, where yellow indicates enrichment and blue indicates depletion. **A)** This plot represents Human samples collected from ventilator-associated pneumonia (VAP) patients, with a diverse range of pathogens identified as the causative agents of infection. The taxa composition on the ANCOM-BC2 reflects the heterogeneity the human samples due to the lower number of taxa presents in the analysis. **B)** This plot represents an experimental VAP model in swine, inoculated with *Pseudomonas aeruginosa*, a common pathogen associated with VAP. The taxa composition on the ANCOM-BC2 reflects the similarity and standardization of the preclinical model microbiome.

and *Macroccoccus* exhibited significant increases over time. *Neisseria* exhibited minor changes across time points (lfc values between +0.23 and +0.56) but failed sensitivity screening, suggesting non-robust results. Similarly, *Globicatella* showed slight variations but was not statistically robust (Supplementary Material Table 6).

Discussion

The combined analysis of alpha diversity, beta diversity, taxonomic profiles, and differential abundance from the ANCOM-BC2 analysis provides valuable insights into the differences between Illumina and Nanopore sequencing platforms and their implications for respiratory microbiome studies both clinical and preclinical.

Each platform presents distinct strengths and limitations that must be considered when selecting an approach for microbiome analysis.

The assessment of alpha diversity revealed distinct patterns between the sequencing platforms. While Observed OTUs consistently showed significant differences, with Illumina capturing greater species richness, Shannon diversity indices were largely comparable between the two platforms. This suggests that, while Illumina excels in detecting a broader range of taxa, the evenness of microbial communities appears to remain unaffected by the sequencing platform. Such findings are consistent with previous studies reporting that Illumina often outperforms long-read technologies in detecting infrequent taxa but shows similarity in community evenness metrics²⁶. Interestingly, Pig samples displayed higher alpha diversity than Human samples across both platforms, aligning with known microbiome studies in livestock, where greater microbial richness is attributed to dietary complexity and environmental exposures²⁷. However, the reliance of Illumina on short reads may introduce biases in resolving complex genomic structures, whereas Nanopore's ability to sequence longer fragments provides improved genomic continuity but at the cost of higher error rates.

PERMANOVA results indicated that host type and sequencing platform collectively explained 32.7% of the variation in microbial beta diversity, with significant differences in Pig samples but not in Human samples. Similar findings have been reported in other studies, where beta diversity metrics like Bray-Curtis and Jaccard dissimilarities revealed platform-dependent differences in community composition, particularly in more complex microbiomes²⁸. In Human samples, however, platform-driven differences in beta diversity were not statistically significant, indicating that microbial community composition may be less affected by sequencing platform in less diverse or simpler microbiomes^{26,29}. Nanopore's greater variability in beta diversity estimates may limit its precision in distinguishing subtle microbial shifts, a drawback that must be considered for studies requiring high-resolution community profiling.

The discrepancies in taxonomic profiles between Illumina and Nanopore platforms highlight the influence of sequencing technology on microbial community analyses. Illumina captured a broader diversity of taxa, often assigning higher proportions of reads to "Other" categories, consistent with its reported sensitivity for rare taxa²⁷. Conversely, Nanopore assigned higher proportions of reads to dominant taxa like *Staphylococcus* and *Prevotella* and identified unique taxa such as *Enterococcus*, corroborating findings that Nanopore's long-read sequencing offers better resolution for dominant taxa³⁰. For Pig samples, Nanopore identified unique taxa like *Enterococcus* and *Klebsiella* that were less prominent in Illumina data. This may highlight Nanopore's strength in resolving certain microbial groups but also raises questions about its amplification biases. While this highlights the advantage of Nanopore in dominant taxa resolution, its susceptibility to homopolymer errors and lower per-base accuracy must be accounted for in taxonomic assignments. Such platform-specific differences could have significant implications for studies focused on microbial dynamics and require careful interpretation.

The ANCOM-BC2 analysis revealed significant declines in taxa such as *Prevotella* and *Bacteroides* in Nanopore samples, reflecting platform biases against these taxa. Similar patterns have been observed in studies comparing Illumina and Nanopore sequencing, where taxonomic resolution and amplification biases differ between platforms²⁸. Conversely, the increased abundance of taxa like *Clostridium sensu stricto 1* and *Enhydrobacter* in Nanopore data underscores its propensity for platform-specific overrepresentation of certain taxa²⁷. Sensitivity analyses confirmed the robustness of trends for major taxa like *Prevotella*, reinforcing the reliability of these observations. However, taxa flagged as failing sensitivity tests require cautious interpretation, as platform errors may confound their biological relevance. Given these discrepancies, studies requiring high taxonomic accuracy may benefit from Illumina, while those prioritizing long-read advantages should implement error correction strategies in Nanopore analyses.

These findings emphasize the importance of aligning sequencing platform selection with study objectives. Illumina's ability to detect a wide array of taxa makes it ideal for comprehensive microbiome profiling, while Nanopore's advantages in dominant taxa resolution and real-time sequencing lend it strength for targeted applications such as pathogen surveillance²⁶. However, platform-specific biases, particularly for taxa like *Prevotella* and *Bacteroides*, underscore the need for caution in interpreting data from a single method³⁰.

The inclusion of both Human and Pig respiratory samples enables a comparative assessment of sequencing platforms across microbiomes of differing complexity. Humans represent a clinical and real-world model, whereas Pigs provide a preclinical and controlled system for studying microbial dynamics^{31,32}. Pigs harbor a more diverse respiratory microbiome due to environmental exposure and a permissive immune system, while Human samples exhibit lower diversity, likely due to host immune modulation³³. This comparative framework underscores the need for sequencing strategies tailored to microbiome complexity, ensuring methodological choices align with host-specific microbial characteristics.

Future research should prioritize newer approaches that integrate multi-omics techniques, improve Nanopore's error correction algorithms, and standardize bioinformatics workflows to minimize platform biases. Including standardized mock communities alongside real-world samples could also help disentangle platform-specific performance from biological variation. Recent advancements in hybrid sequencing, combining short and long reads, have shown promise in enhancing taxonomic resolution and mitigating individual platform limitations³⁴. These methods could facilitate cross-platform validation and improve the reproducibility of microbiome studies. Ultimately, leveraging hybrid approaches may offer the most comprehensive solution, mitigating platform-specific drawbacks while enhancing the accuracy and depth of microbial community analyses.

To conclude, while Illumina remains the preferred technology for general microbiome studies due to its accuracy and taxonomic breadth, Nanopore's advantages in real-time sequencing and dominant taxa resolution make it an attractive alternative for specific applications, particularly where rapid sequencing is required.

Data availability

Data are provided within the manuscript or supplementary information files. Raw sequencing data have been deposited in the NCBI repository under the accession number PRJNA1253920. The data have also been deposited in Zenodo under the DOI <https://doi.org/10.5281/zenodo.15175207> and are available from the corresponding author on reasonable request.

Received: 28 May 2025; Accepted: 3 September 2025

Published online: 29 September 2025

References

- Cribbs, S. K. et al. Jan., 'Correlation of the lung microbiota with metabolic profiles in bronchoalveolar lavage fluid in HIV infection', *Microbiome*, vol. 4, no. 1, pp. 1–11, (2016). <https://doi.org/10.1186/S40168-016-0147-4/TABLES/3>
- Johnson, J. S. et al. 'Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis', *Nature Communications* vol. 10, no. 1, pp. 1–11, Nov. 2019, (2019). <https://doi.org/10.1038/s41467-019-13036-1>
- López-Aladid, R. et al. Mar., 'Determining the most accurate 16S rRNA hypervariable region for taxonomic identification from respiratory samples', *Scientific Reports* 2023 13:1, vol. 13, no. 1, pp. 1–10, (2023). <https://doi.org/10.1038/s41598-023-30764-z>
- Earl, J. P. et al. Oct., 'Species-level bacterial community profiling of the healthy sinonasal microbiome using Pacific Biosciences sequencing of full-length 16S rRNA genes', *Microbiome*, vol. 6, no. 1, pp. 1–26, (2018). <https://doi.org/10.1186/S40168-018-0569-2/METRICS>
- Louca, S., Doebeli, M. & Parfrey, L. W. 'Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem', *Microbiome*, vol. 6, no. 1, pp. 1–12, Feb. (2018). <https://doi.org/10.1186/S40168-018-0420-9/FIGURES/4>
- Fromentin, M. et al. 'The 16S rRNA lung Microbiome in mechanically ventilated patients: a methodological study', *Exp. Lung Res.* **48** (1), 23–34. <https://doi.org/10.1080/01902148.2021.2021327> (2022).
- Rintala, A. et al. Apr., 'Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor', *J Biomol Tech*, vol. 28 1, no. 1, pp. 19–30, (2017). <https://doi.org/10.7171/JBT.17-2801-003>
- 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** (1), 1–13. <https://doi.org/10.1186/S12866-021-02094-5/FIGURES/6> (Dec. 2021).
- Cusco, A., Catozzi, C., Vines, J., Sanchez, A. & Francino, O. 'Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and the 16S-ITS-23S of the rrn operon', *F1000Res*, vol. 7, Oct. (2018). <https://doi.org/10.1101/450734>
- Straub, D. et al. 'nf-core/ampliseq: ampliseq version 2.11.0', <https://doi.org/10.5281/ZENODO.14011895>
- Straub, D. et al. 'Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline', *Front Microbiol*, vol. 11, p. 550420, Oct. (2020). <https://doi.org/10.3389/FMICB.2020.550420/BIBTEX>
- Ewels, P. A. et al. Feb., 'The nf-core framework for community-curated bioinformatics pipelines', *Nature Biotechnology* 2020 38:3, vol. 38, no. 3, pp. 276–278, (2020). <https://doi.org/10.1038/s41587-020-0439-x>
- Dale, R. et al. Jul., 'Bioconda: sustainable and comprehensive software distribution for the life sciences', *Nat Methods*, vol. 15, no. 7, p. 475, (2018). <https://doi.org/10.1038/S41592-018-0046-7>
- Da Veiga, F. et al. Aug., 'BioContainers: an open-source and community-driven framework for software standardization', *Bioinformatics*, vol. 33, no. 16, p. 2580, (2017). <https://doi.org/10.1093/BIOINFORMATICS/BTX192>
- Ewels, P., Magnusson, M., Lundin, S. & Käller, M. 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, vol. 32, no. 19, pp. 3047–3048, Oct. (2016). <https://doi.org/10.1093/BIOINFORMATICS/BTW354>
- Martin, M. 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet J.* **17** (1), 10–12. <https://doi.org/10.14806/EJ.17.1.200> (May 2011).
- Callahan, B. J. et al. 'DADA2: high resolution sample inference from illumina amplicon data', *Nat. Methods*. **13** (7), 581. <https://doi.org/10.1038/NMETH.3869> (Jun. 2016).
- 'Oxford Nanopore Technologies EPI2ME wf-16s'. Accessed: Jan. 23, 2025. [Online]. Available: <https://labs.epi2me.io/workflows/wf-16s/>
- McMurdie, P. J. & Holmes, S. 'Phyloseq: an R package for reproducible interactive analysis and graphics of Microbiome census data', *PLoS One*. **8** (4), e61217. <https://doi.org/10.1371/JOURNAL.PONE.0061217> (Apr. 2013).
- Wickham, H. et al. 'Welcome to the tidyverse', *J. Open. Source Softw.* **4** (43), 1686. <https://doi.org/10.21105/JOSS.01686> (Nov. 2019).
- Oksanen, J. et al. 'vegan community ecology package version 2.5-7 November 2020', (2020).
- Lin, H. & Das Peddada, S. 'Analysis of compositions of microbiomes with bias correction', *Nature Communications* 2020 11:1, vol. 11, no. 1, pp. 1–11, Jul. (2020). <https://doi.org/10.1038/s41467-020-17041-7>
- Wickham, H. 'Data Analysis', pp. 189–201, (2016). https://doi.org/10.1007/978-3-319-24277-4_9
- Garnier et al. 'viridis(Lite) - Colorblind-Friendly Color Maps for R', (2024). <https://doi.org/10.5281/zenodo.4679423>
- 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 Bioinform.* **23** (1), 1–39. <https://doi.org/10.1186/S12859-022-05103-0/FIGURES/10> (Dec. 2022).
- Szoboszlay, M. et al. 'Nanopore Is Preferable over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required', *Microorganisms*, vol. 11, no. 3, Mar. (2023). <https://doi.org/10.3390/MICROORGANISMS11030804>
- Heikema, A. P. et al. Sep., 'Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota', *Genes (Basel)*, vol. 11, no. 9, pp. 1–17, (2020). <https://doi.org/10.3390/GENES11091105>
- Stoeck, T., Katzenmeier, S. N., Breiner, H. W. & Rubel, V. 'Nanopore duplex sequencing as an alternative to illumina miseq sequencing for eDNA-based biomonitoring of coastal aquaculture impacts', *Metabarcoding Metagenom.* **8**, 45–68. <https://doi.org/10.3897/MBMG.8.121817> (2024).
- Sinclair, L., Osman, O. A., Bertilsson, S. & Eiler, A. 'Microbial Community Composition and Diversity via 16S rRNA Gene Amplicons: Evaluating the Illumina Platform', *PLoS One*, vol. 10, Oct. (2014). <https://doi.org/10.1101/010058>
- Yeo, K. et al. 'A comparison between full-length 16S rRNA Oxford nanopore sequencing and illumina V3-V4 16S rRNA sequencing in head and neck cancer tissues', *Arch. Microbiol.* **206** <https://doi.org/10.1101/2024.03.08.584026> (Mar. 2024).
- Zhao, L., Luo, J. L., Ali, M. K., Spiekeroetter, E. & Nicolls, M. R. 'The Human Respiratory Microbiome: Current Understandings and Future Directions', *Am J Respir Cell Mol Biol*, vol. 68, no. 3, pp. 245–255, Mar. (2023). <https://doi.org/10.1165/RCMB.2022-0208TR>
- Pirola, M., Espinosa-Gongora, C., Bogaert, D. & Guardabassi, L. 'The Porcine respiratory microbiome: recent insights and future challenges', *Anim. Microbiome*. **3** (1). <https://doi.org/10.1186/S42523-020-00070-4> (Dec. 2021).
- Moor, J. et al. 'Influence of pig farming on human Gut Microbiota: role of airborne microbial communities', *Gut Microbes*, vol. 13, no. 1, pp. 1–13, (2021). <https://doi.org/10.1080/19490976.2021.1927634>

34. Dommann, J. et al. 'A novel barcoded nanopore sequencing workflow of high-quality, full-length bacterial 16S amplicons for taxonomic annotation of bacterial isolates and complex microbial communities', *mSystems*, vol. 9, Oct. (2024). <https://doi.org/10.1128/MSYSTEMS.00859-24>

Author contributions

G.M. and A.S.C. performed sequencing and quality control of Illumina and Oxford Nanopore datasets. G.M. conducted bioinformatic analyses, including taxonomic classification, alpha and beta diversity assessments, and differential abundance testing. A.P., A.M., B.L., J.C.R., L.B.F., D.C., K.K., and R.C. were responsible for sample collection, processing, and metadata curation. M.F., A.T., and L.F.B. conceptualized the study, supervised the research, and provided critical revisions to the manuscript. G.M. drafted the manuscript, with input from all authors. All authors reviewed and approved the final manuscript.

Funding

The study was supported by the ISCIII-FEDER grant (Code: PI21/00725) awarded to LFB and AT; PYRAMID (Code: PID2022-141924OB-I00) – Agencia Estatal de Investigación to MF; 06/06/0028/CIBER de Enfermedades Respiratorias (CIBERES), an initiative of ISCIII; 2.603/IDIBAPS; HOMILUNG (part of the Horizon Europe programme under grant agreement No. 101137148); La Caixa Health Research 2018 (HR18-00058), co-led by AT; the ICREA Academy award to AT; SGR 01148 (2021) from Generalitat de Catalunya to LFB; and SEPAR grants (Code: 1536) awarded to LFB. Funders had no role in the project design, data collection, data analysis, interpretation, or writing of the manuscript.

Declarations

Human ethics

This study involving human participants was approved by the Internal Review Board of Hospital Clinic, Barcelona (registry number HCB/2021/0343). Written informed consent was obtained from all participants. All methods involving human subjects were performed in accordance with the relevant institutional guidelines and regulations.

Competing interests

A. Torres has received grants from MedImmune, Cubist, Bayer, Theravance and Polyphor, and personal fees as an advisory board member from Bayer, Roche, The Medicines CO and Curetis. He has received bureau fees for keynote speaker presentations from GSK, Pfizer, Astra Zeneca and Biotest Advisory Board, but these were not associated with the study described in this paper. All the remaining authors declare no conflict of interest.

Animal ethics

Animal experiments were conducted in accordance with the European guidelines for the care and use of laboratory animals. The study was approved by the Institutional Animal Care and Use Committee of the University of Barcelona (approval number 159/20). All animal procedures were performed in accordance with institutional and national regulations, and the study is reported in accordance with the ARRIVE guidelines.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-18768-3>.

Correspondence and requests for materials should be addressed to A.T. or L.F.-B.

Reprints and permissions information is available at 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) 2025