UNIVERSITY OF HELSINKI

Quantitative microbiota profiling: Comparison of computationally predicted profiles against quantitative PCR-based reference

Master’s Programme in Microbiology and Microbial Biotechnology
Master's thesis

Author:
M Nugrah Fadillah

Supervisor(s):
Docent Anne Salonen
Dr. Ching Jian
Tuomas Heini

06.2026
Helsinki

Faculty: Faculty of Agriculture and Forestry
Degree programme: Master's Programme in Microbiology and Microbial Biotechnology
Study track: n.a
Author: M Nugrah Fadillah
Title: Quantitative microbiota profiling: Comparison of computationally predicted profiles against quantitative PCR-based reference
Level: Master's thesis
Month and year: 05.2026
Number of pages: 35
Keywords: quantitative microbiome profiling, microbial load, differential abundance analysis, compositional data analysis, gut microbiome
Supervisor or supervisors: Docent Anne Salonen, Dr. Ching Jian, Tuomas Heini
Where deposited: Helsinki University Library
Additional information: n.a

Abstract:

Microbiota profiles generated by high-throughput sequencing are compositional and do not capture bacterial load**.** Studies using quantitative microbiota profiling (QMP) have shown that variation in microbial load is a major source of microbiota variation and can confound disease associations. This thesis evaluated whether sequencing-based computational load proxies and differential abundance approaches can reproduce qPCR-based QMP in a longitudinal cohort. Fecal samples from 95 Finnish healthy adults traveling to Benin, West Africa, were collected before travel, immediately after travel, and one month later (287 samples in total). Shotgun metagenomic sequencing was combined with universal 16S rRNA gene qPCR for QMP. Two machine-learning models and the bacterial-to-host read count ratio derived from the metagenomic data were compared with qPCR-based total bacterial load. Family-level differential abundance results of bacterial taxa were compared across absolute, relative, and several common differential abundance testing methods using qPCR-based quantitative profiles as the reference. Both machine-learning models correlated weakly with qPCR-based bacterial load, and the bacterial-to-host ratio showed no significant correlation. Agreement in differential abundance results was highest when bacterial load was stable and lowest for the contrast with the largest qPCR-detected load decrease. Machine-learning-predicted absolute abundance performed better than most relative-abundance approaches but failed to reproduce the qPCR-based QMP reference. Among methods without measured bacterial load, total-sum-scaled log-transformed relative abundance with a linear mixed model showed the closest agreement. Methods such as centered log-ratio, LinDA, ANCOM-BC2, and DESeq2 showed more missed findings, method-specific findings, or direction errors. In longitudinal studies, or where substantial shifts in microbial load are expected, experimental load measurement is important for quantitative profiling. More complex normalization or statistical testing methods did not consistently improve agreement with quantitative microbiota profiles, and some produced misleading biological interpretations when bacterial load changed.

Abbreviations

ANCOM-BC2: Analysis of Compositions of Microbiomes with Bias Correction (version 2)

AST: Arcsine square-root transformation

B:H: Bacterial-to-host (read count ratio)

CLR: Centered log-ratio (transformation)

CN: Copy number

Ct: Threshold cycle value

DA: Differential abundance

ddPCR: digital droplet PCR

DESeq2: Differential expression analysis for sequence count data (version 2)

DNA: Deoxyribonucleic acid

ETEC: Enterotoxigenic *Escherichia coli*

FDR: False discovery rate

HTS: High-throughput sequencing

LinDA: Linear models for differential abundance analysis

LMM: Linear mixed model

MLP: Machine-learning prediction

PCR: Polymerase chain reaction

QMP: Quantitative microbiota profiling

qPCR: Quantitative (real-time) PCR

rRNA: Ribosomal ribonucleic acid

rrnDB: Ribosomal RNA operon copy-number database

TSS: Total sum scaling

V1/V4/V5: Study visits 1, 4, and 5 (pre-travel, post-travel, and one-month follow-up)

1 Introduction

The collection of microorganisms in the human gut, commonly termed the gut microbiota, plays a critical role in host physiology, immunity, metabolism, and disease susceptibility. Profiling the composition and dynamics of the gut microbiota is essential for understanding host-microbe interactions and identifying microbiota-associated mechanisms underlying disease (Lee et al., 2024; VanEvery et al., 2023). Advances in high-throughput sequencing (HTS), particularly 16S rRNA gene amplicon sequencing and shotgun metagenomics, have enabled large-scale profiling of gut microbial communities without the need for culture-based methods (Gloor et al., 2017). Such HTS-derived data are inherently compositional because each sample is constrained to a fixed total number of sequencing reads (Morton et al., 2019). As a result, the abundance of each feature is expressed relative to all others in the sample, creating dependence between features and potentially generating spurious correlations (Hickman & Korpela, 2025). Sequencing depth varies between samples and mainly reflects technical variations rather than biology, requiring normalization before analysis (Gloor et al., 2017; Jian et al., 2020).

This compositional nature complicates biological interpretation. An increase in the relative abundance of a taxon does not necessarily indicate that its absolute abundance increased; it may simply reflect a decrease in other taxa (Hickman & Korpela, 2025). Relative abundance data also ignore variation in total microbial load, which may itself be biologically important both for the ecosystem and the host because it influences the total microbial cells, metabolites, and other microbiota-derived components (Gloor et al., 2017; Jian et al., 2020; Vandeputte et al., 2017). Compositionality effects also create statistical challenges. Because all features are interdependent, many standard statistical approaches that assume independent observations can produce misleading results. In addition, relative abundances are constrained between 0 and 1 and often deviate from normal distributions, limiting the applicability of methods that rely on normality assumptions. Together, these properties can bias both ecological interpretation and statistical inference (Gloor et al., 2017; Jian et al., 2020).

Recently, experimental, computational, and statistical approaches have been developed to mitigate compositionality effects. Experimental approaches such as quantitative microbiota profiling (QMP) combine HTS data with independent measurements of total microbial load obtained by methods such as flow cytometry, real-time PCR (qPCR), digital droplet PCR

(ddPCR), or by integrating spike-in controls (Jian et al., 2020; Lloréns-Rico et al., 2021; Vandeputte et al., 2017; Wirbel et al., 2025). Studies using QMP have shown that variation in microbial load is a major source of microbiota variation and can confound disease associations (Nishijima et al., 2025; Vandeputte et al., 2017).

However, the additional cost, labor, and experimental complexity associated with QMP have motivated the development of computational approaches aimed at predicting absolute microbial abundance from HTS data. Machine learning-based methods trained on datasets with paired HTS and experimental quantification have shown that microbial load may be inferred from relative abundance profiles. For example, Nishijima et al. (2025) developed a gradient boosting model trained on large cohorts with experimentally measured microbial loads, achieving reasonable performance in internal validation but reduced accuracy in external datasets. Wirbel et al. (2025) reported a strong correlation between total DNA concentration in sequencing extracts and absolute prokaryotic abundance measured by ddPCR and demonstrated high predictive accuracy using DNA concentration as a primary predictor, including in an external Parkinson's disease cohort. An alternative strategy treats host DNA as a natural internal standard by using the ratio of bacterial to host reads in metagenomic data as a proxy for bacterial biomass, an approach particularly applicable to host-associated samples such as stool (Tang et al., 2025).

In parallel, statistical methods have been developed to analyse relative abundance data while accounting for compositionality, including centred log-ratio transformation (also known as compositional data analysis methods) and specialized differential abundance testing methods such as ALDEx2 and ANCOM-BC2 (Fernandes et al., 2013; Gloor et al., 2017; Lin & Peddada, 2024; Lloréns-Rico et al., 2021). While these approaches reduce artifacts introduced by compositionality effects to some extent, they do not recover information about absolute microbial load. A recent benchmarking study using simulated data concluded that experimental approaches such as QMP outperformed statistical approaches in detecting true taxon-load associations and disease-related microbial shifts, particularly in scenarios characterized by unequal microbial loads and dysbiosis (Lloréns-Rico et al., 2021). Compositional data analysis methods were found to be more susceptible to false positives under these conditions, limiting their interpretability in biologically complex settings (Lloréns-Rico et al., 2021). However, these findings have not been validated in real-world data.

Despite increasing recognition of the importance of microbial load and compositionality effects, important gaps remain. Most existing benchmarking studies rely on simulated data or cross-sectional datasets (Lloréns-Rico et al., 2021; Nishijima et al., 2025), and direct comparisons of computational or statistical approaches against QMP are lacking in longitudinal settings. This is a critical limitation because interventions and other longitudinal designs are especially prone to compositionality effects when a total microbial load changes over time (Jian et al., 2020). In addition, within-person temporal variation in quantitative microbiota profiles can exceed between-person variation for most genera, indicating that time-dependent biases may be missed in cross-sectional validation studies (Vandeputte et al., 2021). Longitudinal multi-omics datasets are therefore essential for understanding microbial community dynamics and evaluating method robustness under real-world ecological succession and perturbation scenarios, yet such datasets remain limited (Martínez Arbas et al., 2021).

In this thesis work, I first evaluated computational approaches for predicting total microbial load against qPCR -based bacterial enumeration. I then compared selected computational and statistical approaches proposed to mitigate compositionality effects against qPCR-based QMP for detecting differentially abundant microbial taxa in a longitudinal human cohort. The analyses used fecal microbiota profiles from a sub-cohort of Finnish travellers to Benin, West Africa, who participated in a vaccine study (Kantele et al. 2023, https://pubmed.ncbi.nlm.nih.gov/37099803/), comprising 95 Finnish adults sampled before travel, immediately after travel, and one month later, for a total of 287 fecal samples.

1.1 Aims of the thesis

This thesis assessed whether computational estimates of bacterial load and common differential abundance methods were concordant with qPCR-based QMP in a longitudinal cohort, with two specific aims:

1. To compare computationally predicted total bacterial load (machine-learning predictions and the bacterial-to-host read-count ratio) with the qPCR-based reference.
2. To compare the differentially abundant taxa recovered under different abundance representations and differential abundance testing methods, using the qPCR-based quantitative profile as the reference.

2 Methods

The overall study design and analytical workflow are summarised in Figure 1.

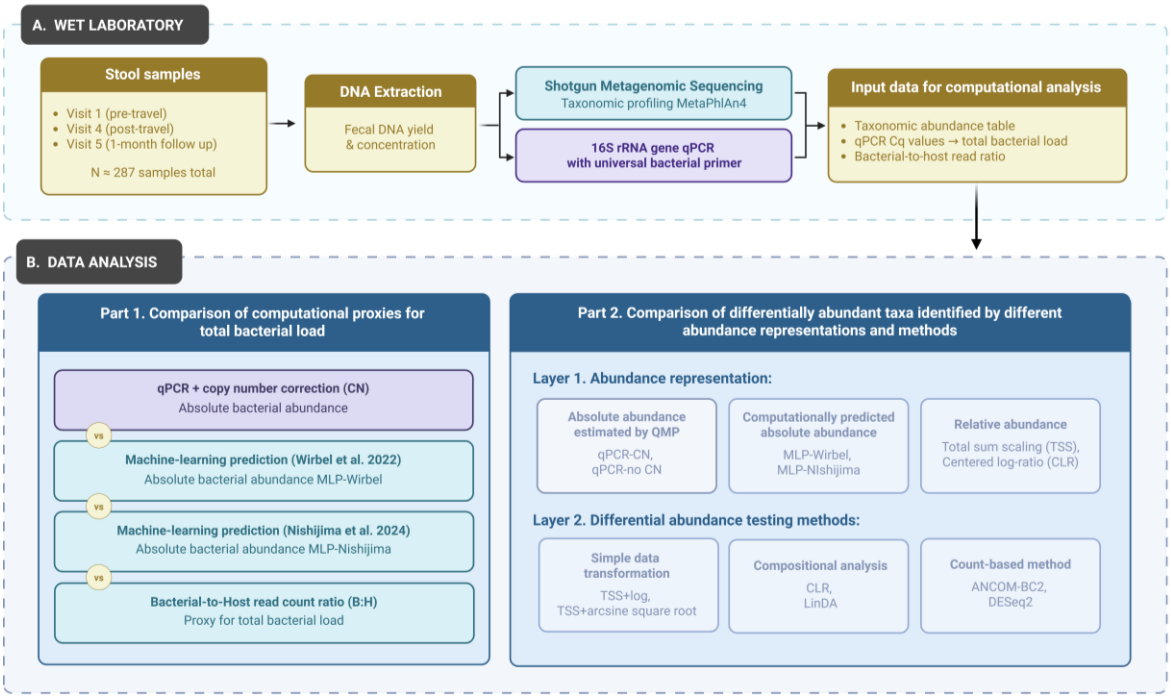


Figure 1. Overview of the wet-lab and data analysis workflow used in this thesis. (A) Wet-lab and data processing workflow. Fecal samples collected at three time points were analyzed by shotgun metagenomic sequencing and universal 16S rRNA gene qPCR. Shotgun sequencing generated raw metagenomic data, while qPCR provided estimates of total bacterial load. Host-derived reads were identified and removed using Hostile, and the remaining microbial reads were taxonomically profiled with MetaPhlAn4. This workflow generated taxonomic abundance tables, qPCR-based total bacterial load estimates, and bacterial-to-host read count ratios as input for data analysis. **(B)** Data analysis. Part 1 compared three computational proxies for total bacterial load against the qPCR reference; Part 2 assessed the concordance of different abundance representations and differential abundance testing workflows in identifying differentially abundant taxa across three paired time-point comparisons. References to all used tools are in the body text.

2.1 Study cohort, sample processing, and metagenomic sequencing

Fecal samples were obtained from a sub-cohort of a randomized controlled trial (Kantele et al., 2023), which evaluated the safety and immunogenicity of ETVAX®, a multivalent oral whole-cell inactivated enterotoxigenic *Escherichia coli* (ETEC) vaccine in Finnish adult travelers to Benin, West Africa. Samples were collected at three time points: before departure (V1, pre-travel), immediately upon return (V4, post-travel), and a one-month follow-up visit (V5). Host-related and clinical outcomes were outside the scope of this thesis, but the included participants were balanced across the randomized group, age, sex, and other potential confounding factors. Fecal DNA extraction, shotgun sequencing, and taxonomic profiling were performed before the start of this thesis work, but for the context are described below. Fecal DNA was extracted using the KingFisher Flex automated

purification system with MagMAX reagents (Thermo Fisher Scientific), as described previously (Jokela et al., 2023). Shotgun metagenomic sequencing was performed as described previously (Dubois et al., 2024). Briefly, sequencing libraries were prepared using the NEBNext Ultra II DNA library Prep Kit for Illumina and sequenced on the NovaSeq 6000 S4 PE150 platform, generating 2 × 150 bp paired-end reads. Sequencing yielded an average of 34 ± 8.2 million reads per sample.

Raw reads were assessed using FastQC, and adapters were removed with cutadapt v3.5 (Martin, 2011). Human-derived reads were identified, counted, and removed using Hostile v0.2.0 (Constantinides et al., 2023). Microbial taxonomic profiles were generated with MetaPhlAn 4.0 using default parameters and the mpa_vJun23_CHOCOPhlAnSGB_202403 database (Blanco-Míguez et al., 2023). After quality assessment, 287 high-quality metagenomic fecal microbiota profiles from 95 participants were retained for further analysis.

2.2 qPCR-based total bacterial enumeration and QMP

2.2.1 qPCR protocol

Total bacterial load in each sample was quantified by qPCR using the universal eubacterial primer pair 331F (5'-TCCTACGGGAGGCAGCAGT-3') and 797R (5'-GGACTACCAGGGTATCTAATCCTGTT-3'), as described by Jian et al. (2020). Reactions were prepared in a total volume of 25 µL containing 5 µL HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne0, 0.25 µM of each primer, 0.02 ng of normalized template DNA, and nuclease-free water. Each sample was run in three technical replicates on a Bio-Rad CFX96 Touch Real-Time PCR Detection System. Thermocycling conditions were as follows: initial denaturation at 95°C for 15 min; 39 cycles of 95°C for 15 s, 50°C for 20 s, and 72°C for 30 s, with fluorescence acquisition at the end of the extension step (Jian et al., 2020).

A standard curve was generated using a purified 16S rRNA gene PCR product amplified from *Bifidobacterium longum* genomic DNA with the same primer pair. The copy number of the purified amplicon was calculated from its measured DNA mass and length, using 660 g/mol per base pair as the average molecular weight of double-stranded DNA and Avogadro's number. Standard dilutions ranging from the 10^2 to 10^7 copies/µL were used to generate the standard curve by plotting threshold cycle (Ct) value against $\log_{10}$-transformed copy numbers. Ct values from fecal DNA samples were then converted to 16S rRNA gene copy

numbers using this curve. After correcting for template input volume, DNA dilution, extraction volume, and fecal sample mass, the results were expressed as 16S rRNA gene copies per gram of wet feces, as described by Jian et al. (2020).

All qPCR assays were performed in triplicate. Precautions were taken to ensure that the data from each triplicate fell within 0.5 Ct, and clear outliers (>2 standard deviations) were removed before calculating the average Ct of each sample. Melting curves and non-template controls were used to assess run reliability. There was no detectable amplification from no-template controls in any of the assays. The amplification efficiencies of all qPCR assays ranged from 92% to 103%, indicating acceptable qPCR performance (Jian et al., 2020).

2.2.2 Calculation of absolute abundance and copy-number correction

Sequencing reads assigned to each taxon were divided by the total reads per sample to obtain relative abundances. These relative abundances were then converted into absolute abundances by multiplying the relative abundance of each taxon by the total bacterial abundance in the sample, as determined by qPCR (Jian et al., 2020). To account for variation in 16S rRNA gene copy numbers between taxa, optional copy-number correction was performed by dividing the absolute abundance of each taxon by the number of 16S rRNA gene copies in its genome, using the 16S copy number database rrnDB (Stoddard et al., 2015).

2.3 Computational proxies for total bacterial load

As indicated in Figure 1 (Part 1 in “B. Data analysis”), two machine learning approaches were used to predict total microbial load from shotgun metagenomic data. In addition, the log-transformed bacterial-to-host read count ratio was used as an independent metagenome-derived proxy for total bacterial load.

2.3.1 Machine-learning-based load prediction of total microbial load and absolute abundance

The Wirbel model, labelled W3 in Table 1 and based on Wirbel et al. (2025) was implemented with the mlr3 framework (v1.5.0) and retrained on the thesis cohort. The model used seven sample-level features derived from shotgun metagenomic sequencing: extracted DNA concentration, bacterial fraction of total sequencing reads, alpha diversity (Shannon index), and the proportions of reads assigned to Bacterial, Archaea, Eukaryota, or

unclassified taxa. The model predicted total 16S rRNA gene copies per sample on a log10 scale. Predictions were back-transformed and divided by stool input weight to obtain 16S rRNA gene copies per gram of stool. W3-based absolute taxon abundances were then calculated by multiplying each MetaPhlAn4 relative abundance proportion by the predicted sample-level total load, following the same scaling approach used for the qPCR-based QMP.

The Nishijima model, labelled W4 in Table 1 and based on Nishijima et al. (2025) was applied using the MLP R package (v.0.1.0) with xgboost v1.7.5.1 as the backend. The model used MetaPhlAn4 species-level relative abundance profiles, expressed as proportions and generated with the mpa_vJun23_CHOCOPhlAnSGB_202403 database. Predictions were generated using the MetaCardis-trained shotgun metagenome model configuration available in the package. This model was originally trained against flow cytometry-based cell count and predicts total microbial cell counts per gram of stool. W4-based absolute taxon abundances were obtained from the package's quantitative microbiome profiling output mode.

2.3.2 Bacterial-to-host read ratio

The bacterial-to-host (B:H) read count ratio was evaluated as a metagenome-derived proxy for total bacterial load, as proposed by Tang et al. (2025). The ratio was calculated by dividing the number of bacterial reads by the number of host-derived reads in each stool metagenome. This approach treats host DNA as an internal reference signal and assumes that host DNA shedding into stool is relatively stable, or changes in proportion to microbial biomass, within the target study population.

2.4 Comparison of differentially abundant taxa identified by different abundance representations and methods

As indicated in Figure 1 (Part 2 in "B. Data analysis") and detailed in Table 1, the analysis had two layers. The first layer compared seven abundance representations (W1–W7) for identifying differentially abundant taxa across three paired time-point contrasts: V1–V4, V4–V5, and V1–V5. Each representation was tested using either a random-intercept linear mixed model (LMM) or a paired Wilcoxon test, resulting in 14 workflows. For these two-time-point contrasts, the LMM tests the same within-participant mean difference as a paired t-test under standard Gaussian assumptions. The second layer fixed the abundance testing methods (M1 – M10) for detecting differentially abundant taxa across three paired time-

point contrasts. These methods were selected for their widespread use in microbiome differential abundance testing and their applicability to longitudinal data (Nearing et al., 2022; Pelto et al., 2025; Yang & Chen, 2023). The qPCR-based QMP workflow (applying the copy-number correction with an LMM (W1/M1) was used as the reference in both layers.

Table 1. Abundance representation workflows and differential abundance testing methods evaluated in this study.

Layer 1. Abundance representation

ID	Workflow	Type	Input	Norm.	Trans.
W1	**qPCR (CN-corrected) ★**	Measured absolute	16S qPCR + relative profiles	–	log
W2	qPCR (no CN-correction)	Measured absolute	16S qPCR + relative profiles	–	log
W3	MLP Wirbel	Predicted absolute	HTS + DNA concentration †	–	log
W4	MLP Nishijima	Predicted absolute	HTS (taxonomic profiles) †	–	log
W5	TSS + AST	Relative	HTS (taxonomic profiles)	TSS	AST
W6	TSS + log	Relative	HTS (taxonomic profiles)	TSS	log
W7	CLR	Relative (compositionality-aware)	HTS (taxonomic profiles)	CLR	–

Layer 2. Differential abundance testing method

ID	Method	Type	Norm.	Trans.	Covar.	Compositionality-aware
M1	**qPCR · LMM ★**	Absolute	–	log	Yes	N/A
M2	TSS + AST · LMM	Relative + simple transformation	TSS	AST	Yes	No
M3	TSS + log · LMM	Relative + simple transformation	TSS	log	Yes	No
M4	TSS + AST · Wilcoxon	Relative + simple transformation	TSS	AST	No	No
M5	TSS + log · Wilcoxon	Relative + simple transformation	TSS	log	No	No
M6	CLR · LMM	Compositional data analysis	CLR	–	Yes	Yes
M7	CLR · Wilcoxon	Compositional data analysis	CLR	–	No	Yes
M8	LinDA	Compositional data analysis	CLR‡	–	Yes	Yes
M9	ANCOM-BC2	Count-based	–	–	Yes	Yes
M10	DESeq2	Count-based	–	–	Yes	No

All differential abundance methods except qPCR · LMM (M1) are applied to the same taxonomic count table derived in Layer 1. Norm., normalization; Trans., transformation; Covar., covariate adjustment; qPCR, quantitative PCR; CN, copy number; HTS, high-throughput sequencing; MLP, machine-learning prediction; TSS, total sum scaling (each taxon count divided by its sample total to give proportions, i.e., relative abundance); AST, arcsine square root transformation; log, log transformation; CLR, centered log-ratio transformation; LMM, linear mixed model. † MLP workflows predict total microbial load from sequencing-derived inputs using a model trained on paired absolute-load measurements (see Section 2.3.1). ‡ LinDA applies an internal CLR-like bias correction. ★ Reference workflow/method. Compositionality-aware indicates whether a method is designed to account for the constant-sum (compositional) constraint of the data.

2.4.1 Preparation of abundance tables, data normalization, and/or transformation

Before differential abundance testing, species-level tables were aggregated to the family level by summing relative abundances or read counts across species assigned to the same family. This aggregation was used to maintain the analysis at a taxonomic resolution appropriate to the scope of this Master's thesis. For the abundances-representation comparison, taxa were retained only if they were present in more than 10% of the sample across all model-specific input tables, ensuring a common feature set. A minimum relative abundance threshold of 0.2% was applied. Absolute abundance tables (W1–W4) were log-transformed after adding a pseudo-count of 1. Total-sum-scaled (TSS), in which each taxon's read count is divided by

the total read count of its sample so that the taxa within a sample sum to 1 (i.e., proportions). These TSS relative abundances were transformed using either arcsine square root transformation (AST, W5) or log transformation (W6). Centred log-ratio transformation (CLR, W7) was applied to relative abundances after adding a pseudo-count of 1×10^{-6}, using the per-sample geometric mean of all taxa as the reference.

2.4.2 Differential abundance testing methods

LMMs were used for M1–M3 and M6. Models were fitted with time point as a fixed effect and participant identity as a random intercept. Each model was applied to its corresponding transformed abundance table, including qPCR-based QMP absolute abundances, metagenome-derived absolute abundance estimates, and transformed relative abundance data. Paired Wilcoxon signed-rank tests were used for M4, M5, and M8. These tests were applied within each paired time-point contrast using the corresponding transformed abundance tables described in Section 2.4.1. *LinDA* (M7; Zhou et al., 2022) was applied to TSS-normalized relative abundances using MicrobiomeStat v1.4.0, with *feature.dat.type* = "*proportion*". Participant identity was included as a random effect. Internal prevalence and abundance filters were disabled because taxon filtering had already been applied before analysis. *ANCOM-BC2* (M9; Lin & Peddada, 2024) was applied to read count estimates from MetaPhlAn4 using *ANCOMBC* v2.13.1. Time point was included as the fixed effect and participant identity as a random effect. Internal prevalence and library-size filters were disabled because filtering had already been performed. *DESeq2* (M10; Love et al., 2014) was applied to read count estimates from MetaPhlAn4 using a paired design with participant identity and time point as model terms. Size factors were estimated using the poscounts methods.

2.4.3 Statistical analysis

Statistical analysis and visualization were performed in R version 4.5.3 using stats v4.5.3, lme4 v2.0.1, lmerTest v3.2.1, ANCOMBC v2.13.1, DESeq2 v1.46.0, MicrobiomeStat v.1.4.0, dplyr, tidyr, ggplot2, and ComplexHeatmap.

Agreement between each computational microbial load proxy and the qPCR-based bacterial enumeration reference was assessed using Spearman's rank correlation coefficient (ρ), calculated from pairwise complete observations. All load estimates were log-transformed before correlation analysis and visualization. Changes in total bacterial load across V1, V4,

and V5 were tested using the Friedman test, with Kendall's W reported as the effect size. Pairwise differences between V1–V4, V4–V5, and V1–V5 were tested using paired Wilcoxon signed-rank tests.

For differential abundance analyses, overlap in significant taxa (bacterial families) was compared among workflows or methods using the Jaccard index. The Jaccard index was calculated as the number of shared significant taxa divided by the number of significant taxa in either workflow/method. A value of 1 indicates identical sets of significant taxa, whereas a value of 0 indicates no overlap. Pairwise Jaccard indices were visualized as heatmaps.

Because the Jaccard index does not show direction or strength of change, each candidate workflow or method was also compared with qPCR-based QMP reference using standardized effect sizes. These were calculated as the sign of the model coefficient multiplied by $-\log_{10}$ of the FDR-adjusted *P* value, and were capped at ± 4 for visualization. Scatterplots were used to show agreement with the reference on a common scale. Taxa were colored by significance status and direction of change to distinguish concordant changes, opposite-direction changes, candidate-only findings, reference-only findings, and non-significant taxa. This allowed the main source of disagreement with the qPCR-based QMP reference to be identified visually.

Where applicable, *P* values were adjusted for multiple testing across taxa within each time-point contrast using the Benjamin-Hochberg false discovery rate (FDR). Taxa were considered differentially abundant at FDR < 0.05.

2.5 Use of artificial intelligence tools

Claude (Sonnet 4.6, Anthropic) was used for R code writing, particularly for improving figure aesthetics. Grammarly (v1.2.268.1900, Grammarly Inc.) was used to check grammar and improve readability throughout the manuscript. All AI-assisted outputs were reviewed, edited, and verified by the author to ensure accuracy and consistency with the original analyses and interpretation.

2.6 Code availability

The code to replicate the analyses in this thesis can be found at: Nugrah-Fadillah/quantitative-microbiota-profiling: Master's thesis analysis: Quantitative microbiota profiling, University of Helsinki 2026.

3 Results

3.1 Comparison of computational proxies for total bacterial load

Across all samples (N = 287), qPCR-based bacterial enumeration showed weak but significant correlations with the two machine-learning proxies for total bacterial load: Wirbel: ρ = 0.286, p < 0.001; Nishijima: ρ = 0.319, p < 0.001 (Figure 2A). The bacterial-to-host read (B:H) ratio was not significantly correlated with qPCR-based total bacterial load (ρ = 0.061, p = 0.307; Figure 2A).

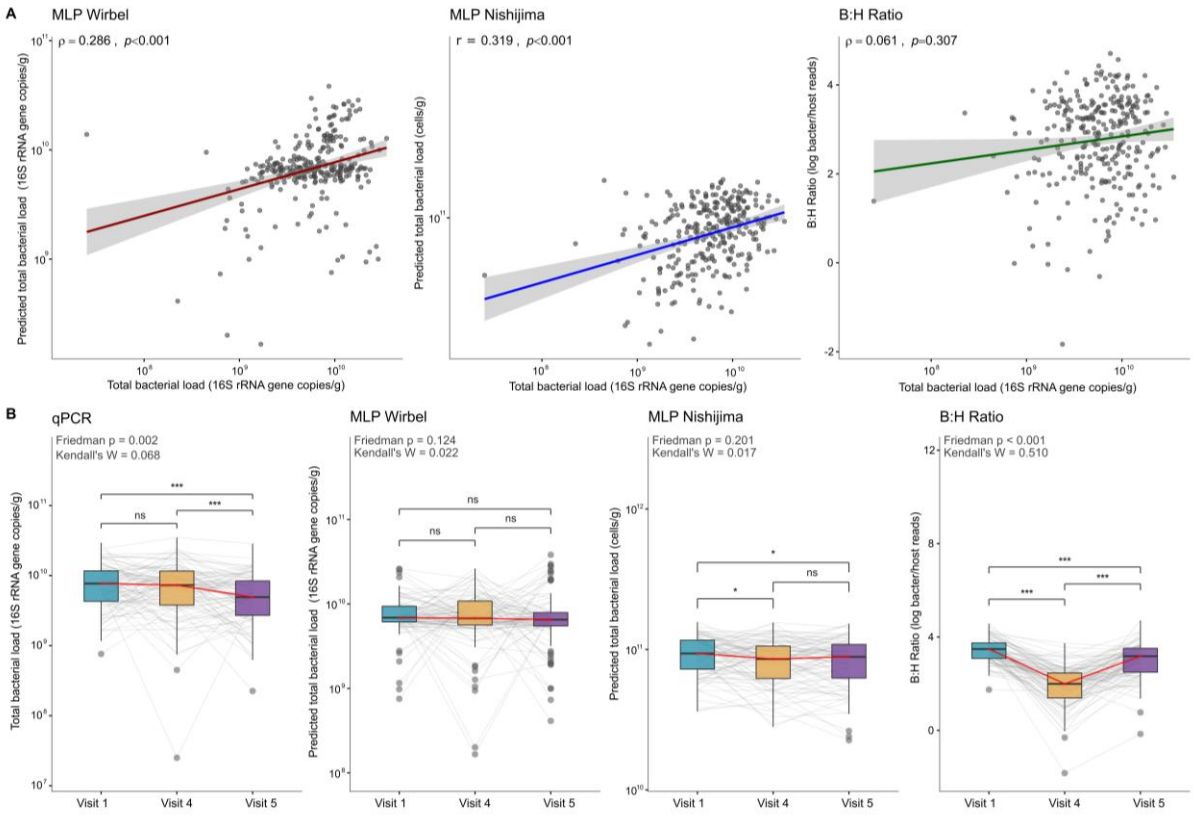


Figure 2. Comparison of computational proxies for total bacterial load against qPCR-based bacterial enumeration. (A) Spearman correlations between qPCR-based total bacterial load and each computational proxy across all 287 samples. Load values were log-transformed before analysis and visualization. **(B)** Longitudinal changes in qPCR-based bacterial load, machine-learning-predicted load, and the bacterial-to-host read count ratio across V1, V4, and V5 in participants with complete data (N = 95). Grey lines link individual participants, and red lines show the median trend. Significance labels show paired Wilcoxon signed-rank test results: ns, not significant; * p < 0.05; ** p < 0.01; *** p < 0.001.

Longitudinal patterns differed between the proxies and the qPCR reference (Figure 2B). qPCR-based total bacterial load changed significantly across V1, V4, and V5 (Friedman test: p = 0.002, Kendall's W = 0.068). Pairwise tests showed no significant change from V1 to V4, but bacterial load was lower at V5 than at both V1 and V4. Neither machine-learning proxy detected a significant longitudinal change (Wirbel: p = 0.124, W = 0.022; Nishijima:

p = 0.201, W = 0.017). In contrast, B:H ratio changed strongly across visits (Friedman p < 0.001, W = 0.510) and differed in all pairwise comparisons, but this pattern did not match the qPCR reference.

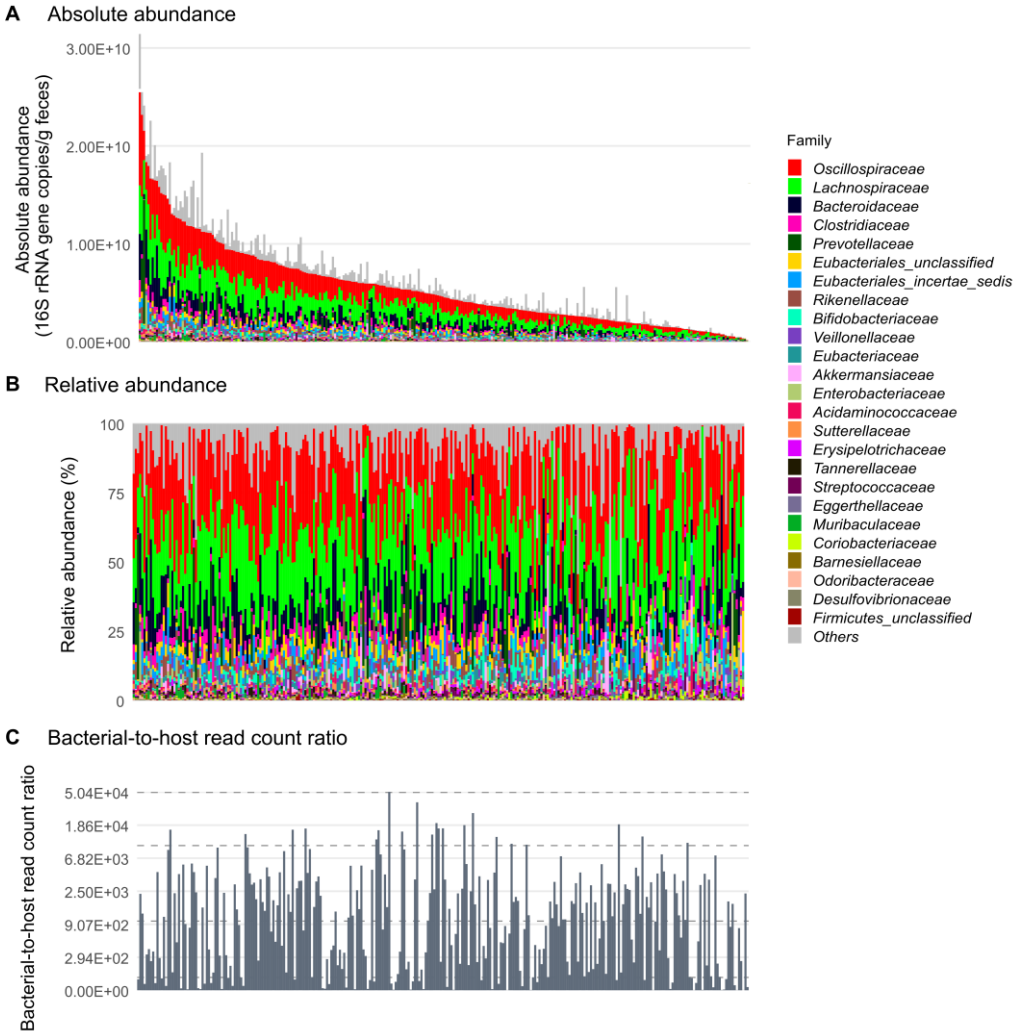


Figure 3. Family-level microbiota profiles and bacterial-to-host read count ratios aligned by sample and ranked by qPCR-based total bacterial load. The same sample order is used across all three panels to compare absolute abundance, relative abundance, and the B:H ratio within each sample. **(A)** qPCR-based QMP profiles showing absolute abundances, expressed as 16S rRNA gene copies per g feces. **(B)** Relative abundance profiles from shotgun metagenomic sequencing, expressed as percentage of total microbial reads. **(C)** Bacterial-to-host read count ratio for the same samples. The 25 most abundant families based on mean relative abundance are shown; remaining families are grouped as “Others”.

The same disagreement was visible when comparing family-level abundance profiles (Figure 3). qPCR-based QMP profiles and relative abundance profiles gave different sample rankings and community patterns. For example, *Oscillospiraceae* and *Lachnospiraceae* had higher absolute abundances in high-load samples, but their relative abundances did not follow the same pattern (Figure 3). The B:H ratio also did not follow the rank order of either relative or absolute abundance profiles and varied widely across samples. These differences

show that different representations of the same metagenomic data can lead to different views of community structure, supporting further comparison of how abundance representations and differential abundance testing methods affect the detection of differentially abundant families over time.

3.2 Concordance of abundance representations and testing

Differential abundance was assessed at the family level across three paired longitudinal contrasts: V1–V4, V4–V5, and V1–V5. Candidate workflows and differential abundance (DA) testing methods were compared with qPCR-CN·LMM reference (or each other) using (i) Jaccard overlap of statistically significant families, (ii) standardized effect-size plots, and (iii) summaries for selected high- and low-abundance families. Concordance was first examined across abundance representations, then across DA testing methods, following the structure outlined in Section 2.4.

Figure 4A shows Jaccard overlap between each abundance representation and the qPCR-CN·LMM reference. Concordance was highest for the V1–V4, where qPCR-based total bacterial load did not change significantly (Figure 2B). In contrast, the MLP-based absolute abundance workflows had the highest overlap with the reference, with a Jaccard value of 0.91 for both MLP-Wirbel·LMM and MLP-Nishijima·LMM. TSS-based relative abundance workflows also showed high overlap, with Jaccard values of 0.76 for TSS+AST·LMM and 0.78 for TSS+log·LMM. CLR-based workflows showed the lowest overlap with the reference and with the other workflows.

The largest disagreement occurred for the V1–V5, where Jaccard overlap with the reference fell to 0.07 for TSS+AST·LMM and 0.51 for TSS+log·LMM. The MLP-based workflows remained slightly more concordant, with Jaccard values of 0.63 for MLP-Wirbel·LMM and 0.61 for MLP-Nishijima·LMM. Thus, as expected, disagreement between absolute and relative abundance representations generally increased when total bacterial load changed more strongly. It is worth noting, however, that the simple TSS+log·LMM workflow consistently retained moderate-to-high overlap with the qPCR-based absolute abundance reference across different contrasts.

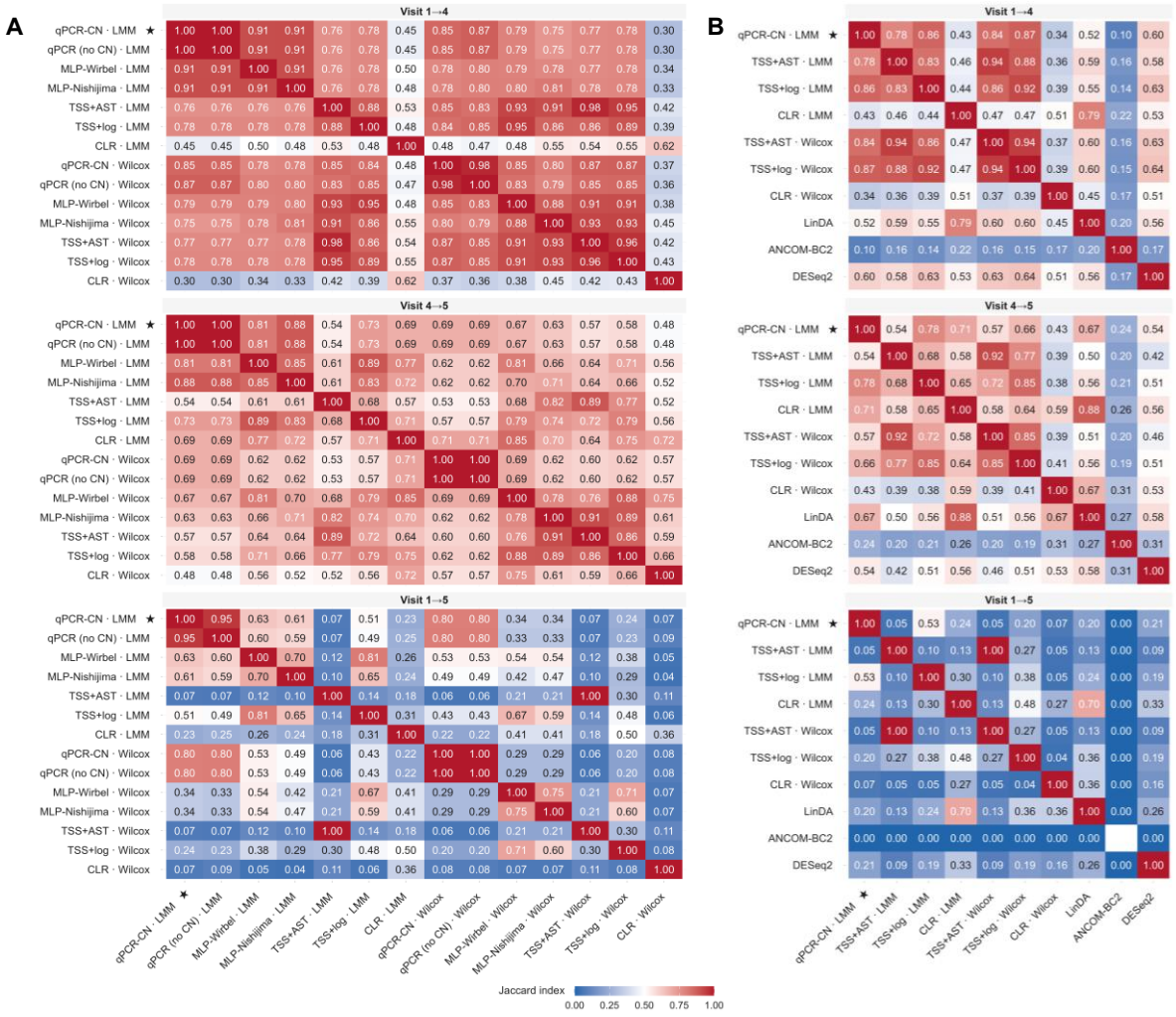
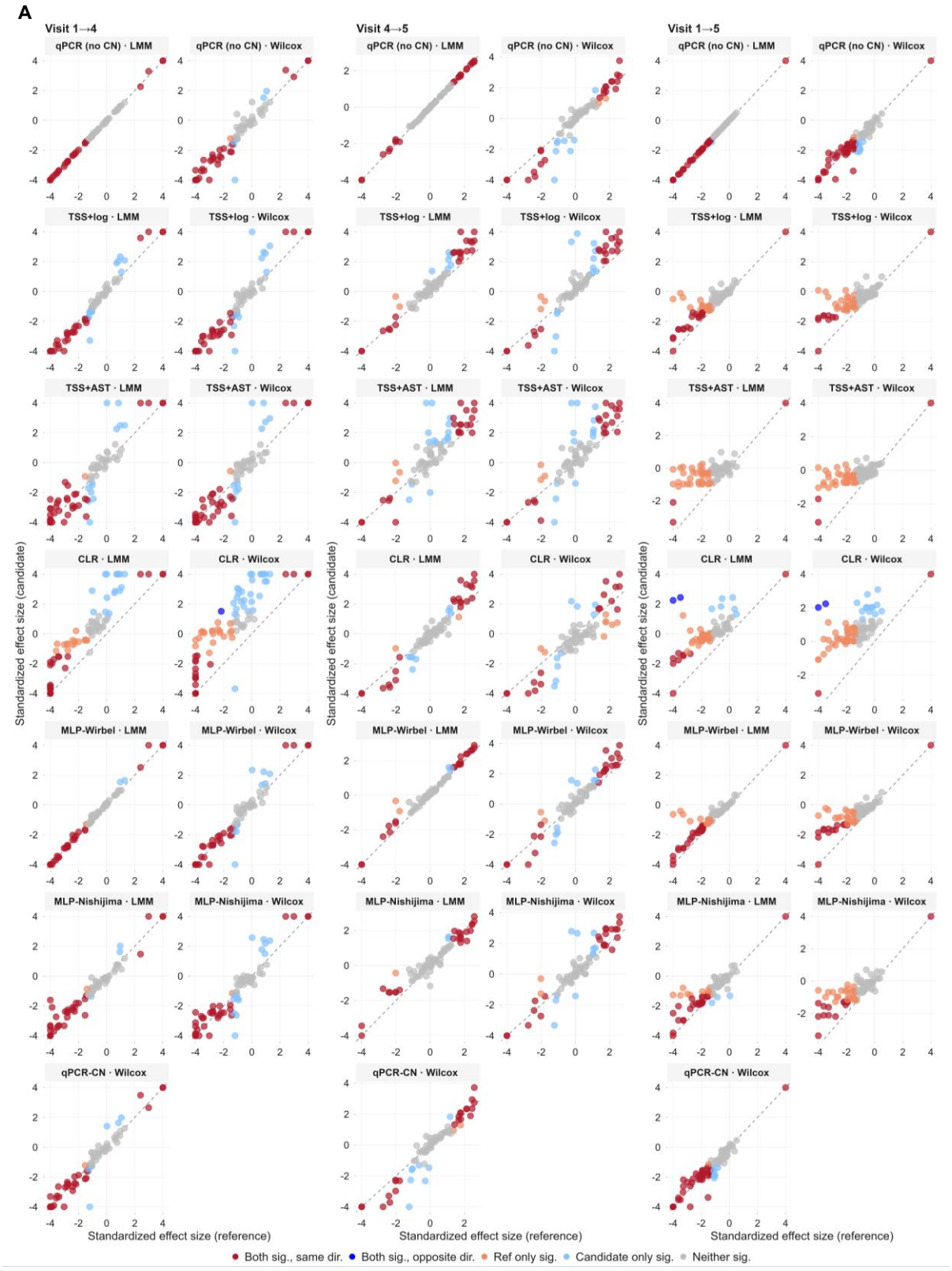


Figure 4. Jaccard overlap of differentially abundant families across abundance representations and DA testing methods. Jaccard overlap was calculated for each pair of workflows or methods within each longitudinal contrast as the number of shared statistically significant families divided by the number of families significant in either workflow or method (FDR-p < 0.05). (A) Pairwise overlap across abundance representations, including qPCR-based QMP, machine-learning-predicted absolute abundance, TSS-normalized relative abundance, and CLR-transformed relative abundance. (B) Pairwise overlap across DA testing methods. ★ Reference workflow/method.

Figure 4b compares DA testing methods against the same qPCR-CN·LMM reference. Among these non-absolute abundance methods, TSS+log·LMM was the most concordant, including at V1–V5, where its Jaccard value was 0.53. Other DA methods showed lower overlap at this contrast. DESeq2 decreased from 0.60 at V1-V4 to 0.21 at V1–V5. LinDA showed moderate overlap at V1–V4, with a Jaccard value of 0.52, but lower concordance at V1–V5, with a value of 0.20. ANCOM-BC2 showed consistently lower overlap with the reference, with Jaccard values of 0.10 at V1–V4, 0.24 at V4–V5, and 0.00 at V1–V5, where it did not identify any significant families.

Because Jaccard overlap only compares which families were significant, standardized effect sizes were calculated and used to compare the direction and relative strength of change against the qPCR-CN·LMM reference.



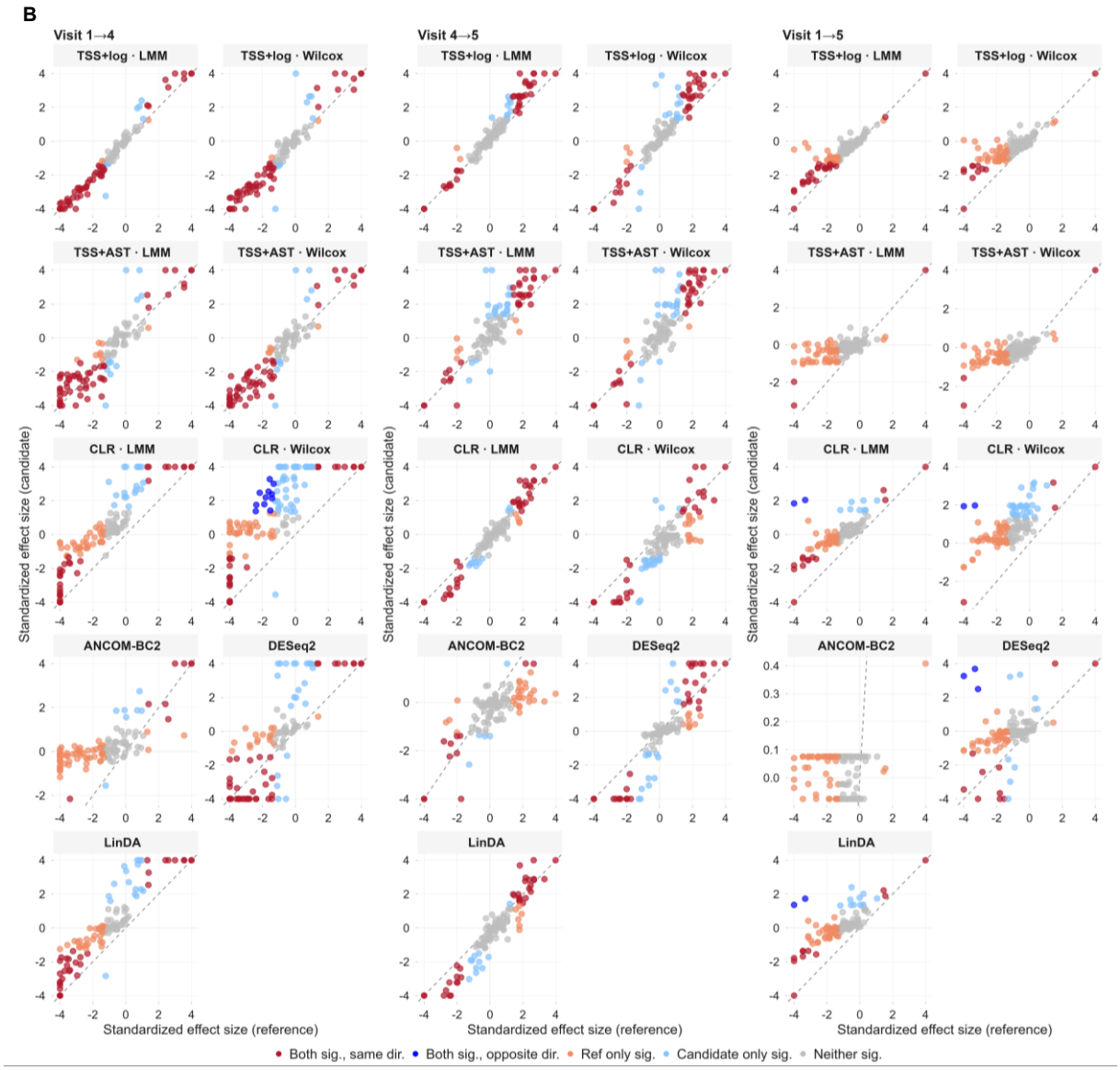


Figure 5. Effect size agreement between candidate differential abundance methods and the qPCR-based absolute abundance reference. Standardized effect sizes for all tested bacterial families comparing each candidate method (y-axis) against qPCR-CN·LMM (x-axis) at three longitudinal contrasts. Color indicates direction (dir.) and concordance status. (A) Representation effect agreement across candidate methods. (B) Method effect agreement across candidate methods.

In the abundance-representation layer (Figure 5A), qPCR(no CN)·LMM closely matched qPCR-CN·LMM across contrasts, indicating that 16S rRNA gene copy-number correction had little effect on the DA results. MLP-based workflows and TSS+log·LMM followed the reference more closely than TSS+AST·LMM and CLR-based workflows. At V1–V5, where qPCR-based total bacterial load changed most, disagreement was mainly driven by families detected only by the qPCR-CN·LMM reference and by larger deviations from the 1:1 concordance line.

In the DA-method layer (Figure 5B), TSS+log·LMM showed the closest agreement with the qPCR-CN·LMM reference among the selected DA methods. DESeq2 produced more candidate-only significant families, whereas LinDA and ANCOM-BC2 showed weaker agreement with the reference. Across both layers, most disagreements were reference-only or candidate-only findings rather than opposite-direction changes. The main exception was CLR·Wilcox, which identified a larger proportion of significant families in the opposite direction from the qPCR-CN·LMM reference.

To test whether disagreement with the qPCR-CN·LMM reference varied with abundance, directional agreement was examined separately for selected high- and low-abundance families, defined by the mean relative abundance across all samples within the abundance-representation (or DA-method) layer (Figure 6).

Among high-abundance families, agreement varied by longitudinal contrast and workflow (Figure 6A). At V1–V4, *Bacteroidaceae* showed the clearest agreement, with most workflows and methods detecting the same significant increase as the qPCR-CN·LMM reference. Other high-abundance families showed weaker agreement: several workflows and methods called *Oscillospiraceae* or *Clostridiaceae* significant, although they were not significant in the reference. At V4–V5, *Bacteroidaceae* was consistently detected as significantly decreased, whereas the significant decrease in *Lachnospiraceae* detected by the reference was missed by several candidate workflows/methods. The strongest disagreement among high-abundance families occurred at V1–V5, the contrast with the largest decrease in qPCR-based total bacterial load. At this contrast, several candidate workflows and methods failed to detect high-abundance families that decreased significantly in the reference. CLR-based workflows, LinDA, and DESeq2 also showed directional disagreement for some high-abundance families, calling them significantly increased when the reference detected decreases. ANCOM-BC2 showed limited agreement overall, with few significant calls among high-abundance families. Among low-abundance families, agreement was higher at V1–V4, where several workflows and methods detected significant changes in the same direction as the qPCR-CN·LMM reference (Figure 6B). However, the overall pattern did not show a simple difference between high- and low-abundance families. Workflows/method-dependent disagreement appeared to depend more on the longitudinal contrast and the workflow/method than on family abundance alone.

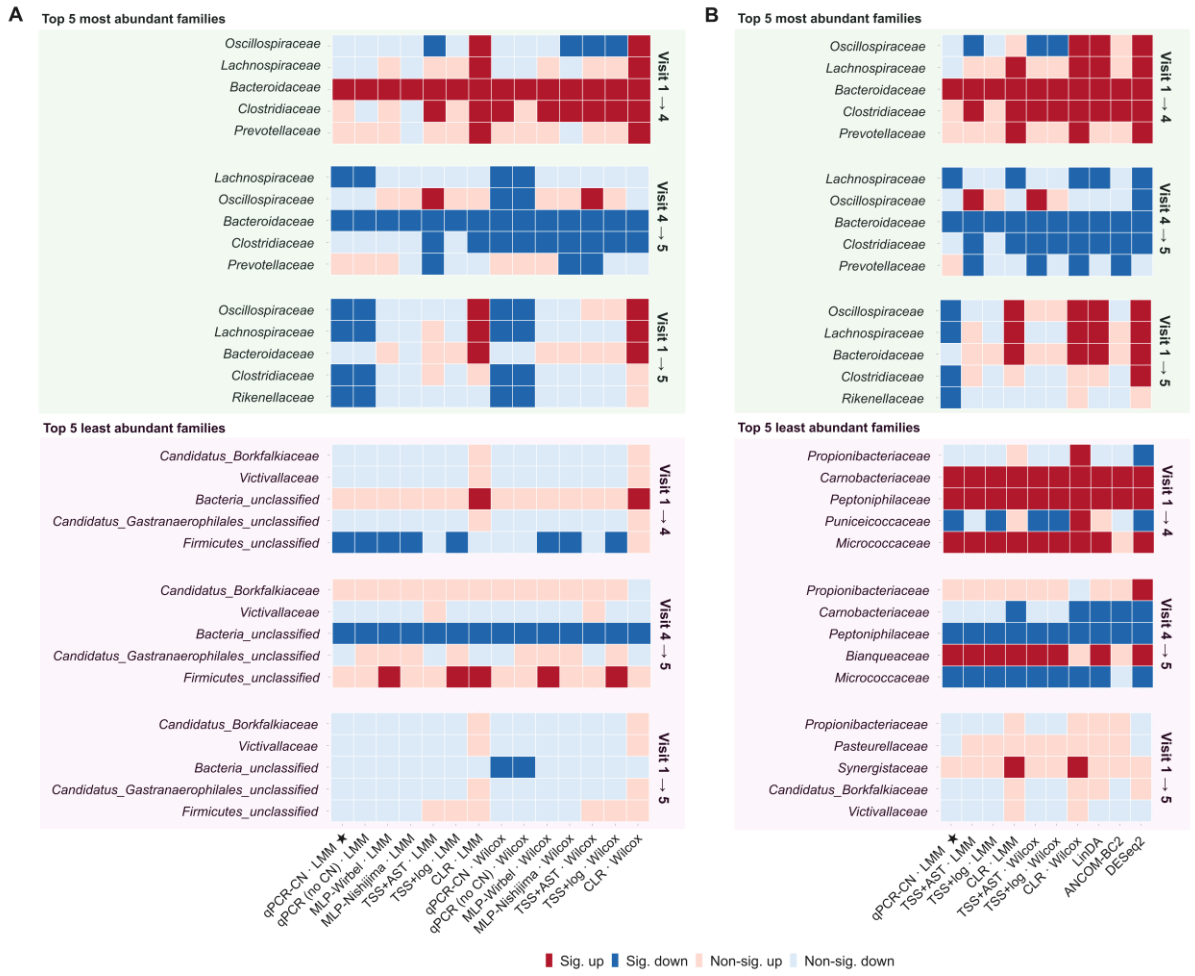


Figure 6. Direction of differential abundance results for selected high- and low-abundance families. (A) Abundance-representation layer. **(B)** DA-method layer. Significance and direction of change are shown for the five highest-abundance families and five lowest-abundance families, ranked by mean relative abundance across all samples within each analysis layer. Results are shown for three paired contrasts: V1–V4, V4–V5, and V1–V5.

4 Discussion

This study demonstrates that differences in total bacterial load can significantly affect which bacterial families are identified as differentially abundant over time when using relative or computationally predicted absolute abundances from sequencing data, which lack experimentally measured bacterial load. Agreement with the qPCR-based QMP reference was highest when bacterial load was stable and lowest when bacterial load decreased most. Machine-learning-predicted absolute abundance yielded slightly closer results than most relative-abundance-based workflows but still failed to match qPCR-based QMP. Among approaches without directly measured bacterial load, total-sum-scaled, log-transformed relative abundance with a linear mixed model showed the closest agreement.

4.1 Limited transferability of machine-learning-predicted bacterial load

The Wirbel and Nishijima models showed weak but significant correlations with qPCR-based total bacterial load across all samples, but neither detected a significant overall change across visits. The two models were trained under different assumptions, and the sets of input features may be strongly influenced by cohort-specific characteristics as well as technical variation.

The Wirbel model uses sample-level sequencing and DNA features, including bacterial and host read fractions, kingdom-level read proportions, Shannon diversity, and extracted DNA concentration (Wirbel et al., 2025). These features can be affected by DNA extraction, host DNA shedding, and changes in gut physiology during travel, such as stress and diarrhea (Boolchandani et al., 2022; McArdle & Kaforou, 2020; Sui et al., 2020). The Nishijima model uses species-level relative abundance to predict microbial cell counts per gram of stool (Nishijima et al., 2025). This assumes that the relationship between community composition and microbial load in the training cohorts also applies to the present cohort. In this cohort, drastic travel-related microbiota shifts may have weakened that relationship (Mlangeni et al., 2025).

Despite these limits, machine-learning-predicted absolute abundance showed high agreement with the qPCR-based reference for differentially abundant families at V1–V4, where qPCR-based total bacterial load was stable. Agreement decreased at V1–V5, where bacterial load decreased most, and missed or discordant family-level calls became more apparent. These results suggest that machine-learning microbial load prediction may be

useful when measured bacterial load is unavailable, but it should not be treated as equivalent to QMP, especially when total bacterial load is expected to drastically change over time.

4.2 Decoupling of bacterial-to-host (B:H) read count ratios from total bacterial loads

The B:H ratio showed no significant correlation with qPCR-based total bacterial load but showed the greatest longitudinal change among the load proxies. As mentioned previously, the B:H ratio assumes that host DNA input is relatively stable, so that changes in the ratio reflect changes in bacterial load (Tang et al., 2025). This assumption is unlikely to hold in our cohort. Travel-related stress, diarrhea, inflammation, and altered gut transit may change epithelial shedding and host DNA release into stool (La Torre et al., 2023; Leigh et al., 2023; Procházková et al., 2023). The ratio can also be affected by DNA extraction efficiency, sample collection, and storage, because host and bacterial cells may not be recovered with equal efficiency (Doyle et al., 2025; Maghini et al., 2024). Therefore, the B:H ratio is not a reliable predictor of changes in bacterial biomass.

4.3 Total-sum-scaled log-transformed relative abundance best approximated qPCR-based absolute-abundance results

Among methods without measured bacterial load, total-sum-scaled log-transformed relative abundance generally gave good agreement with qPCR-based absolute-abundance reference. With a linear mixed model, it agreed better with the reference than centered log-ratio methods, LinDA, ANCOM-BC2, and DESeq2.

Microbiota differential abundance analysis has long been difficult because sequencing data are sparse, high-dimensional, and compositional (Armstrong et al., 2022; Kumar et al., 2024). These data properties are thought to make elementary DA methods (e.g., t-test) inappropriate (Kleine Bardenhorst et al., 2021), but more complex methods are not necessarily advantageous. RNA-seq-derived methods such as DESeq2 are often used for microbiota data, but recent comparisons have argued that DA methods can give very different results on the same datasets, and that count-based models can produce unstable or inflated findings when their assumptions do not fit microbiota data (Hawinkel et al., 2020).

The same problem applies to compositional data analysis methods, such as centered log-ratio methods (see also Section 4.4). In this study, compositional data analysis methods did not improve agreement with the qPCR-based QMP reference. Instead, they showed lower

concordance and, in some contrast, opposite directions for several high-abundance families. This agrees with recent work using biological ground-truth datasets, where commonly used compositional data analysis methods did not consistently outperform simpler approaches and sometimes produced directionally wrong results (Gamboa-Tuz et al., 2025).

Pelto et al. (2025) reached a similar conclusion using 69 datasets from 53 human gut microbiome studies. Their study was mainly based on cross-sectional case-control comparisons rather than longitudinal load changes, but it still showed that simple analyses outperformed compositional methods (Pelto et al., 2025). The present longitudinal results thus extend these findings: total-sum-scaled log-transformed relative abundance gave reasonable agreement when bacterial load was stable, but lost sensitivity when total bacterial load changed over time. At the strongest load-change contrast, it mainly missed families detected by the qPCR-based QMP reference rather than calling them in the opposite direction. This is less misleading than directional contradiction, but still represents a clear loss of sensitivity compared with QMP.

4.4 Centered log-ratio method showed direction errors during broad community shifts

Centered log-ratio transformation expresses each family relative to the geometric mean of all families in the same sample (Nearing et al., 2022). This can reduce some compositional effects, but it does not restore bacterial load information. When many families change at the same time, the geometric mean also changes, and the direction of a centered log-ratio change may differ from the direction of an absolute abundance change. This was seen in the V1–V5 contrast. Centered log-ratio methods called several high-abundance families, including *Bacteroidaceae*, *Oscillospiraceae*, and *Lachnospiraceae*, as increased, whereas the qPCR-based QMP reference detected decreases. LinDA, which fits linear models on centered log-ratio-transformed data with bias correction (Zhou et al., 2022), showed a similar but weaker pattern.

Zero replacement may further affect centered log-ratio results. The transformation requires positive values, so zero must be replaced before analysis. This can affect zero or low-abundance families and can make their transformed values partly depend on the sample geometric mean (Austin & Korem, 2025).

Our findings are consistent with recent simulation work showing that centered log-ratio transformation does not necessarily improve microbiota differential abundance analysis compared with relative abundance, and can increase false positives or false negatives compared with absolute abundance data (Hickman & Korpela, 2025). In our cohort, the problem was most visible during the broad post-travel community shift, where centered log-ratio methods sometimes gave significant changes in the opposite direction from the qPCR-based QMP reference.

4.5 ANCOM-BC2 showed low sensitivity, while DESeq2 showed over-detection

ANCOM-BC2 showed very low agreement with qPCR-based QMP and detected few significant families. The bias correction in ANCOM-BC2 estimates the sampling fraction by pooling across taxa under the assumption that most remain stable between compared conditions, a condition Lin and Peddada (2024) state as necessary for reliable inference. In the present cohort, widespread community shifts across all longitudinal intervals challenged this assumption, causing the bias correction to over-correct and suppress the true differential abundance signal. The near-zero detection was not due to the sensitivity score filter, which was not applied in this analysis. Across 53 real microbiome datasets, Pelto et al. (2025) also found that ANCOM-BC2 achieved the lowest replication percentage of all tested methods at 35%, with a conflict percentage of 3%, indicating that the few taxa it called significant were rarely directionally concordant across datasets.

DESeq2 showed a different pattern. It detected more candidate-only significant families and showed weaker standardized effect-size agreement with the qPCR-based reference, especially in the V1–V5 contrast. In agreement with our findings, Nearing et al. (2022), using 38 real cross-sectional microbiome datasets, found that DESeq2 produces high false positive rates relative to more conservative methods; Pelto et al. (2025) reported a Spearman correlation of 0.30 between DESeq2 effect size estimates and the reference across split-data analyses, the lowest among all tested methods.

DESeq2 was developed for RNA sequencing and estimates size factors under the assumption that most features do not change. When many microbes shift together, size factor estimation can become distorted. This can inflate some signals, weaken others, and reduce agreement with the absolute-abundance reference (Hickman & Korpela, 2025; Pelto et al., 2025).

4.6 Taxa-specific concordance was not explained by mean relative abundance

Bacteroidaceae showed the clearest concordance across abundance representations and DA methods, likely because its travel-associated changes were large and relatively consistent across participants (Boolchandani et al., 2022; Henares et al., 2024; Mlangeni et al., 2025). Moreover, members of the family, including *Bacteroides fragilis*, *B. vulgatus*, and *Odoribacter splanchnicus*, have been associated with travellers' diarrhea, and their bloom during diarrheal conditions may reflect a role in regulating inflammation and restoring microbial homeostasis in the human gut during travel (Boolchandani et al., 2022).

In contrast, *Oscillospiraceae* and *Lachnospiraceae* were also highly abundant but showed weaker agreement in effect size and direction across methods. Previous travel studies have reported mixed patterns for these families (Boolchandani et al., 2022; Henares et al., 2024; Mlangeni et al., 2025). In a previous study based on our ETVAX® cohort, several *Oscillospiraceae* taxa, formerly classified within *Ruminococcaceae*, decreased in relative abundance after travel (Mlangeni et al., 2025). This pattern was not reproduced by QMP in the present study, suggesting that the earlier relative-abundance signal may partly reflect compositionality effects. Part of the inconsistency may also come from family-level aggregation. Abundant bacterial families such as *Lachnospiraceae* and *Oscillospiraceae* include genera that can respond differently to travel, diarrhea, or gut transit changes (Procházková et al., 2023). Opposite genus-level change can therefore weaken or obscure the family-level signal.

Overall, mean relative abundance alone did not explain agreement across methods. Concordance depended more on whether a family showed a large, consistent change across participants, and whether that change remained detectable after aggregation at the family level.

4.7 Limitations

This work has several limitations. First, the 16S rRNA gene copy-number correction relies on incomplete reference databases and uncertain taxonomic matching. This problem has been recognized for years and remains unresolved (Louca et al., 2018). In this study, copy-number correction had little effect at the family level, but uncertainty may be larger at finer taxonomic levels. Future work should test newer copy-number prediction methods (Gao & Wu, 2023) and assess whether correction changes results at genus or species level. Second,

this preliminary analysis was limited to bacterial families. Family-level analysis reduces complexity, but it can hide opposing changes among genera or species within the same family. This is especially relevant for large families such as *Lachnospiraceae* and *Oscillospiraceae*. Future analyses should compare family-, genus-, and species-level results to determine whether method agreement improves or worsens at finer resolution. Third, the present cohort represents a complex adult travel setting. Travel may involve dietary changes, pathogen exposure, diarrhea, medication use, inflammation, stress, and altered gut transit. These factors were not considered in the current analysis. Future studies can test whether diarrhea, antibiotic use, stool consistency, or transit-related changes explain the observed shifts in bacterial load and bacterial-to-host read ratios. Fourth, qPCR-based QMP may have inherent biases, though these biases are largely shared with HTS (Jian et al., 2021). Specifically, qPCR measures 16S rRNA gene copies, not viable cells, and can be affected by primer bias, DNA extraction, PCR inhibition, and copy-number correction (Jian et al., 2020, 2021). Flow cytometry and spike-in methods provide useful alternatives, but they measure different targets. Flow cytometry counts stained cells or particles and can be affected by staining, gating, debris, and cell viability, as shown in a recent study to overestimate microbial biomass (Wang et al., 2025). Spike-in methods can match qPCR-based load estimates well in some studies (Wang et al., 2025), but they also depend on when and how the spike-in is added and how closely it follows extraction and sequencing efficiency of the sample bacteria. Future work should compare qPCR, spike-in, and flow cytometry in the same samples rather than assuming one method is universally correct. Finally, the results should be tested in other microbial communities. Lower-complexity communities, such as vaginal or infant gut microbiota, may help separate compositional effects from broad community shifts because fewer dominant taxa drive the profile. Including both lower-complexity and adult perturbation datasets would clarify when relative-abundance methods are acceptable and when measured bacterial load is necessary.

5 Conclusion

This thesis shows that ignoring total bacterial load can distort microbiota differential abundance results. When bacterial load changes, relative-abundance analysis can make the same bacterial family appear increased, decreased, or unchanged depending on the abundance representation and testing method.

A key finding is that more complex methods were not necessarily better. Among approaches without measured bacterial load, total-sum-scaled log-transformed relative abundance gave the closest agreement with qPCR-based quantitative microbiota profiles. In contrast, centered log-ratio methods and several specialized differential abundance methods showed more missed findings, method-only findings, or wrong-direction results.

These results support the use of experimental quantitative microbiota profiling in longitudinal studies. Measured bacterial load is needed to separate true biomass changes from compositional shifts, especially in studies of disease mechanisms, treatment effects, and biomarker discovery. Without it, method-dependent errors can lead to misleading biological conclusions and slow progress in the field (Jian, 2026).

Conflict of Interest

The research was conducted in the absence of any potential conflict of interest.

Funding

This master's thesis was funded by HiLIFE Research Trainee Scholarships and by the Research Program Unit funds at the Faculty of Medicine to Anne Salonen.

Ethical Approval

The study protocol was approved by the Ethics Committee of HUS (HUS/2331/2016) and the Finnish Medicines Agency (EudraCT2016-002690-35). All participants provided written informed consent at their first study visit.

Acknowledgments

I am grateful to my supervisors, Anne Salonen, Ching Jian, and Tuomas Heini. I thank Anne Salonen for her guidance throughout the entire thesis process and for the scientific direction that kept the project on track from start to finish. I especially want to thank Ching Jian for his mentorship in bioinformatics and scientific writing; much of the analysis and clarity in this thesis came from his patient and generous help. I thank Tuomas Heini for his help and insight in the wet-lab work.

I would also like to thank all members of the Microbe Inside group, whose discussions, feedback, and everyday support made the work both possible and enjoyable.

Finally, I thank my friends and colleagues from the MMB master's program for their company and support throughout these studies.

Reference

Armstrong, G., Rahman, G., Martino, C., McDonald, D., Gonzalez, A., Mishne, G., & Knight, R. (2022). Applications and Comparison of Dimensionality Reduction Methods for Microbiome Data. *Frontiers in Bioinformatics*, *2*. https://doi.org/10.3389/fbinf.2022.821861

Austin, G. I., & Korem, T. (2025). Compositional transformations can reasonably introduce phenotype-associated values into sparse features. *MSystems*, *10*(5). https://doi.org/10.1128/msystems.00021-25

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., … Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, *41*(11), 1633–1644. https://doi.org/10.1038/s41587-023-01688-w

Boolchandani, M., Blake, K. S., Tilley, D. H., Cabada, M. M., Schwartz, D. J., Patel, S., Morales, M. L., Meza, R., Soto, G., Isidean, S. D., Porter, C. K., Simons, M. P., & Dantas, G. (2022). Impact of international travel and diarrhea on gut microbiome and resistome dynamics. *Nature Communications*, *13*(1), 7485. https://doi.org/10.1038/s41467-022-34862-w

Constantinides, B., Hunt, M., & Crook, D. W. (2023). Hostile: accurate decontamination of microbial host sequences. *Bioinformatics*, *39*(12). https://doi.org/10.1093/bioinformatics/btad728

Dubois, L., Valles-Colomer, M., Ponsero, A., Helve, O., Andersson, S., Kolho, K.-L., Asnicar, F., Korpela, K., Salonen, A., Segata, N., & de Vos, W. M. (2024). Paternal and induced gut microbiota seeding complement mother-to-infant transmission. *Cell Host & Microbe*, *32*(6), 1011-1024.e4. https://doi.org/10.1016/j.chom.2024.05.004

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., & Gloor, G. B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS ONE*, *8*(7), e67019. https://doi.org/10.1371/journal.pone.0067019

Gamboa-Tuz, S. D., Ramos, M., Franzosa, E., Huttenhower, C., Segata, N., Oh, S., & Waldron, L. (2025). *Commonly used compositional data analysis implementations are not advantageous in microbial differential abundance analyses benchmarked against biological ground truth*. https://doi.org/10.1101/2025.02.13.638109

Gao, Y., & Wu, M. (2023). Accounting for 16S rRNA copy number prediction uncertainty and its implications in bacterial diversity analyses. *ISME Communications*, *3*(1). https://doi.org/10.1038/s43705-023-00266-0

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. In *Frontiers in Microbiology* (Vol. 8, Number NOV). Frontiers Media S.A. https://doi.org/10.3389/fmicb.2017.02224

Hawinkel, S., Rayner, J. C. W., Bijnens, L., & Thas, O. (2020). Sequence count data are poorly fit by the negative binomial distribution. *PLOS ONE*, *15*(4), e0224909. https://doi.org/10.1371/journal.pone.0224909

Henares, D., Monsalvez, V., Brotons, P., Machado, M. L., Capilla, S., Gomila-Grange, A., Bierge, P., Cubero, M., Q. Pich, O., Requena-Méndez, A., Muñoz-Almagro, C., & Gasch, O. (2024). Human gut microbiota composition associated with international travels. *Travel Medicine and Infectious Disease*, *61*, 102747. https://doi.org/10.1016/j.tmaid.2024.102747

Hickman, B., & Korpela, K. (2025). Impact of data compositionality on the detection of microbiota responses. *Gut Microbes*, *17*(1). https://doi.org/10.1080/19490976.2025.2590841

Jian, C. (2026). Navigating the translational labyrinth of the gut microbiota. *Scientific Reports*, *16*(1), 13779. https://doi.org/10.1038/s41598-026-49409-y

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., & Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS ONE*, *15*(1). https://doi.org/10.1371/journal.pone.0227285

Jian, C., Salonen, A., & Korpela, K. (2021). Commentary: How to Count Our Microbes? The Effect of Different Quantitative Microbiome Profiling Approaches. *Frontiers in Cellular and Infection Microbiology*, *11*. https://doi.org/10.3389/fcimb.2021.627910

Jokela, R., Ponsero, A. J., Dikareva, E., Wei, X., Kolho, K.-L., Korpela, K., de Vos, W. M., & Salonen, A. (2023). Sources of gut microbiota variation in a large longitudinal Finnish infant cohort. *EBioMedicine*, *94*, 104695. https://doi.org/10.1016/j.ebiom.2023.104695

Kantele, A., Riekkinen, M., Jokiranta, T. S., Pakkanen, S. H., Pietilä, J. P., Patjas, A., Eriksson, M., Khawaja, T., Klemets, P., Marttinen, K., Siikamäki, H., Lundgren, A., Holmgren, J., Lissmats, A., Carlin, N., & Svennerholm, A. M. (2023). Safety and immunogenicity of ETVAX®, an oral inactivated vaccine against enterotoxigenic Escherichia coli diarrhoea: a double-blinded, randomized, placebo-controlled trial amongst Finnish travellers to Benin, West Africa. *Journal of Travel Medicine*, *30*(7). https://doi.org/10.1093/jtm/taad045

Kleine Bardenhorst, S., Berger, T., Klawonn, F., Vital, M., Karch, A., & Rübsamen, N. (2021). Data Analysis Strategies for Microbiome Studies in Human Populations—a Systematic Review of Current Practice. *MSystems*, *6*(1). https://doi.org/10.1128/mSystems.01154-20

Kumar, B., Lorusso, E., Fosso, B., & Pesole, G. (2024). A comprehensive overview of microbiome data in the light of machine learning applications: categorization, accessibility, and future directions. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1343572

La Torre, D., Van Oudenhove, L., Vanuytsel, T., & Verbeke, K. (2023). Psychosocial stress-induced intestinal permeability in healthy humans: What is the evidence? In *Neurobiology of Stress* (Vol. 27). Elsevier Inc. https://doi.org/10.1016/j.ynstr.2023.100579

Lee, J.-Y., Bays, D. J., Savage, H. P., & Bäumler, A. J. (2024). The human gut microbiome in health and disease: time for a new chapter? *Infection and Immunity*, *92*(11). https://doi.org/10.1128/iai.00302-24

Leigh, S.-J., Uhlig, F., Wilmes, L., Sanchez-Diaz, P., Gheorghe, C. E., Goodson, M. S., Kelley-Loughnane, N., Hyland, N. P., Cryan, J. F., Clarke, G., & Cryan Her, J. F. (2023). The impact of acute and chronic stress on gastrointestinal physiology and function: a microbiota-gut-brain axis perspective The stress response: from adaptation to disease. *The Journal of Physiology*, *601*, 4491–4538. https://doi.org/10.1113/JP281951#support-information-section

Lin, H., & Peddada, S. Das. (2024). Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nature Methods*, *21*(1), 83–91. https://doi.org/10.1038/s41592-023-02092-7

Lloréns-Rico, V., Vieira-Silva, S., Gonçalves, P. J., Falony, G., & Raes, J. (2021). Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23821-6

Louca, S., Doebeli, M., & Parfrey, L. W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome*, *6*(1), 41. https://doi.org/10.1186/s40168-018-0420-9

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Martínez Arbas, S., Busi, S. B., Queirós, P., de Nies, L., Herold, M., May, P., Wilmes, P., Muller, E. E. L., & Narayanasamy, S. (2021). Challenges, Strategies, and Perspectives for Reference-Independent Longitudinal Multi-Omic Microbiome Studies. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.666244

McArdle, A. J., & Kaforou, M. (2020). Sensitivity of shotgun metagenomics to host DNA: abundance estimates depend on bioinformatic tools and contamination is the main issue. *Access Microbiology*, *2*(4). https://doi.org/10.1099/acmi.0.000104

Mlangeni, T., Jian, C., Häkkinen, H. K., de Vos, W. M., Salonen, A., & Kantele, A. (2025). Travel to the tropics: Impact on gut microbiota. *Travel Medicine and Infectious Disease*, *66*, 102869. https://doi.org/10.1016/j.tmaid.2025.102869

Morton, J. T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L. S., Edlund, A., Zengler, K., & Knight, R. (2019). Establishing microbial composition measurement standards with reference frames. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10656-5

Nearing, J. T., Douglas, G. M., Hayes, M. G., MacDonald, J., Desai, D. K., Allward, N., Jones, C. M. A., Wright, R. J., Dhanani, A. S., Comeau, A. M., & Langille, M. G. I. (2022). Microbiome

differential abundance methods produce different results across 38 datasets. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-28034-z

Nishijima, S., Stankevic, E., Aasmets, O., Schmidt, T. S. B., Nagata, N., Keller, M. I., Ferretti, P., Juel, H. B., Fullam, A., Robbani, S. M., Schudoma, C., Hansen, J. K., Holm, L. A., Israelsen, M., Schierwagen, R., Torp, N., Telzerow, A., Hercog, R., Kandels, S., … Thiele, M. (2025). Fecal microbial load is a major determinant of gut microbiome variation and a confounder for disease associations. *Cell*, *188*(1), 222-236.e15. https://doi.org/10.1016/j.cell.2024.10.022

Pelto, J., Auranen, K., Kujala, J. V., & Lahti, L. (2025). Elementary methods provide more replicable results in microbial differential abundance analysis. In *Briefings in Bioinformatics* (Vol. 26, Number 2). Oxford University Press. https://doi.org/10.1093/bib/bbaf130

Procházková, N., Falony, G., Dragsted, L. O., Licht, T. R., Raes, J., & Roager, H. M. (2023). Advancing human gut microbiota research by considering gut transit time. *Gut*, *72*(1), 180–191. https://doi.org/10.1136/gutjnl-2022-328166

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., & Schmidt, T. M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, *43*(D1), D593–D598. https://doi.org/10.1093/nar/gku1201

Sui, H., Weil, A. A., Nuwagira, E., Qadri, F., Ryan, E. T., Mezzari, M. P., Phipatanakul, W., & Lai, P. S. (2020). Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00953

Tang, G., Carr, A. V., Perez, C., Ramos Sarmiento, K., Levy, L., Lampe, J. W., Diener, C., & Gibbons, S. M. (2025). Metagenomic estimation of absolute bacterial biomass in the mammalian gut through host-derived read normalization. *MSystems*, *10*(8). https://doi.org/10.1128/msystems.00984-25

Vandeputte, D., De Commer, L., Tito, R. Y., Kathagen, G., Sabino, J., Vermeire, S., Faust, K., & Raes, J. (2021). Temporal variability in quantitative human gut microbiome profiles and implications for clinical research. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27098-7

Vandeputte, D., Kathagen, G., D'Hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., Wang, J., Tito, R. Y., De Commer, L., Darzi, Y., Vermeire, S., Falony, G., & Raes, J. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, *551*(7681), 507–511. https://doi.org/10.1038/nature24460

VanEvery, H., Franzosa, E. A., Nguyen, L. H., & Huttenhower, C. (2023). Microbiome epidemiology and association studies in human health. *Nature Reviews Genetics*, *24*(2), 109–124. https://doi.org/10.1038/s41576-022-00529-x

Wang, S., Healy, D., Patangia, D., Uniacke-Lowe, S., Kamilari, E., Kozak, I. M., Yang, B., Dempsey, E. M., Stanton, C., & Ross, R. P. (2025). Assessment of absolute abundance in mother-infant gut microbiome using marine-sourced bacterial DNA spike-in and comparison with conventional quantification methods. *Microbiome Research Reports*, *4*(2). https://doi.org/10.20517/mrr.2024.94

Wirbel, J., Andermann, T. M., Brooks, E. F., Evans, L., Groth, A., Dvorak, M., Chakraborty, M., Palushaj, B., Reynolds, G. Z. M., Porter, I. E., Al Malki, M., Rezvani, A., Gooptu, M., Elmariah, H., Runaas, L., Fei, T., Martens, M. J., Bolaños-Meade, J., Hamadani, M., … Bhatt, A. S. (2025). Accurate prediction of absolute prokaryotic abundance from DNA concentration. *Cell Reports Methods*, *5*(5). https://doi.org/10.1016/j.crmeth.2025.101030

Yang, L., & Chen, J. (2023). Benchmarking differential abundance analysis methods for correlated microbiome sequencing data. In *Briefings in Bioinformatics* (Vol. 24, Number 1). Oxford University Press. https://doi.org/10.1093/bib/bbac607

Zhou, H., He, K., Chen, J., & Zhang, X. (2022). LinDA: linear models for differential abundance analysis of microbiome compositional data. *Genome Biology*, *23*(1), 95. https://doi.org/10.1186/s13059-022-02655-5