# Namco: A microbiome explorer

Alexander Dietrich[1,*], Monica Steffi Matchado[1,*], Maximilian Zwiebel[1], Benjamin Ölke[1], Michael Lauber[1], Ilias Lagkouvardos[2], Jan Baumbach[4,5], Dirk Haller[2,3], Beate Brandl[2], Thomas Skurk[2], Hans Hauner[2,6], Sandra Reitmeier[2,3,†], Markus List[1,†]

[1]Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany
[2]ZIEL - Institute for Food & Health, Technical University of Munich, 85354 Freising, Germany
[3]Chair of Nutrition and Immunology, TUM School of Life Sciences, Technical University of Munich, 85354 Freising, Germany
[4]Institute of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark
[5]Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany
[6]Institute of Nutritional Medicine, TUM School of Medicine, Technical University of Munich, Munich, Germany

[†] joint last author; [*] joint first author

## Abstract

**Background:** 16S rRNA gene profiling is currently the most widely used technique in microbiome research and allows for studying microbial diversity, taxonomic profiling, phylogenetics, functional and network analysis. While a plethora of tools have been developed for the analysis of 16S rRNA gene data, only few platforms offer a user-friendly interface and none comprehensively cover the whole analysis pipeline from raw data processing down to complex analysis. **Results:** We introduce Namco, an R shiny application that offers a streamlined interface and serves as a one-stop solution for microbiome analysis. We demonstrate Namco's capabilities by studying the association between rich fibre diet and the gut microbiota composition. Namco helped to prove the hypothesis that butyrate-producing bacteria are prompted by fibre-enriched intervention. **Conclusion:** Namco provides a broad range of features from raw data processing and basic statistics down to machine learning and network analysis, thus covering complex data analysis tasks that are not comprehensively covered elsewhere. Namco is freely available at https://exbio.wzw.tum.de/namco/.

## 1. Background

Over the past decade, microbiome research has contributed to our understanding of human health and microbiome-associated diseases with implications for diagnosis, prevention and treatment [1]. Several studies have discovered important links between the gut microbiome and human diseases including diabetes [2], cancer [3], inflammatory bowel disease [4] and brain disorders [5]. Currently, microbiome datasets are generated using either targeted (amplicon) gene sequencing to characterize the microbial composition and phylogeny or shotgun metagenomics to study, in addition, the gene-coding and functional potential of the microbiome. Due to the comparably lower sequencing costs, sequencing of the 16S rRNA gene is the most commonly available method. Analysis of 16S (or 18S in the case of eukaryotes) rRNA genes can be grouped into four steps. First, amplicons are clustered into operational taxonomic units (OTUs) [6] or amplicon sequence variants (ASVs) [7]. For OTUs, sequencing errors are addressed by choosing a similarity threshold of typically 97% for clustering whereas the latter employs a denoising strategy to identify the unique error-corrected sequence of an organism [8]. Several benchmark studies proved that denoising methods provide better resolution and accuracy than OTU clustering methods [7, 9]. Along with taxonomic profiling, various measures of alpha and beta diversity are typically computed to study the microbial diversity within and between samples, respectively. Additional analysis steps include, for instance, (i) differential abundance analysis to identify bacteria and/or functions that differ between groups of interest, (ii) in silico inference of metagenomes for functional profiling and (iii) microbial association analysis through correlation, co-occurrence and network inference methods.

Many open-source tools have been developed for microbiome data analysis (Table 1). *Mothur* [10], *QIIME2* [11] and *DADA2* [8] offer processing of raw sequencing files through clustering and annotation of 16S rRNA genes and provide OTU or ASV tables which serve as an input for further downstream analysis. The *QIIME2* [11] pipeline offers more than 20 plugins for downstream analysis including q2-sample-classifier for supervised classification and regression analysis [12], q2-longitudinal for time-series analysis [13], and plugins for compositional data analysis [14]. In addition, a plethora of R packages have been implemented to perform statistical analysis and high quality visualizations on amplicon tables, like *Phyloseq* [15, 16] and *themetagenomics* [17]. Even though R scripts leveraging these tools offer a powerful approach to analyze microbial data, their application can be demanding for users without scripting knowledge and bioinformatics training. Hence, there is a need for user-friendly tools that can support end-to-end exploration of microbiomes. To this end, several web-based tools have been developed including *microbiomanalyst* [18], *IMNGS* [19], *iMAP* [20], *MGRAST* [21], *wiSDOM* [22], *VAMPS* [23], *Shiny-phyloseq* [16] etc. However, most of these tools (i) cover only the downstream part of the analysis and often omit raw data processing, (ii) offer only standard analysis and are thus not sufficient for more complex data sets, (iii) omit functional profiling or use outdated approaches such as *Tax4FUN* [24] and *PICRSUt1* [25] which were outperformed by *PICRUSt2* [26], (iv) do not offer confounder analysis (v) offer no support for time-series and (vi) lack support for machine learning and (vii) do not construct microbial association networks and differential networks on different taxonomic ranks.

To address these limitations, we introduce Namco, an R shiny application that offers a streamlined and user-friendly interface and serves as a one-stop solution for microbiome analysis. Namco provides a broad range of features from raw data processing and basic statistics down to machine learning and network analysis, thus covering complex data analysis tasks that are not comprehensively covered elsewhere (see Table 1 for a comparison with other tools). Namco's thoroughly documented and easy-to-use graphical interface is intended to eliminate the use of command-line arguments during the data processing, making advanced microbiome analysis accessible to a broad range of biomedical researchers. Privacy legislation such as the European Union's General Data Protection Regulation (GDPR) can prevent users from uploading their data to web tools such as Namco. Hence, we make Namco available under an open source license and release it as a Docker container, which can be executed locally (e.g. using Docker Desktop) or can be safely deployed on a local protected server.

## 2. Material and methods

### 2.1. Input file formats

For raw data processing, Namco accepts both single and paired-end FASTQ files which are processed internally based on *DADA2* [8]. Alternatively, users can start their analysis with a previously generated OTU or ASV table which can either include relative abundance values or read counts. For simplicity, we refer to these as features throughout the manuscript. Features should be labelled either by their taxonomy or alternatively by a unique id which Namco can map to a separate tabular file with taxonomic labels that can be uploaded separately by the user. Optionally, users can upload a meta data file containing additional information for each sample which can be leveraged for groupwise differences, correlation, confounder or condition-specific analysis, for instance. Finally, users can optionally provide a tree file in newick format for phylogenetic tree analysis, diversity analysis or to create ecologically-organized heatmaps. The entire workflow of Namco is presented in Figure 1.

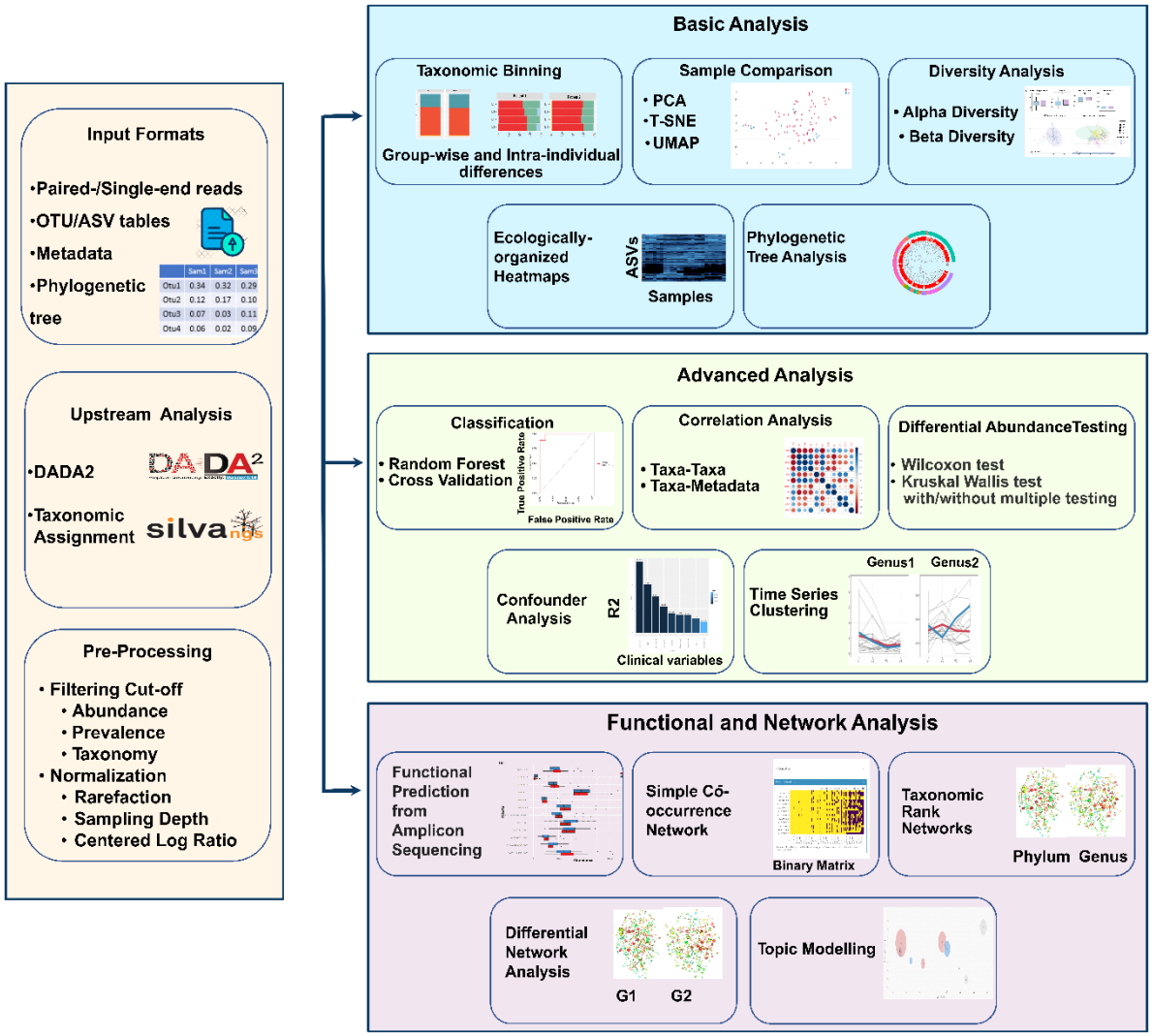


**Figure 1**: Overall workflow of Namco. Namco provides a comprehensive end-to-end analysis of microbiome data including raw FASTQ processing and filtering, down to statistical, functional and network analysis. It also provides various tables and visualization options and allows users to navigate through different data analysis tasks.

**Table 1: Comparisons of Namco with other web-based tools for microbiome data analysis**

| Features | Namco | Microbiome Analyst | MG-RAST | VAMPS | Shiny-phyloSeq | animalcules | GenePiper | METAGEN assist | iMAP | wiSDOM | EzMAP | Metavizr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Raw sequence processing | ☑ | ☐ | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | ☑ | ☐ | ☑ | ☑ |
| Filtering | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ |
| Normalization | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☐ | ☐ | ☐ | ☐ |
| Taxonomic/sample overview | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ |
| Alpha-Diversity | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☑ | ☑ | ☐ |
| Beta-Diversity | ☑ | ☑ | ☑ | ☐ | ☐ | ☑ | ☑ | ☑ | ☐ | ☐ | ☑ | ☐ |
| Statistical tests between taxa | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☑ | ☑ | ☑ | ☐ |
| Statistical tests between samples | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | ☐ | ☑ | ☑ | ☐ |
| Rarefaction | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☑ | ☐ | ☐ | ☑ | ☐ | ☐ |
| Confounder Analysis | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Ordination methods | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ |
| Clustering | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ |
| Time Series Analysis | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Functional prediction | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ |
| Pathway visualization | ☐ | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Taxonomic differential analysis | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☐ | ☑ | ☐ | ☐ | ☑ | ☑ |
| Machine learning & classification | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☐ | ☑ | ☐ | ☑ | ☐ | ☐ |
| Taxon Set-Enrichment analysis | ☐ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Integration with public data | ☐ | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ |
| Co-occurrence network | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Topic modelling | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Single network Analysis | ☑ | ☐ | ☐ | ☐ | ☑ | ☐ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Differential networks analysis | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Web-tool | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☐ | ☑ | ☐ | ☐ |
| Open source | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | ☐ | ☑ | ☑ | ☑ | ☑ |
| Availabe as Docker image | ☑ | ☐ | ☐ | ☐ | ☐ | ☑ | ☐ | ☐ | ☑ | ☐ | ☐ | ☐ |

## 2.2. Denoising and taxonomic assignment

Namco accepts both single- and paired-end reads in FASTQ format and performs denoising using the *DADA2* [8] algorithm. Users can change the default parameters such as trimming length depending on the amplicon length and the quality score threshold. The SILVA (v. 138) [27] database is used as reference for taxonomic classification. Namco stores the *DADA2* output as a phyloseq object using the *Phyloseq* R [15] package and passes it down to further downstream analysis. Alternatively, users can perform their own upstream process elsewhere and upload abundance tables and metadata as inputs to Namco.

## 2.3. Data overview and filtering

The data overview section in Namco summarizes sample details, total number of features identified during the denoising step, and the number of groups provided by the meta file. This gives an overall picture of the input data and will help in preparing the input data for further processing. Normalization and filtering are considered as crucial steps in microbial analysis [28, 29]. By default, Namco applies 0.25% of the abundance filter on the DADA2-generated features and normalizes abundance to 10,000 reads before downstream analysis. Filtering the abundance of both OTUs and ASVs > 0.25% [30] was identified as an effective threshold to prevent the identification of spurious taxa to a large extent. Alternatively, users can also choose different normalization methods and filtering percentages such as sampling depth, rarefaction and centered log ratio transformation [31]. In addition to the normalization methods, Namco offers different filtering options based on sample prevalence and relative abundance as microbiome data are very

sparse in nature and often have zero counts in most samples. These rare taxa are caused by sequencing artefacts, contamination and/or sequencing errors [30].

## 2.4. Basic analysis

### 2.4.1. Visualization of taxonomic binning, sample comparison and ecologically-organized heatmap

Namco integrates R-scripts from *Rhea* [32] and phyloseq [15] to perform taxonomic profiling and diversity analysis and provides different options to visualize the distribution of dominant taxa at different ranks (domain, phylum, class, order, family and genus) among groups using barplots. In addition, taxonomic distribution can also be inferred based on intra-individual differences by visualizing taxa for individual samples. Users can download feature tables of relative abundances aggregated at different taxonomic levels and export any of the generated plots. Relative similarity of datasets and metadata can also be visualized using PCA, UMAP and t-SNE. The advanced heatmap option in Namco creates a heatmap using ordination methods such as NMDS and PCA to organize the rows and columns instead of hierarchical clustering approach which gives an overview of the abundance of features across sample groups which are very high/low abundant.

### 2.4.2. Diversity analysis

Alpha-diversity quantifies the diversity of the microbiome within a group. Namco supports five common alpha-diversity measures, namely *Shannon entropy* [33] *and Simpson index* [34] together with their counterparts accounting for the effective number of species [35] as well as richness. Users can select different categories from the metadata to visualize alpha-diversity and determine significant differences via a Wilcoxon test. Beta-diversity analysis explains the variation between groups and relies on a phylogenetic tree as input along with the feature table to calculate dissimilarity. Namco supports the most common distance metrics including *weighted* and *unweighted unifrac distances*, *generalized unifrac* [36], *variance adjusted unifrac distance* and *Bray Curtis dissimilarity* [37]. Calculation of *unifrac distances* is only possible if a tree file was uploaded. The results are presented using *multi-dimensional scaling (MDS)* and *non-metric multidimensional scaling (NMDS)*. In addition, it is also possible to visualize the distance as a hierarchically clustered dendrogram which helps to identify closely related samples. Significance between groups is determined by a permutational multivariate analysis of variances using the adonis function of the *vegan* R-package [38]. P-values are corrected for multiple testing following Benjamini-Hochberg (BH) [39].

## 2.5. Differential analysis

### 2.5.1. Differential abundance testing using simple statistical tests and association analysis

A key aim in microbial research is to identify differences in microbial composition between conditions or phenotypes. Namco reports statistically significant features between the sample groups using the non-parametric Wilcoxon test (*SIAMCAT* R-package [40]) which was shown to reliably control the false discovery rate in differential abundance analysis [41]. Users can choose groups which should be compared against each other and adjust the significance level as well as other filtering parameters. Differential abundance can be calculated at different taxonomic levels such as phylum or genus, where Namco aggregates the feature table accordingly. Namco shows the distribution of microbial relative abundance along with the significance and a generalized fold change [42] as a nonparametric measure of effect size. In addition to the Wilcoxon test, Namco also offers Kruskal-Wallis test with multiple testing to find significant differences between more than two groups.

### 2.5.2. Correlation analysis

Namco provides correlation analysis to reveal significant associations between taxa or between taxa and metadata such as continuous experimental variables. Namco further considers relative abundances of features at different levels (phylum, class, order etc).

### 2.5.3. Topic modelling

Topic modeling was originally designed to uncover hidden thematic structures in document collections [43]. This concept was adapted to metagenomic analyses to explore co-occurring taxa as topics and to find topics associated with provided sample metadata [44]. Namco employs the *themetagenomic R* [44] package to predict topics and to study their association with sample metadata which can be continuous, binary, categorical, or factor covariates.

## 2.6. Functional profiling

Microbial composition varies widely between individuals, making the robust identification of phenotype-associated microbial features challenging. One can hypothesize that the functional potential of the microbiome is more robustly associated with a phenotype than the microbial composition. While investigating the functional potential of the microbiome is not directly feasible with 16S rRNA gene sequencing data, several tools have been proposed for inferring the functional profile with the help of reference sequencing databases. To this end, Namco adapts the *PICRUSt2* [26] approach, which showed improved accuracy and flexibility compared to related tools including *PICRUSt, Tax4FUN2, Piphillin* and *PanFP* [26]. Namco also performs differential analysis on the predicted KEGG orthologs, enzyme classification numbers and pathways using *Aldex2* [45], which was recently reported to perform best for this type of analysis [46]**.** The relative abundances of significant KEGG annotation terms are plotted in a barplot along with the p-value.

### 2.7. Phylogenetic tree analysis

Phylogenetic analysis belongs to the basic steps to get an overview of the evolutionary relations between features. Namco displays the provided/calculated phylogenetic tree in circular or rectangular format. In addition, users can add two heatmap layers as taxonomic ranks and/or a meta-group. The meta-group heatmaps are colored by the abundance of the features in the corresponding meta-group.

### 2.8. Network analysis

Several methods for inferring and analysing microbial co-occurrence networks were developed to study the role of microbial interactions in association with the host [47, 48]. Namco implements multiple strategies for network construction. As a simple approach, the feature abundance matrix is converted into a binary indicator matrix using an abundance cutoff (presence / absence) (default cut-off is 1: all features with a value less than 1 are considered absent, while the rest is considered present.) This cutoff can be adapted manually to get a more strict binary representation of the abundance matrix. Next, the number of co-occurring feature pairs are counted across samples for each groups (e.g. case and control-pairwise) and the difference in co-occurrence counts as well as the log2 fold-change between the groups is calculated and displayed as a network where nodes represent features and edges represent frequent group-specific interactions. For more advanced approaches, Namco employs the *NetComi* [49] R package, where users can build microbial association networks at different taxonomy levels using nine different network construction algorithms. In addition, *NetComi* offers a method for differential network analysis between two conditions to identify pairs of taxa differentially associated between two groups.

### 2.9. Confounder analysis and explained variation

Confounding variables may mask the actual relationship between the dependent and independent variables in a study [50]. Especially microbiome composition is associated with several host variables including body mass index (BMI), sex, age and geographical location, among others [51]. Namco utilizes the permutational multivariate analysis of variances (adonis function of the *vegan R*-package) [38] to rule out confounding factors using available information from the user-provided metadata table. The explained variation of covariates is determined by $R^2$ values which are considered significant at p-value $\leq 0.05$.

### 2.10. Classification based on random forest

Beyond differential abundance analysis, an important question is if a classification model can be trained on a minimal set of features to robustly predict the outcome (e.g. disease state or treatment response). Such models highlight the potential of microbiome data for prognostic and diagnostic purposes through biomarkers and surrogate endpoints. Namco allows users to build classification models and to identify important features using machine learning algorithms. Within Namco, random forest (*ranger* [52] R package) is used as a classification tool, since it has shown good performance even on comparably small sample sizes in microbial data analysis [53, 54]. By

default, Namco splits the data into training and test sets and performs 10-fold repeated cross-validation. Experienced users can modify advanced parameters such as the ratio of training and test sets, the number of cross validation folds, the resampling method and the number of decision trees. The results are summarized in a confusion matrix and a receiver-operator-characteristic (ROC) plot which helps in evaluating the model performance. The most informative features that were used for classification can be extracted as biomarker candidates for hypothesis generation and further research.

### 2.11. Time series analysis and clustering

Time series analysis in Namco allows users to determine how microbial communities including taxa, OTU / ASV, and other features like richness change over time. For instance, time series analysis helps to study the microbial changes in response to a treatment over multiple timepoints or during different stages of host development. Namco offers different options to modify the inputs for time series line charts including displays of changes in either relative abundance or absolute abundance or richness.

## 3. Use case

To illustrate the broad utility of Namco, we analyzed human fecal samples from an interventional cross-over study. The study's aim was to develop healthier convenience food products with an increased fibre content and to foster customer acceptance of such products. Here, we analyzed if the stool microbiota is altered by the fibre-rich diet.

### 3.1. Ethics statement

The study protocol was approved by the ethical committee of the Faculty of Medicine of the Technical University of Munich in Germany (approval no. 529/16S). The guidelines of the International Conference on Harmonization of Good Clinical Practice and the World Medical Association Declaration of Helsinki (in the revised version of Fortaleza, Brazil 2013) were considered. All study participants have given written informed consent. The study was registered at the German Clinical Trial Register (DRKS00011526).

### 3.2. Study design

The human intervention study was a single-blinded, controlled cross-over study. Volunteers were recruited from a cohort of middle-aged subjects who were broadly phenotyped within the *enable* nutrition cluster, 50% were males and 50% females. Inclusion criteria were 40-65 years old volunteers with an elevated waist circumference. For detailed information on inclusion and exclusion criteria of the enable cohort see Brandl et al [55]. Study participants were invited four times to the study center. During the first visit baseline characteristics were collected. In general, the intervention of meatloaf in a bun and pizza was performed as described in Rennekamp et al [56].

### 3.3. Phenotypic characteristics of the study group

The study group was age and sex-matched (N = 11 females, N = 10 males) and received the same intervention and placebo (Table 2). Baseline measurements were performed in the morning after an overnight fasting. Body composition and body weight were measured by using a Seca Medical Body Composition Analyser, mBCA 515, Seca GmbH & Co. KG, Hamburg, Germany). Body height was measured in a standing position without shoes using a stadiometer (Seca GmbH & Co. KG, Hamburg, Germany). BMI was calculated as weight (kg)/height (m2). Waist circumference was measured at the midpoint between the lowest rib and the iliac crest with a measuring tape (Seca GmbH & Co. KG, Hamburg, Germany).

**Table 2. Overview of the study group characteristics. Indication of the mean values and the standard deviation for the participants is given, besides significant difference in traits between the sexes.**

| | **Mean** | **SD** | **Differences between sexes** |
|---|---|---|---|
| Weight [kg] | 90.14 | 11.42 | 0.0080 (**) |
| Height [m] | 1.73 | 0.08 | 1.35e-05 (***) |
| BMI [kg/m2] | 30.12 | 2.41 | 0.8490 (ns) |
| Fat-free mass [%] | 62.98 | 6.74 | 2.40e-08 (***) |
| Fat mass [%] | 37.02 | 6.74 | 2.40e-08 (***) |
| Skeletal muscle mass [kg] | 27.55 | 6.28 | 9.34e-07 (***) |
| Visceral fat [kg] | 3.24 | 1.32 | 4.55e-05 (***) |
| Waist circumference [cm] | 101.3 | 7.26 | 0.0058 (**) |

### 3.4. Sample preparation

The participants were asked to visit the study center sober (10 hours before the visit) and received the intervention or placebo meal in the study center. Additionally, a capsule with food coloring was administered. The intake of the dye stains the stool green which helps to associate collected samples and food intake.

The time of the meal, as well as the time of the excretion, were recorded and showed a mean transit time of 34.74 ± 24.69 h. Since a coloring capsular was administered together with the meals, the stool sample can be assigned to the meal. The dye causes a visible green coloration of the sample, and a recognizable coloration was noted in the data.

Participants consumed two different types of food (meatloaf in a bun and a pizza) both, either enriched with fibre (intervention) (IM) or not (placebo) (M). The first interventional meal

(meatloaf in a bun, IM1) the white bread roll in the fibre-enriched meal contained additional 5.7% wheat fibre (VITACEL® WF600) and the meatloaf (Leberkas) a mixture of 3.1% wheat fibre and 4.5% resistant dextrin.
The second intervention (pizza, IM2) was also fibre- enriched containing up to 20 g fibre with 3.0 % wheat fibre, 2.4% powdered cellulose and 2.1% inulin (Table 3). The intervention meals thus constituted a major part of the recommended daily fibre intake. As the fibre content is above 6 g per 100 g, the food products are considered to count as high fibre products.

**Table 3. Nutritional values per serving for the intervention (enriched) and the placebo (standard) meatloaf and salami pizza meal as well as the difference between intervention and placebo meal**

| | **Portion meatloaf with bun 240 g** | | **Portion salami pizza 320 g** | |
|---|---|---|---|---|
| | **Enriched** | **Standard** | **Enriched** | **Standard** |
| Energy [kcal] | 413 | 587 | 829 | 876 |
| Fat [g] | 13 | 35 | 41 | 45 |
| Carbohydrate [g] | 47 | 47 | 75 | 83 |
| Total fibre [g] | 19 | 2.9 | 20 | 6 |

### 3.5. Sample preparation and sequencing

For the analysis of gut microbiota, the 16S rRNA gene was sequenced at the ZIEL Core Facility Microbiome, Technical University Munich, Germany. Detailed description of the samples preparation and sequencing are described elsewhere [57]. Briefly, sample DNA was isolated following an in-house developed protocol. Targeting the V3V4 region of the 16S rRNA gene, samples were amplified and purified. Pooled amplicons were paired-end sequenced on an Illumina MiSeq. Sequencing data is available under BioProject ID PRJNA774891

## 4. Research question

An increased fibre intake was shown to be protective against the development of cardiovascular [58, 59] and malignant diseases [60, 61] and there are specific health claims associated with specific types of fibres. In this study, we examined the presence of butyrate-producing bacteria which could be promoted by the fibre-enriched intervention and thus prove that dietary aspects can have a permanent effect on the gut microbial composition.

## 5. Results

### 5.1. Diversity analysis

We studied changes in the microbial composition following dietary intervention using Namco. Paired-end FASTQ files were processed using the DADA2 denoising step embedded in Namco with the default parameters. During the DADA2 step, a 0.25% abundance-based filter was applied to reduce sparsity [30]. ASVs were normalized to 10,000 reads before downstream analysis and outliers were removed. Additionally, a prevalence filter cutoff of 10% was introduced. No significant differences were observed in alpha-diversity measures between the IM and M groups: Shannon, richness, Simpson Index, effective Shannon entropy or effective Simpson entropy. Likewise, no significant clustering was found in beta-diversity including unweighted and weighted Unifrac between IM and M (Figure 2).

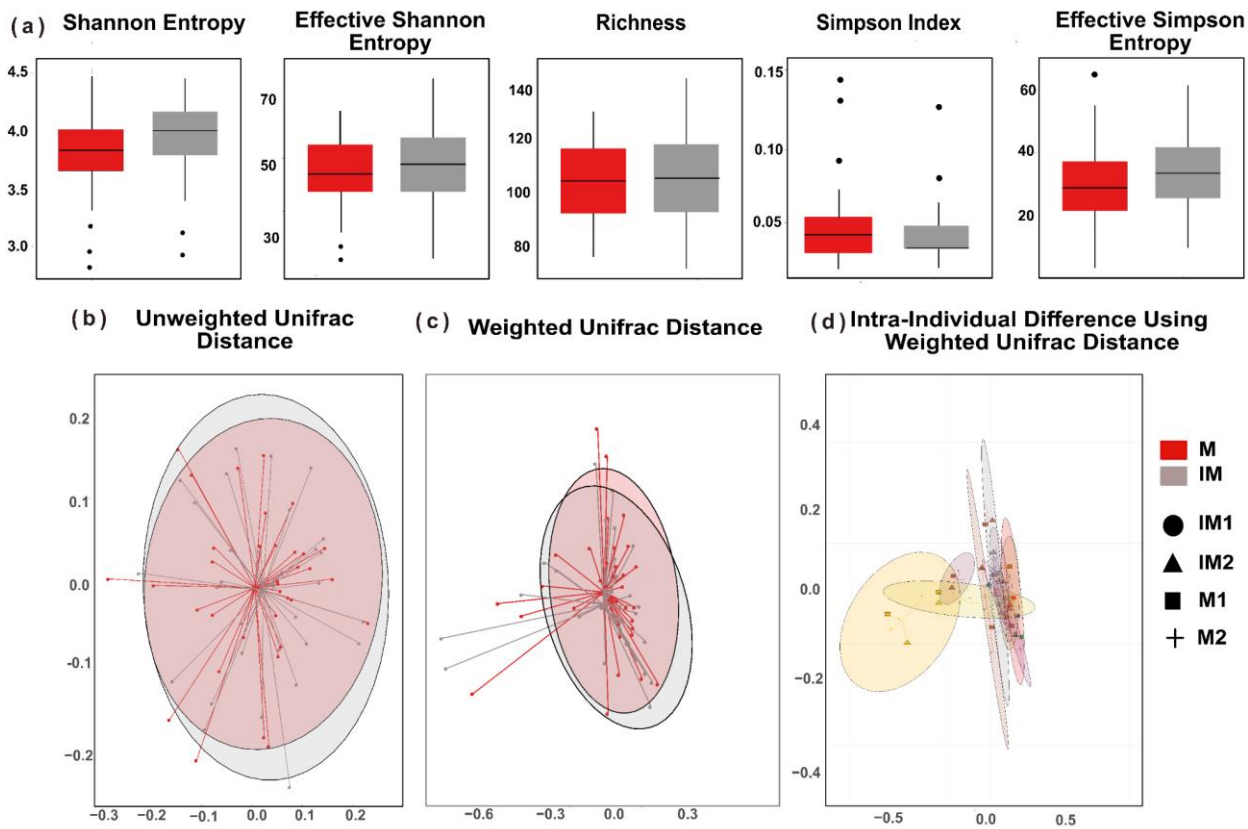


**Figure 2 (a):** Alpha diversity measures associated with intervention (IM) and non-intervention meals (M). There was no significant difference identified between the groups. (b) nMDS visualizations of beta diversity analysis using the unweighted or weighted (c) unifrac distance. (d) nMDS visualizations of beta diversity for intra-individual patients across two intervention meals and their respective control using weighted unifrac distance.

## 5.2.Taxonomic distribution

Overall, 6 phyla and 77 genera were observed in all groups. Dominating phyla in both intervention (IM1 and IM2) and non-intervention groups (M1 and M2) were Firmicutes and Bacteroidota, contributing up to 90% to the total bacterial composition. The relative abundance of Firmicutes was found to be slightly higher in the IM group compared to the M group. Bacteroidota were slightly more abundant in the M group. Other phyla such as Actinobacteria, Verrucomicrobia and Proteobacteria showed < 5% mean relative abundance between IM and M groups (**Figure 3 (a))**. *Ruminococcaceae Incertae Sedi*s was significantly different between the IM and M groups. Overall, except *Ruminococcaceae Incertae Sedi*s, no other significant difference was observed at the phylum level between IM and M groups. However, intra-individual heterogeneity was observed at the abundance on phylum level **(Figure 3 (b))** at each intervention. On the phylum level, we observed differences in the relative abundance of Firmicutes and Actinobacteria from 34% to 77% and from 0.46% to 10.56%, in the IM and M, respectively. Similarly, the relative abundance of Proteobacteria also varied from 0.03% to 5.6 % and 0.07% to 8.6% in the IM and M, respectively . On the genus level, most of the individuals showed a uniform distribution except for one individual who showed a high level of *Prevotella* (54% of relative abundance).

The top 20 genera are shown in **Figure 3(c)** and intra-individual differences are plotted in **Figure 3(d)**. *Bacteroides was* the most abundant genus in both IM and M groups followed by *Faecalibacterium, Prevotella and Agathobacter.* Due to the similarity observed at the genus level between IM and M meals, a non-parametric paired Wilcoxon test was applied on the relative abundance to identify microbial changes between (IM1, IM2) and their respective controls (M1 and M2). Samples with missing information regarding what kind of meals were adminintrated were removed during the following analysis. In total, five genera: *Anaerostipes, Ruminococcaceae Incertae Sedi*s, *Parabacteroides, Fusicatenibacter and Butyricicoccus* were significantly different in abundance prior to multiple corrections between intervention meals and normal meals. After multiple testing correction with BH, only *Ruminococcaceae Incertae Sedi*s remained significant. Relative abundance of *Anaerostipes, a* butyrate-producing bacterium, was found to be higher in IM2 compared to M2. *Anaerostipes* is a gram positive and anaerobic bacterial from the family of *Lachnospiraceae* and found to be highly expressed in a normal healthy gut [62]. As a validation, previous studies stated that the abundance of *Anaerostipes* increases with fibre-rich diets and negatively correlated with BMI [63, 64]. *Ruminococcaceae Incertae Sedi*s and *Parabacteroides* also showed a significant difference between IM1 and M1 **(Figure 4)**. *Ruminococcaceae* is also known for producing short-chain fatty acids (SCFA) including butyrate which promotes a healthy bowel [65] and is nominally protective of weight gain [66]. In addition, *Ruminococcus bromii* is reported as the key species in fermenting resistant starch, which in turn helps in conferring health benefits including weight control and protection against diabetes [67]. *Parabacteroides* have been reported to have metabolic benefits and negative correlation with BMI [68, 69]. One of the species of *Parabacteroides (P. distasonis)* is also reported to be part of the core gut microbiome [70–72] and has the ability to produce succinic acid and to promote the increase of bile acid in terms of regulating host metabolism [69, 73]. Ezeji et al. also found *Parabacteroides* to be enriched in fibre-rich dietary intervention groups [74]. At the genus level, dietary fibre intervention significantly promoted the growth of beneficial genera *Anaerostipes, Ruminococcaceae Incertae*

*Sedi*s and *Parabacteroides.* Additionally, *Fusicatenibacter* was found significantly higher in IM1 than M2 and *Butyricicoccus* was found higher in the M1 group compared to IM2.

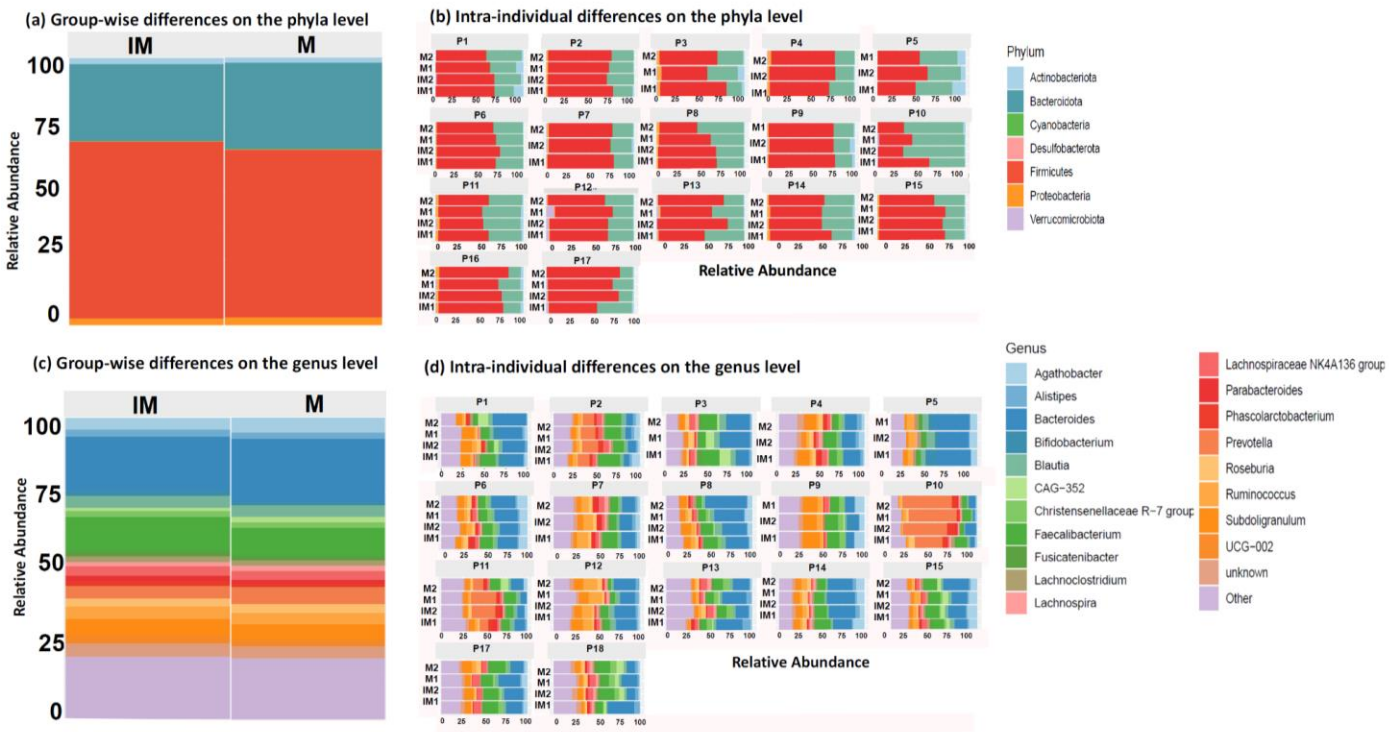


**Figure 3:** (a, c) Relative abundances of phyla and genus between intervention and non-intervention groups. (b,d) Bar plots show inter-individual variation in the gut microbiome between intervention and non-intervention meals.

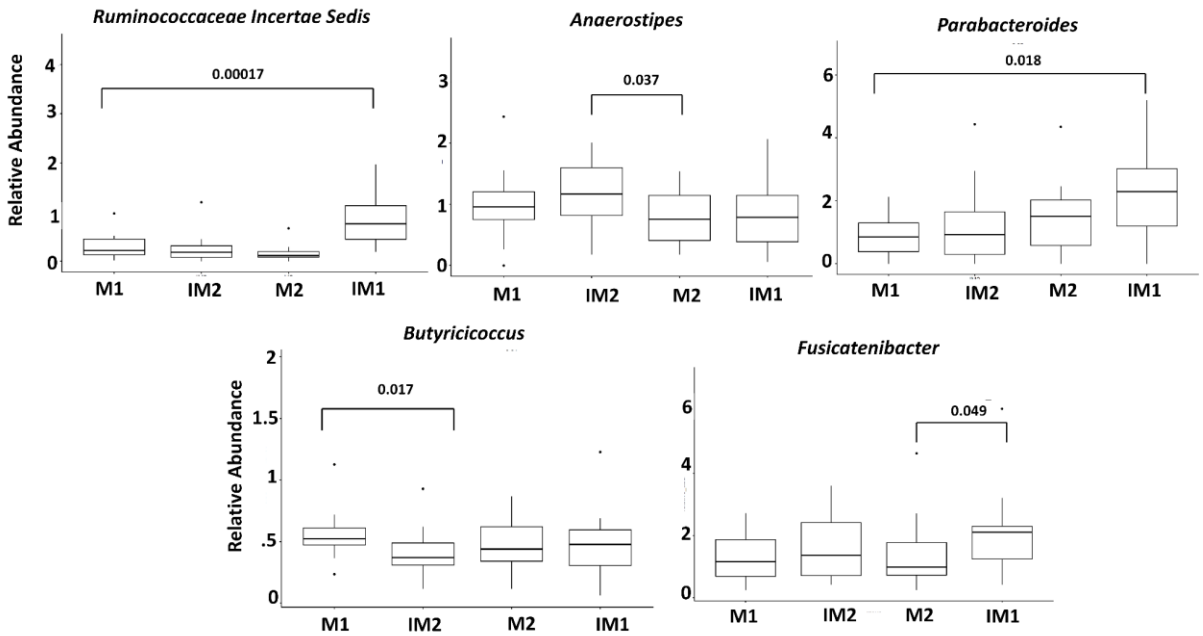


**Figure 4:** Boxplots representing the significant differences between mean proportions of genera between four meal groups. Significance was tested using non-parametric Wilcoxon Rank test. IM1 and IM2 represent the first and second interventional meals, respectively. M1 and M2 represent first and second non-intervention meals, respectively.

### 5.3. Correlation of gut microbial composition and clinical metadata

To study possible associations between features and continuous variables of interest such as fat-free mass [%], fat mass [%], skeletal muscle mass [kg], BMI and age, spearman correlation was calculated at the phyla level **(Figure 5 (a))**, showing that formerly known as Deltaproteobacteria) was negatively correlated with BMI followed by Verrucomicrobiota, Firmuicutes and Actinobacteriota. On the genus level **(Figure 5 (b))**, the genera *Phascolarctobacterium, Lachnospira, Lachnospiraceae FCS020 group, Prevotella, Alistipes and Oscillospiraceae UCG-005* were significantly but positively correlated with BMI, whereas *Bacteroides* and *Ruminococcus* and *Lachnoclostridium* were negatively correlated with BMI. *Prevotella, Lachnospiraceae FCS020 group* and *Phascolarctobacterium, w*hich were positively correlated with BMI, were also slightly less abundant in the IM group compared to the M groups. *Anaerostipes*, [*Eubacterium] ruminantium group* and *Lachnospira* were also positively associated with fat mass [%]. Conversely, *Rikenellaceae RC9 gut group and Clostridia UCG−014,* were negatively associated with fat mass percentage.

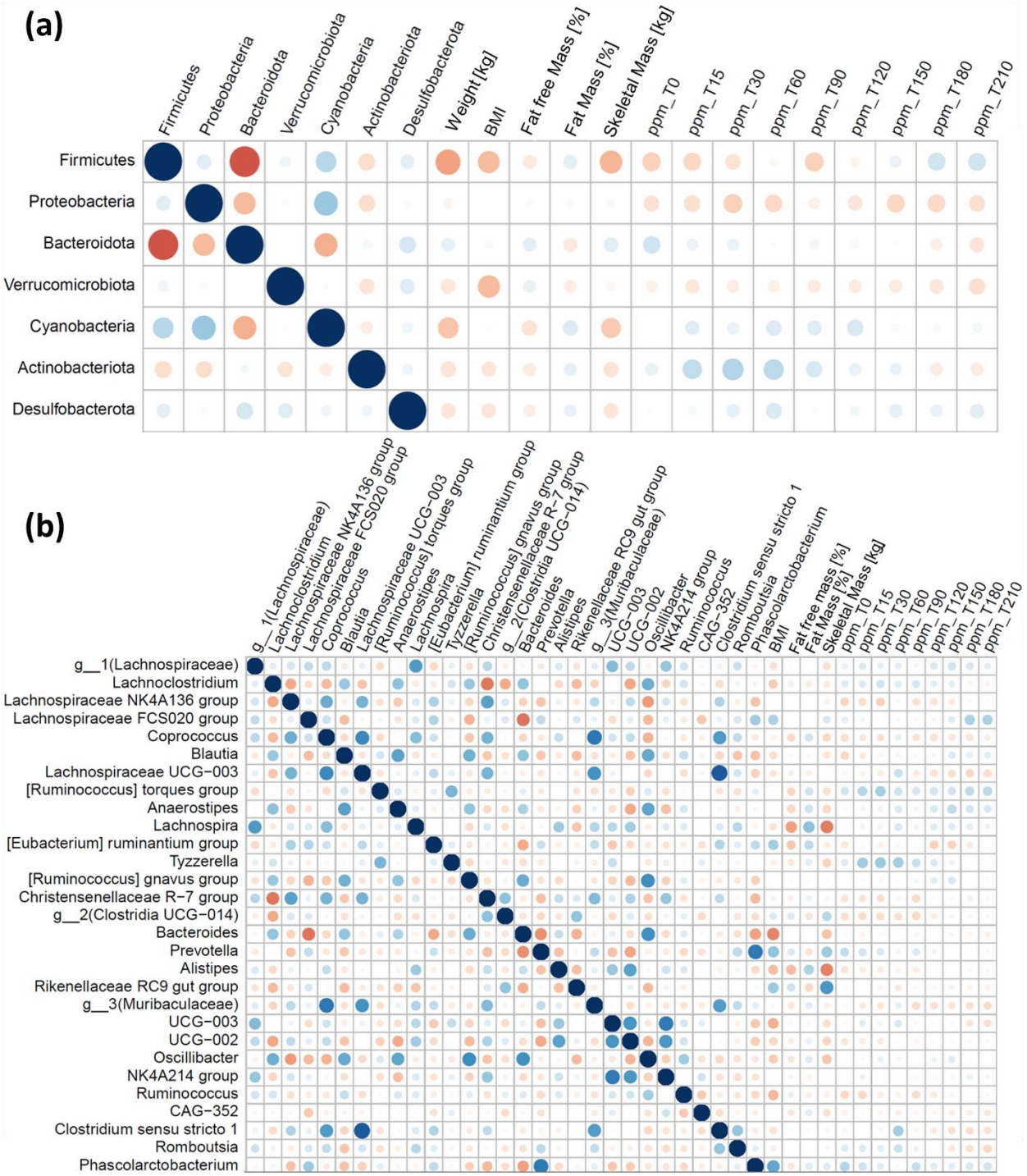


**Figure 5:** Correlation between gut microbial composition and clinical variables at phyla (a) and (b) genus level

### 5.4. Functional analysis

The built-in PICRUSt2 option of Namco was used to infer the functional differences of the microbial communities between the IM and M groups. Similar to differential abundance analysis at taxa level, significant differences were calculated with a paired Wilcoxon rank test between the IM and M groups. In total, 82 KEGG Orthologues (KO) were significantly different between the groups without correction for multiple testing. Significant KO terms with p-value < 0.05 were grouped according to KO categories in order to understand their functions. The majority of the 76 KO terms belonged to metabolic categories (level 1) and were further divided into 11 sub-categories (level 2) including carbohydrate metabolism, amino acid metabolism, energy metabolism (Oxidative phosphorylation), lipid metabolism, metabolism of cofactors and vitamins and glycan biosynthesis and metabolism. Among these, the KO term K00845 (glucokinase) is part of the amino sugar and nucleotide sugar metabolism which was enriched in the IM groups but not in the M groups. Previous studies suggested that high fibre intake shows a positive impact on glucose and fat metabolism in humans [75]. In support of that, ATP-binding cassette (ABC) transporters such as K02018, K10823, K15580 and K15583 were also upregulated in IM **(Figure 6).** Previous studies suggested that *Firmicutes*, which were slightly more abundant in the IM group, encodes ABC transporters which belong to transport ATPase groups on the bacterial plasma membrane. These transporters, which are essential to transfer glucose to the other side of the plasma membrane [76] also help in transporting anti-inflammatory butyrate from bacterial digestion of dietary fibres [77]. Differences in the abundance of ABC transporters and glucokinase are shown in Figure 7. Aspartate aminotransferase (AST) (K00812) was downregulated in the IM group compared to the M group. AST is an important biomarker for liver damage. Few studies showed that fibre-riched diets reduce the level of AST [78]. We also identified twelve pathways as significantly different between IM and M groups, namely fatty acid elongation saturated (FASYN-ELONG-PWY), superpathway of N-acetylglucosamine, N-acetylmannosamine and N-acetylneuraminate degradation (GLCMANNANAUT-PWY), lipid IVA biosynthesis(NAGLIPASYN-PWY), O-antigen building blocks biosynthesis (E. coli) (OANTIGEN-PWY), acetylene degradation (P161-PWY), polyisoprenoid biosynthesis (E. coli) (POLYISOPRENSYN-PWY), urate biosynthesis/inosine 5'-phosphate degradatio (PWY-5695), Kdo transfer to lipid IVA III (Chlamydia) (PWY-6467), guanosine ribonucleotides de novo biosynthesis(PWY-7221), superpathway of GDP-mannose-derived O-antigen building blocks biosynthesis (PWY-7323), superpathway of UDP-N-acetylglucosamine-derived O-antigen building blocks biosynthesis (PWY-7332) and superpathway of thiamin diphosphate biosynthesis I(THISYN-PWY) **(Figure 7).** We repeated this differential analysis using the ALDEx2 option provided in Namco. ALDEx2 applies CLR transformation to the raw counts to address potential biases introduced through compositionality. Following this approach, no significant KO terms were identified. We hypothesize that CLR transformation might increase the specificity in functional analysis at the cost of sensitivity, suggesting that users need to carefully reflect their method choice when interpreting their results.

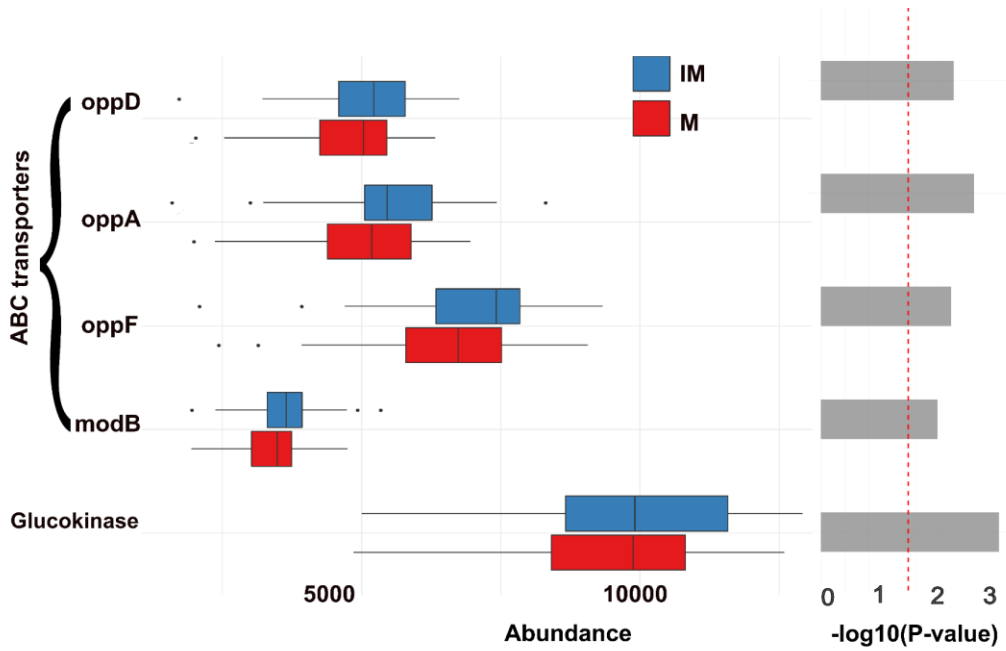


**Figure 6:** Bar plot showing the difference in the relative abundance of ABS transporters and glucokinase between IM and M group. p-values were calculated using the Wilcoxon Rank test on abundance values.

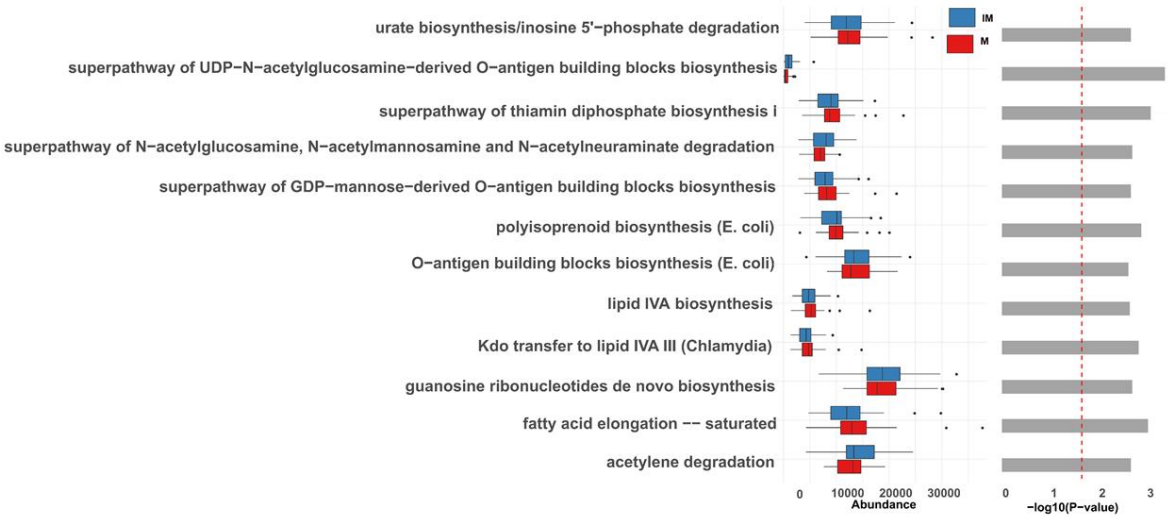


**Figure 7**: Bar plot showing pairwise-group comparison of significant pathways. Wilcoxon test was performed on relative abundance and extended error bar plots were used for the comparison between IM and M groups. Only predicted functions and pathways with $p < 0.05$ are shown. Bar plots on the left side display the mean proportion of each KO term with -log10 p-value while the right display the mean proportion of each KEGG pathway with -log10 p-values.

### 5.5. Network analysis

Co-occurrence networks were analyzed in Namco to characterize bacterial interactions between IM and M groups using the SPRING metric for network construction as the default option. To focus on the most abundant ASVs, an abundance cutoff of 0.25% and a prevalence cutoff of 10% were applied. The 270 remaining ASVs were used as input for network comparison. The co-occurrence network between IM and M at the genus level was illustrated in **Figure 8.** A genus level network was generated using The SPRING [79] method as association measure (nlambda and replication numbers were set to 50 and 100, respectively). Eigenvector centrality was used for defining hubs and for scaling node sizes. Comparisons of all global measures for degree and eigenvector centrality, respectively, are given in Table 2. For none of the four centrality measures, any significant differences are observed. The two networks shared similar properties and no node hubs were identified in both groups. The largest differences in closeness centrality between IM and M groups belonged to the genus *CAG-56*, *Eubacterium coprostanoligenes*, *Lachnospira Lachnospiraceae UCG-004*, and *Ruminococcaceae Incertae Sedis*. Among these genus, only *Ruminococcaceae Incertae Sedis* was found to be significantly abundant in the IM group from previous analysis. Overall, there is no significant difference observed at the genus level network between the IM and M groups.

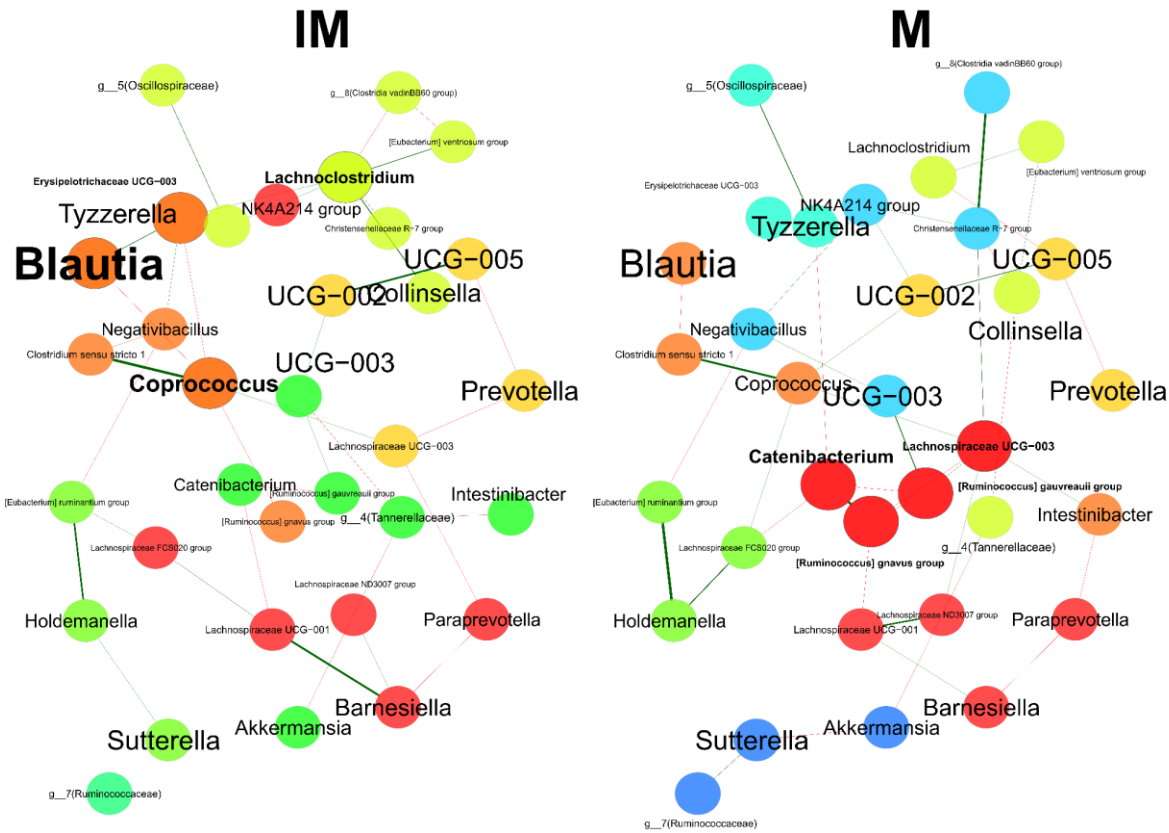


**Figure 8:** Bacterial associations on the genus level for the intervention (IM) and non-intervention (M) groups using SPRING method. Green edges represent positive associations and red edges represent negative associations. Node colors represent clusters determined by using greedy modularity optimization. Networks are shown with only the 50 nodes with the highest degree and 50 edges with highest weight.

## 6. Conclusion and Outlook

To gain biologically and clinically relevant insights from the large amounts of available microbiome sequencing data, a plethora of algorithms, statistical methods and software packages have been developed. We implemented Namco to serve as a one-stop data analysis platform that performs both raw data processing and basic as well as advanced downstream analyses of microbial datasets. Namco integrates previously available tools into a single coherent computational workflow and allows the user to construct, analyse and understand microbial composition in a fast and reproducible manner. This platform is intended to eliminate the use of command-line arguments during data processing. Namco is accessible in the web browser and hence does not require the installation of any software packages. Namco also allows saving results from each analysis as a R session which can be used to resume at any time, thus simplifying sharing of research results. Since Namco is available as a Docker image, it can be conveniently installed locally or on a clinical server behind a firewall to facilitate the GDPR-compliant analysis of sensitive data without the need for an upload to the public Namco instance. Detailed comparisons between other web-based tools (Table 1) showed that Namco offers a unique set of functions such as, for instance, time series clustering, function profiling using PICRUSt2, confounder analysis and topic modelling.

We considered dietary intervention study as a case study to explore the features of Namco. We examined the association of rich fibre dietary intake with the gut microbiota composition through basic and advanced analysis in Namco. We investigated differences in relative abundance between IM and M groups, where we compared the top abundant taxa and also studied the intra-individual variation in gut microbiome with respect to fibre-enriched diets. After exploring the datasets in terms of relative abundance and diversity analysis, we found that genera with significant differences between IM and M groups were involved in the producing butyrate which is a SCFA that helps to maintain the homeostasis of gut via anti-inflammatory and antimicrobial actions [80–82]. Namco did not only provide information about differential abundant microbial composition, but also helped in determining the significantly different KO terms and pathways. Namco revealed that the IM group showed a positive association with the presence of the glucokinase which belongs to amino sugar and nucleotide sugar metabolism. High fibre intake showed a positive impact on glucose metabolism in humans. Studies showed that long term intake of fibre improves the glucose homeostasis [83]. In addition to that, Namco also identified ABC transporters which play a major role in transmission of glucose through plasma membranes as significantly correlated with the IM group [76]. It was also possible to study microbial interactions by generating differential microbial co-occurrence networks on the genus level using Namco. The topological features of the resulting differential network from IM and M groups showed only a slight difference in the estimated associations. Overall, Namco provided a much-needed interface to analyze the microbial community data in a more intuitive way.

In summary, we present Namco, a shiny R application dedicated to provide end to end microbiome analysis for 16S rRNA gene analysis. We incorporate leading R packages for both upstream and downstream analysis in an efficient framework for researchers to characterize and understand the

microbial community structure in their data, leading to valuable insights into the connection between the microbial community and phenotypes of interest. In the future, we plan to further expand Namco with support for novel analysis techniques and for correlation of microbial abundances with other data sources such as metabolomics and transcriptomics.

## 7. Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German 582 Research Foundation) – Projektnummer 395357507 – SFB 1371. We also wanted to thank Luise Rauer and Claudia Hülpüsch from IEM Augsburg for constructive feedback during the development of Namco. This project (JB) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777111. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains. In addition, JB was partially funded by his VILLUM Young Investigator Grant nr.13154. Parts of the study received funding by a grant of the German Ministry for Education and Research (BMBF, 01EA1409C). The preparation of this paper was supported by the enable Cluster.

## 8. Conflict of interest

MLI receives consulting fees from mbiomics GmbH outside this work.

## 9. Availability of data and materials

Tool home page: https://exbio.wzw.tum.de/namco/
Browser: Google Chrome, Microsoft Internet Explorer, Mozilla Firefox.
Source code: https://github.com/biomedbigdata/namco
license: GPLv3

## 10. Abbreviations

**ABC:** ATP-Binding Cassette
**AST:** Aspartate Aminotransferase
**ASV:** Amplicon Sequence Variants
**BMI:** Body Mass Index
**GDPR:** General Data Protection Regulation
**KO:** KEGG Orthology
**IM:** Intervention Meals
**M:** non-intervention Meals
**MDS:** Multi-Dimensional Scaling
**NMDS:** Non-metric Multidimensional Scaling
**OTU:** Operational Taxonomic Units
**SCFA:** Short Chain Fatty Acid

## 11. Authors' contributions

**BB, TS and HH** designed and performed the experiments shown in the use case. **AD, MZ, MLA and BÖ** implemented the first version of Namco which was extended and improved by **AD, MLI** and **MM**. **MM** analyzed the data, wrote the first draft of the paper, prepared figures and/or tables. **MLI** and **SR** were involved in conceptualization and supervision. All authors were involved in manuscript review and editing and approved the final manuscript.

## 12. References


1. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet. 2012;13:260–70.

2. Devaraj S, Hemarajata P, Versalovic J. The human gut microbiome and body metabolism: implications for obesity and diabetes. Clin Chem. 2013;59:617–28.

3. Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA, Knight R. The microbiome and human cancer. Science. 2021;371. doi:10.1126/science.abc4552.

4. Glassner KL, Abraham BP, Quigley EMM. The microbiome and inflammatory bowel disease. J Allergy Clin Immunol. 2020;145:16–27.

5. Morais LH, Schreiber HL 4th, Mazmanian SK. The gut microbiota-brain axis in behaviour and brain disorders. Nat Rev Microbiol. 2021;19:241–55.

6. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.

7. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–43.

8. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

9. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 2015;9:968–79.

10. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41.

11. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37:852–7.

12. Bokulich NA, Dillon MR, Bolyen E, Kaehler BD, Huttley GA, Caporaso JG. q2-sample-classifier: machine-learning tools for microbiome classification and regression. J Open Res Softw. 2018;3. doi:10.21105/joss.00934.

13. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, et al. q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. mSystems. 2018;3. doi:10.1128/mSystems.00219-18.

14. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. Microb Ecol Health Dis. 2015;26:27663.

15. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217.

16. McMurdie PJ, Holmes S. Shiny-phyloseq: Web application for interactive microbiome analysis with provenance tracking. Bioinformatics. 2015;31:282–3.

17. Woloszynek S, Mell JC, Zhao Z, Simpson G, O'Connor MP, Rosen GL. Themetagenomics: Exploring Thematic Structure and Predicted Functionality of 16s rRNA Amplicon Data. bioRxiv. 2019;:678110. doi:10.1101/678110.

18. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res. 2017;45:W180–8.

19. Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, et al. IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci Rep. 2016;6:33721.

20. Buza TM, Tonui T, Stomeo F, Tiambo C, Katani R, Schilling M, et al. iMAP: an integrated bioinformatics and visualization pipeline for microbiome data analysis. BMC Bioinformatics. 2019;20:374.

21. Wilke A, Bischof J, Gerlach W, Glass E, Harrison T, Keegan KP, et al. The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Res. 2016;44:D590–4.

22. Su S-C, Galvin JE, Yang S-F, Chung W-H, Chang L-C. wiSDOM: a visual and statistical analytics for interrogating microbiome. Bioinformatics. 2021. doi:10.1093/bioinformatics/btab057.

23. Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren AM, et al. VAMPS: a website for visualization and analysis of microbial population structures. BMC Bioinformatics. 2014;15:41.

24. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics. 2015;31:2882–4.

25. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013;31:814–21.

26. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. 2020;38:685–8.

27. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal

RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41 Database issue:D590–6.

28. Cao Q, Sun X, Rajesh K, Chalasani N, Gelow K, Katz B, et al. Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. Front Microbiol. 2020;11:607325.

29. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. Methods for normalizing microbiome data: An ecological perspective. Methods Ecol Evol. 2019;10:389–400.

30. Reitmeier S, Hitch TCA, Treichel N, Fikas N, Hausmann B, Ramer-Tait AE, et al. Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. ISME Communications. 2021;1:1–12.

31. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017;8:2224.

32. Lagkouvardos I, Fischer S, Kumar N, Clavel T. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. PeerJ. 2017;5:e2836.

33. Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;27:379–423.

34. Simpson EH. Measurement of Diversity. Nature. 1949;163:688–688.

35. Chao A, Chiu C-H, Jost L. Phylogenetic diversity measures based on Hill numbers. Philos Trans R Soc Lond B Biol Sci. 2010;365:3599–609.

36. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71:8228–35.

37. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. Ecol Monogr. 1957;27:325–49.

38. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, et al. The vegan package. Community ecology package. 2007;10:719.

39. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.

40. Wirbel J, Zych K, Essex M, Karcher N, Kartal E, Salazar G, et al. Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. Genome Biol. 2021;22:93.

41. Hawinkel S, Mattiello F, Bijnens L, Thas O. A broken promise: microbiome differential abundance methods do not control the false discovery rate. Brief Bioinform. 2019;20:210–21.

42. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25:679–89.

43. Blei DM, Lafferty JD. A correlated topic model of Science. aoas. 2007;1:17–35.

44. Woloszynek S, Mell JC, Zhao Z, Simpson G, O'Connor MP, Rosen GL. Exploring thematic structure and predicted functionality of 16S rRNA amplicon data. PLoS One. 2019;14:e0219235.

45. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One. 2013;8:e67019.

46. Nearing JT, Douglas GM, Hayes M, MacDonald J, Desai D, Allward N, et al. Microbiome differential abundance methods produce disturbingly different results across 38 datasets. bioRxiv. 2021;:2021.05.10.443486. doi:10.1101/2021.05.10.443486.

47. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science. 2012;336:1268–73.

48. Matchado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, et al. Network analysis methods for studying microbial communities: A mini review. Comput Struct Biotechnol J. 2021;19:2687–98.

49. Peschel S, Müller CL, von Mutius E, Boulesteix A-L, Depner M. NetCoMi: network construction and comparison for microbiome data in R. Brief Bioinform. 2020. doi:10.1093/bib/bbaa290.

50. Clayden A. Causal relationships in medicine: A practical system for critical appraisal. Ann Intern Med. 1991;114:916.

51. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020;587:448–54.

52. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, Articles. 2017;77:1–17.

53. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, et al. Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat Med. 2021;27:321–32.

54. Polster SP, Sharma A, Tanes C, Tang AT, Mericko P, Cao Y, et al. Permissive microbiome characterizes human subjects with a neurovascular disease cavernous angioma. Nat Commun. 2020;11:2659.

55. Brandl B, Skurk T, Rennekamp R, Hannink A, Kiesswetter E, Freiherr J, et al. A Phenotyping Platform to Characterize Healthy Individuals Across Four Stages of Life - The Enable Study. Front Nutr. 2020;7:582387.

56. Rennekamp R, Brandl B, Giesbertz P, Skurk T, Hauner H. Metabolic and satiating effects and consumer acceptance of a fibre-enriched Leberkas meal: a randomized cross-over trial. Eur J Nutr. 2021;60:3203–10.

57. Reitmeier S, Kiessling S, Clavel T, List M, Almeida EL, Ghosh TS, et al. Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. Cell Host Microbe. 2020;28:258–72.e6.

58. McRae MP. Dietary Fiber Is Beneficial for the Prevention of Cardiovascular Disease: An Umbrella Review of Meta-analyses. J Chiropr Med. 2017;16:289–99.

59. Anderson JW, Baird P, Davis RH Jr, Ferreri S, Knudtson M, Koraym A, et al. Health benefits of dietary fiber. Nutr Rev. 2009;67:188–205.

60. Aune D, Chan DSM, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ. 2011;343:d6617.

61. Xu X, Zhu Y, Li J, Wang S. Dietary fiber, glycemic index, glycemic load and renal cell carcinoma risk. Carcinogenesis. 2019;40:441–7.

62. da Cruz AG, Senaka Ranadheera C, Nazzaro F, Mortazavian A. Probiotics and Prebiotics in Foods: Challenges, Innovations, and Advances. Academic Press; 2021.

63. Bailén M, Bressa C, Martínez-López S, González-Soltero R, Montalvo Lominchar MG, San Juan C, et al. Microbiota features associated with a high-fat/low-fiber diet in healthy adults. Front Nutr. 2020;7:583608.

64. Aoe S, Nakamura F, Fujiwara S. Effect of Wheat Bran on Fecal Butyrate-Producing Bacteria and Wheat Bran Combined with Barley on Bacteroides Abundance in Japanese Healthy Adults. Nutrients. 2018;10. doi:10.3390/nu10121980.

65. Van den Abbeele P, Belzer C, Goossens M, Kleerebezem M, De Vos WM, Thas O, et al. Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. ISME J. 2013;7:949–61.

66. Menni C, Jackson MA, Pallister T, Steves CJ, Spector TD, Valdes AM. Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. Int J Obes . 2017;41:1099–105.

67. Higgins JA, Higbee DR, Donahoo WT, Brown IL, Bell ML, Bessesen DH. Resistant starch consumption promotes lipid oxidation. Nutr Metab . 2004;1:8.

68. Kverka M, Zakostelska Z, Klimesova K, Sokol D, Hudcovic T, Hrncir T, et al. Oral administration of Parabacteroides distasonis antigens attenuates experimental murine colitis through modulation of immunity and microbiota composition. Clin Exp Immunol. 2011;163:250–9.

69. Wang K, Liao M, Zhou N, Bao L, Ma K, Zheng Z, et al. Parabacteroides distasonis Alleviates Obesity and Metabolic Dysfunctions via Production of Succinate and Secondary Bile Acids. Cell Rep. 2019;26:222–35.e5.

70. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012;490:55–60.

71. Cuffaro B, Assohoun ALW, Boutillier D, Súkeníková L, Desramaut J, Boudebbouze S, et al. In Vitro Characterization of Gut Microbiota-Derived Commensal Strains: Selection of Parabacteroides distasonis Strains Alleviating TNBS-Induced Colitis in Mice. Cells. 2020;9. doi:10.3390/cells9092104.

72. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. 2016;352:560–4.

73. Ezeji JC, Sarikonda DK, Hopperton A, Erkkila HL, Cohen DE, Martinez SP, et al. Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. Gut Microbes. 2021;13:1922241.

74. Tian T, Zhang X, Luo T, Wang D, Sun Y, Dai J. Effects of Short-Term Dietary Fiber Intervention on Gut Microbiota in Young Healthy People. Diabetes Metab Syndr Obes. 2021;14:3507–16.

75. Shortt C, Hasselwander O, Meynier A, Nauta A, Fernández EN, Putz P, et al. Systematic review of the effects of the intestinal microbiota on selected nutrients and non-nutrients. Eur J Nutr. 2018;57:25–49.

76. Sun B, Hou L, Yang Y. Effects of altered dietary fiber on the gut Microbiota, short-chain fatty acids and cecum of chickens during different growth periods. Preprints. 2020. doi:10.20944/preprints202002.0109.v1.

77. Andersen V, Svenningsen K, Knudsen LA, Hansen AK, Holmskov U, Stensballe A, et al. Novel understanding of ABC transporters ABCB1/MDR/P-glycoprotein, ABCC2/MRP2, and ABCG2/BCRP in colorectal pathophysiology. World J Gastroenterol. 2015;21:11862–76.

78. Abazarfard Z, Eslamian G, Salehi M, Keshavarzi S. A Randomized Controlled Trial of the Effects of an Almond-enriched, Hypocaloric Diet on Liver Function Tests in Overweight/Obese Women. Iran Red Crescent Med J. 2016;18:e23628.

79. Yoon G, Gaynanova I, Müller CL. Microbial Networks in SPRING - Semi-parametric Rank-Based Correlation and Partial Correlation Estimation for Quantitative Microbiome Data. Front Genet. 2019;10:516.

80. Corrêa-Oliveira R, Fachi JL, Vieira A, Sato FT, Vinolo MAR. Regulation of immune cell function by short-chain fatty acids. Clin Transl Immunology. 2016;5:e73.

81. Donohoe DR, Garge N, Zhang X, Sun W, O'Connell TM, Bunger MK, et al. The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell Metab. 2011;13:517–26.

82. Wang G. Human antimicrobial peptides and proteins. Pharmaceuticals . 2014;7:545–94.

83. Fukagawa NK, Anderson JW, Hageman G, Young VR, Minaker KL. High-carbohydrate, high-fiber diets increase peripheral insulin sensitivity in healthy young and old adults. Am J Clin Nutr. 1990;52:524–8.