Analysing microbiome intervention studies: Does choice of a statistical method affect biological interpretation?

Maryia Khomich (✉ marykhomich@gmail.com)

Nofima AS <https://orcid.org/0000-0002-6840-5739>

Ingrid Måge

Nofima AS

Ingunn Berget

Nofima AS

Ida Rud

Nofima AS

Research

Keywords: gut microbiome, dietary intervention trials, compositional data, centred log-ratio transformation, differential abundance, multivariate, ANOVA

Posted Date: May 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-30540/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

The diet plays a major role in shaping gut microbiome composition and function in both humans and animals, and dietary intervention trials are often used to investigate and understand these effects. A plethora of statistical methods for analysing differential abundance of microbial taxa exists, and new methods are constantly being developed, but there is a lack of benchmarking studies and clear consensus on the best multivariate statistical practices. This makes it hard for a biologist to decide which method to use.

Results

We compared the outcomes of a wide range of ANOVA-like statistical methods to explore to what extent the choice of method affects the biological inferences made. The comparison is based on five published dietary intervention studies representing different subjects and study designs. We found that the methods producing outputs at the community level were in agreement regarding both effect size and statistical significance. However, the methods that provided ranking of operational taxonomic units (OTUs) gave incongruent results, implying that the choice of method is likely to influence the biological interpretations.

Conclusions

We concluded that inferences at the OTU level should be made with caution, and that a combination of several statistical methods allows to interpret the outcomes with higher confidence and simultaneously account for the limitations of each method.

Background

The microbiome has emerged as an important link to health and disease [1]. Microbiome analysis methods are rapidly advancing, in particular in areas such as compositional data analysis, multi-omics, and data integration [2, 3]. A clear understanding of the type of data being analysed is crucial, given the growing number of studies uncovering the key role of microbiome, its composition and functions following diet intervention or medical treatment [4]. At present, analysis of complex microbial data benefits from adapting the multivariate statistical toolbox from ecology and environmental sciences, and a proper choice of statistical tools is becoming increasingly important [5–7]. However, a lack of benchmarking studies and clear consensus on the best multivariate statistical practices make comparisons across microbiome data sets difficult [3, 8].

Different statistical methods have different properties, and the choice of the method should depend on the scientific question, experimental design, data characteristics, and expected relationships among the

variables. In reality, the choice is often biased by the research groups' tradition and familiarity with specific "toolboxes". Main differences between existing statistical approaches for analysing microbiome data are related to: (1) explorative versus confirmative; (2) univariate versus multivariate; (3) parametric versus nonparametric; (4) linear versus nonlinear; (5) compositional versus non-compositional; (6) distance-based versus count/abundance-based; and (7) incorporating phylogenetic information into the analysis or not [9]. Here, we focus on statistical methods for analysing data from dietary intervention studies. In contrast to observational studies, these are usually small in sample size but performed in (semi)-controlled environments and tailored to a specific research hypothesis. The studies often include multiple experimental factors, possibly with more than two levels, and it is therefore natural to turn to ANOVA-like methods. From a biologist's point of view, it is important that the methods are easy to interpret, both on the multivariate (microbial community) and univariate (microbial taxa or operational taxonomic units, OTUs) levels.

Distance-based methods

The distance-based methods are multivariate by definition, since multiple variables (microbial OTUs) are used to calculate pairwise distances between samples. Among distance-based methods, permutational multivariate analysis of variance (PERMANOVA) is the most widely used and more powerful than analysis of similarities (ANOSIM) to detect changes in community structure [10–12]. Both methods may be implemented with any dissimilarity metric. Among abundance-based beta diversity indices, Bray-Curtis is the most common choice due to its theoretical properties and empirical accuracy [13, 14]. The most widely applied phylogenetic beta diversity indices is UniFrac-type metrics [14–16]. However, UniFrac is unsuitable as a distance metric for studies with small sample size, which is usually the case for dietary intervention trials [17, 18]. Both PERMANOVA and ANOSIM test differences at the community level but do not provide any information at the OTU level. Similarity percentage analysis (SIMPER) is a method that can be applied for *post hoc* testing to evaluate the relative contribution of each analysed microbial taxon (i.e. OTU) to the overall average Bray-Curtis dissimilarities by pairwise comparison of two or more groups [12]. To the best of our knowledge, no such method exists for the other distance metrics.

Distance-based methods have their strengths and weaknesses that are important to account for beforehand. ANOSIM cannot deal with multifactorial designs, and both ANOSIM and PERMANOVA have problems detecting differences unless they are present in taxa with high variability [19]. Newer methods aimed at assigning more interpretable effect sizes are under constant development [6, 20]. For example, the more flexible PERMANOVA-S is an extension to existing distance-based methods that can adjust for covariates and simultaneously incorporate multiple distance metrics [21]. However, these methods do not consider covariance/correlation between microbial species, and they encounter significant power loss if *all* microbial species are used for distance calculations [22].

Abundance-based methods

The abundance-based methods can be either univariate (analysing each OTU individually) or multivariate (focusing on covariance structure between OTUs). To account for compositionality in microbiome data,

compositional data analysis (CoDA), based on transformation into a log-ratio coordinate space, is often used [23, 24]. In this study, the centred log-ratio (clr) transformation is applied [24, 25]. After transformation, any suitable univariate or multivariate method can be used. Univariate analysis of variance (ANOVA) is a well-established family of methods that can incorporate complex data structures such as random effects, longitudinal or repeated measurements, nested designs etc. Non-parametric methods (such as Wilcoxon signed-rank test) are popular for assessing differences between groups when sample size is small and normality assumptions are not met. Popular methods tailored for high-throughput sequencing microbiome data are ANOVA-like differential expression analysis (ALDEx2) and the analysis of composition of microbiomes (ANCOM) [23, 26]. Both approaches use the log-ratio transformation of microbiome data prior to univariate assessment of statistical significance and differ in how zero values are handled [4, 27]. ALDEx2 uses a Dirichlet-multinomial probability distribution to estimate abundances from counts and calculates the false discovery rate (FDR) based on Monte Carlo simulations. In ANCOM, on the other hand, the compositional nature of the data is considered by testing the log-ratio for all pairs of OTUs, and then counting the number of tests where the log-ratio is significantly different from zero. This number (W-stat) can be used to obtain a ranking of OTUs most likely to differ between the groups.

The drawback of univariate methods is that they treat all species as independent variables, not taking the covariance between OTUs into account. Such methods may fail to detect community-level differences [28]. The classical generalisation of ANOVA to multiple variables (MANOVA) cannot be used when the number of variables exceeds the number of samples, as it suffers from the problem of singularity of covariance matrices and assumptions that are not fulfilled [29, 30]. Novel statistical ANOVA-like methods include fifty-fifty multivariate analysis of variance (FFMANOVA) [31] and ANOVA-simultaneous component analysis (ASCA) [30]. Both methods are based on principal component analysis (PCA), they can handle multiple collinear responses and can be used on any data types. ASCA has gained popularity in metabolomics research [32–34]. The output of ASCA are scores and loadings related to the experimental factors, which can be visualised in the same way as for PCA to better understand covariance patterns within the data. In addition, the hypothesis testing can be performed at the multivariate level through permutation tests. Both ASCA and FFMANOVA methods have successfully been applied on microbiome data [35–38]. Linear discriminant analysis effect size (LEfSe) is a stepwise approach that combines univariate analysis with multivariate discriminant analysis [39], but it is not well adapted to experimental designs with several multilevel factors. The method has found a wide use in microbiome research due to its easy-to-use and interpret visualisation [40, 41].

An overview of the different methods compared in this study is given in Fig. 1 and Table 1. As can be seen from Fig. 1, PERMANOVA and ANOSIM provide results only at the community level, while SIMPER, ANCOM and ALDEx2 report results for single OTUs only. ASCA and FFMANOVA, on the other hand, give results both at the community and OTU levels. For FFMANOVA, this is done by rotation test procedure which adjusts the p-values for multiple testing [42]. For ASCA, OTU ranking is based on ASCA loadings for each experimental factor or by partial least squares discriminant analysis (PLS-DA) for pairwise comparisons (for details see Methods section).

Method comparison

The aim of the method comparison was to investigate if the choice of a statistical method affects biological inference. At the community level, methods were compared with respect to effect sizes (expressed as percentage of explained variance) and corresponding p-values. At the OTU level, the ranking of OTUs was compared by Spearman's rank correlation and by investigation of scatterplots between the different ranking metrics.

We focused on dietary intervention trials since the link between health balance and disease is of major interest to the microbiome research community, precision medicine and food science. We used five recently published data sets as basis for the comparisons (Table S1). The following criteria for studies to be included were considered: (1) at least two-factorial experimental design with a minimum of two factor levels; (2) either human or animal gut microbiome surveys; and (3) a taxonomic assignment at the OTU level reported. Diet is the main factor of interest in all five studies, and we restricted our comparisons to this factor.

Results

Community level

Four of the methods can be used to test the association between diet and the overall microbiome composition, namely distance-based ANOSIM and PERMANOVA and compositional ASCA and FFMANOVA. The results across studies are summarised in Table 2.

Data set 1. The effect of dietary fibres inulin (IN), cellulose (CE) or brewers spent grain (BSG) on the overall caecal microbiota composition in mice from study by Moen et al. [38] accounted for 34–37% of the explained variation according to the FFMANOVA, ASCA and PERMANOVA. In general, the three methods produced similar results, with slightly smaller p-values by FFMANOVA and ASCA.

Data set 2. Lai et al. [43] investigated the effect of diet (the main variable), exercise and their interaction on the overall faecal microbiota in sedentary and exercised mice fed high fat or normal fat diet (four groups in total). Similarly, all tested methods, except ANOSIM, have produced congruent results (effect of diet 27–31%), with smaller p-values by FFMANOVA and ASCA.

Data set 3. In a longitudinal study by Le Sciellour et al. [44] the authors tested the effect of dietary fibre content on faecal microbiota in growing-finishing pigs fed alternately a low-fibre and a high-fibre diet during four successive 3-week periods. In this survey, the effect of diet was small (2%) compared to the effect of diet in data sets 1 and 2. Similarly, FFMANOVA and ASCA have produced slightly lower p-values than PERMANOVA, but all three methods were in agreement on the effect of diet.

Data set 4. In a longitudinal study by Wang et al. [45] the objectives were to determine the impact of beta glucan on the composition of faecal microbiota in mildly hypercholesterolemic individuals. The

individuals received for 5-week either a treatment containing 3 g high molecular weight (HMW), 3 g low molecular weight (LMW), 5 g LMW barley beta glucan or wheat and rice (control group) [45]. The effect of diet accounted for ~ 2% of the explained variation reported by FFMANOVA, ASCA and PERMANOVA. Diet was significant on a 5% level for PERMANOVA and ASCA ($p = 0.036$ and $p = 0.038$, respectively) and on a 10% level for FFMANOVA ($p = 0.067$). However, different conclusions were drawn with respect to time. Significant results at the community level were obtained only for FFMANOVA ($p = 0.007$).

Data set 5. Birkeland et al. [46] assessed the effect of prebiotic fibres or a control supplement on faecal microbiota composition in human subjects with type two diabetes. The interaction effect of treatment and day accounted for 1–2% of the explained variation according to the FFMANOVA, ASCA and PERMANOVA. Significant results at the community level were obtained only for FFMANOVA ($p < 0.001$).

To summarise, PERMANOVA, FFMANOVA and ASCA gave almost identical results regarding effect sizes and statistical significance across studies (Table 2). ANOSIM provided the most different results, likely due to less resolution in the analysis since the multifactorial nature of the studies cannot be taken into consideration by this approach. Secondly, there is a considerable difference in effect sizes of the main factor of interest (diet) between animal (2–37%) and human (1–2%) dietary intervention trials, with effect sizes being very small in human studies. Lastly, three of the studies had crossover designs, allowing for estimation of interindividual variation. This variation was considerably higher for trials involving human subjects (54–74%) compared to the animal study (26%).

OTU level

Six of the methods, namely SIMPER, ASCA, FFMANOVA, LefSe, ANCOM and ALDEx2, can be used to make biological inference for individual OTUs. The methods give different outputs which can be used to identify differentially abundant OTUs and/or rank the OTUs according to effect sizes (see Methods for details). The summary tables and scatterplots are given as Appendix 1 and Figure S1, respectively.

Agreement between the methods was investigated by calculating Spearman's rank correlation between all pairs of output metrics (Fig. 2). The congruence between the methods was higher in animal studies (Moen, Lai and Le Sciellour) than in human studies (Wang and Birkeland), as expected. In general, animal studies are more controlled, and interindividual variation is smaller compared to studies involving human subjects. Interestingly, ASCA and FFMANOVA are the two methods that had highest agreement across all five data sets, while results obtained by pairwise comparisons of ALDEx2 versus the other methods were generally less correlated (Fig. 2). LefSe was evaluated only for the data sets Moen and Lai, since very few OTUs were flagged as significantly different between the conditions for the remaining three data sets. A high congruence between ASCA and ANCOM results was observed only for animal studies (data sets Moen, Lai and Le Sciellour) compared to human studies (Wang and Birkeland). For easier visualisation, SIMPER was excluded as it showed low agreement with the other tested methods.

Gut microbiome data are typically represented by a few dominant OTUs (relative abundance > 1%) and a majority of low-abundant OTUs (< 0.05%). One is therefore interested in ranking OTUs independent of

their average abundance, to be able to detect biologically relevant changes in low-abundant OTUs. Our results show that ranking of OTUs is highly dependent on abundance for SIMPER and ANCOM (Fig. 3 and Figure S2). Notably, ranking based on SIMPER was completely determined by abundance, since dominant OTUs had larger influence on the Bray-Curtis distance, as had previously been reported [19]. SIMPER is therefore not suited for identifying changes in low-abundant OTUs. For ANCOM, the results indicated that highly abundant OTUs had either very high or very low W-stat, while low-abundant OTUs always had medium-to-low W-stat. To the best of our knowledge, this finding has not been reported before and shows that ANCOM, like SIMPER, is not able to identify changes in low-abundant OTUs. Tentatively, this trend could be due to different distributions of log-ratios for highly and low-abundant OTUs, but it is beyond the scope of the present study to investigate this further.

Discussion

Diet is often considered to be a driver of microbiome variation [47]. In observational population-based studies, diet consistently accounts for only a small proportion of microbiome variation, and this is partly due to large interindividual differences, small sample sizes and limitations in study designs such as potentially insufficient washout periods in crossover studies [47–51]. In general, higher interindividual variation is observed in gut microbiota of human subjects compared to animal species [52]. Nevertheless, use of animal models to study the causal role of gut microbiota in health and disease is an established practice although animal models lack the specific interactions present in the complex system of human organism [53].

Univariate and multivariate analyses provide information at different levels. Biologists will often find that the outputs of univariate analyses are easier to interpret compared to those generated by multivariate analyses, though assumptions are similar for both method types [9]. In general, multivariate methods provide a more holistic overview of differences between samples and account for correlations and interactions between the variables, whereas univariate methods are well suited to point out the differences for specific microbial groups. The two levels therefore provide complementary information, and we think that it is generally of biological interest to report differences at both levels.

Multivariate ANOVA methods, namely FFMANOVA and ASCA, consider the underlying covariance between multiple variables in comparison to traditional ANOVA. In complex microbial data sets, where the variables (i.e. OTUs) are not independent, FFMANOVA and ASCA can serve as suitable tools to make biological inference at both the community and OTU levels. Our results show that these methods performed similarly on the example data (similar community-level effect sizes and similar ranking of OTUs; see Table 2 and Fig. 2). Distance-based PERMANOVA provided similar results at the community level, but the results should be complemented by other methods to gain insight into OTUs that are affected.

ASCA, FFMANOVA and LEfSe (through the LDA step) are based on the covariance structure between OTUs, and therefore depend on the relative scaling of the OTUs. It is usual practice in many areas to scale

all variables to equal variance, giving them equal weight in the model. Another option, which retains more of the original variability, is pareto scaling. An overview of different scaling options used in metabolomics, also relevant for microbiota, can be found in [54]. To prevent that the results are dominated by a few highly-abundant OTUs, data should be scaled when analysing abundances. Cln-transformation puts the variables on comparable levels, and the need for scaling is less obvious. However, the highly-abundant OTUs might still have slightly higher variance, and scaling should be considered depending on the data characteristics.

The tools tested in the present study vary in how flexible they are regarding possible comparisons, namely two-group comparisons (SIMPER, ASCA and LEfSe), possibility to adjust for confounder (ANCOM, FFMANOVA and ASCA) or to specify complex models with interactions (ALDEx2, FFMANOVA and ASCA). These aspects should be considered when selecting methods, as different study designs might require different types of statistical models and tests.

Although the Spearman's rank correlation indicated good agreement for the animal studies, little overlap between lists of "significant" results could be detected. There can be several reasons for this. One important aspect is that different criteria must be used to define "significance" or generally "importance". FFMANOVA and ALDEx2 provide p-values, whereas more heuristic tools must be applied with ASCA, SIMPER and ANCOM. Also note that for some methods (FFMANOVA, ALDEx2 and ANCOM), the ranking is related to *all* levels of the experimental factors, whereas the other methods use pairwise comparisons only. Moreover, differences in sample collection, sample preparation and sequencing contribute to additional variability, which, in turn, affects the validity of the results [55]. Hence, comparisons across different studies with similar interventions would be even more difficult.

Newly introduced statistical tools can produce contrasting results compared to the existing methods, but such comparisons tend to be biased in favour of the new approach [56]. Past benchmarking studies [57–59] have widely reported varying results from different tools, which was also confirmed in our study. Currently, there is no consensus for the best existing tool for detecting differentially abundant microbial taxa, and there is no reason to believe that one single method is best in all cases. It is therefore good scientific practice to compare and report outputs from several methods. The strategy for making inference from multiple outputs depends on whether it is more important to reduce the number of false negatives or false positives. If it is important not to miss out any possible findings (typically in early stages of research), any OTU listed as significant/important by any method should be investigated further. To obtain robust results, on the other hand, OTUs should be reported as differentially abundant only if they were selected as "significant" by several methods [56].

Finally, microbial data at the OTU level are zero-inflated, and rare OTUs should be removed prior to downstream statistical analyses. We have observed that the threshold for filtering out OTUs can significantly affect the results, both at the community and OTU levels, which is a topic for further research.

Conclusions

In the present study, we compared the performance of several multivariate ANOVA-like statistical methods taking several dietary intervention microbiome data sets as examples. At the community level, the different methods came to similar conclusions. However, at the OTU level the agreement between methods varied considerably. Some of the methods provided output metrics that were dependent on average abundance, making it impossible to detect differences in low-abundant OTUs. This suggests that choice of statistical method might affect biological inference at the OTU level. Ranking of OTUs obtained with different methods correlated better for animal studies than for human studies, possibly due to lower interindividual variation in animal studies. Based on our results, we suggest that microbiome surveys should report effect sizes at the community level before making any inferences at lower taxonomic levels. We also advise to combine several multivariate statistical methods as this allows to interpret the results with higher confidence and simultaneously account for the limitations of each method, when applied separately.

Methods

Experimental design and data characteristics. The dietary intervention data sets were selected based on the study design, with a minimum of two independent variables (for example, diet and dose; see Table S1 for details). Prior to the statistical analyses, the data were filtered to keep the OTUs that were present: (1) with relative abundance more than 0.005% in an individual, and (2) in at least 50% of individuals and in one of the groups. For each study in total 507, 561, 560, 397 and 216 OTUs passed this filter and were subsequently used in downstream analyses (Table S1 and Appendix 2). All statistical analyses were performed in R version 3.6.2 [60] unless otherwise specified and were run with 999 permutations.

Zero-value replacements need to be done prior to clr-transformation [24, 25]. This was done by applying function *cMultRepl* with a setting *method* = 'CZM' in the R package *zCompositions* [61].

ANOSIM and **PERMANOVA** were run using the functions *anosim* and *adonis* in the package *vegan* [62], with a setting *method* = *Euclidean* and the clr-transformed data as input.

SIMPER was run on filtered relative abundance data using function *simper* from the R package *vegan*. The average contribution of each OTU to the Bray-Curtis dissimilarity between selected pairwise comparisons of diet levels (specified below) was used as ranking metric.

FFMANOVA was performed by using the function *ffmanova* implemented in the *ffmanova* R package [31] on clr-transformed and centred data. FDR adjusted p-values (setting *pAdjFDR*) were used as ranking metric.

ASCA was run on clr-transformed and centred data using an in-house implementation in MATLAB (R2018b, The MathWorks Inc.); p-values at the community level were calculated by permutation tests using 999 permutations. For study designs with two diet levels, the ASCA loadings were used to rank

OTUs. For designs with more than two diet levels, pairwise comparisons were performed by PLS-DA [63] on the ASCA diet effect matrix plus residuals, and the regression coefficients were used as ranking metric.

LEfSe was run on clr-transformed data in the Galaxy web application and workflow framework [39]. The settings used were as follows: per-sample normalisation = *yes*, $\alpha = 0.05$ for both Kruskal-Wallis and Wilcoxon significance tests, LEfSe threshold = 2 (on a $\log_{10}$ scale), *one-against-all* strategy for multiclass analysis, class variable = *Diet*, subclass variable = *Dose/ Time/ Exercise/ Period*, and subject variable = *Individual*.

ANCOM was run by using the ANCOM 2.0 source code implemented in R [26], using filtered relative abundance data as input. The W-stat was used to rank OTUs, indicating the number of significantly different pairwise log-ratios while adjusting for false discoveries by applying a Benjamini-Hochberg correction at 0.05 level of significance.

ALDEx2 was run using the functions *aldex.clr* and *aldex.glm* from package ALDEx2 v.1.18.0. Raw counts were used as input, and we used p-values for selected factor level contrasts (specified below) to rank OTUs.

Factor level comparisons. For study designs with more than two diet groups, only one of the pairwise comparisons was made for the methods based on the pairwise group comparison, namely ALDEx2, ASCA, LEfSe and SIMPER. The following pairs with the most contrasting outcomes were compared: (1) BSG group vs. IN group [38]; (2) control group vs. 3 g HMW [45] and (3) Placebo 6 weeks group vs. Fibre 6 weeks group [46].

Abbreviations

ANOVA: Analysis of variance; ALDEx2: ANOVA-like differential expression tool for high-throughput sequencing data; ANCOM: Analysis of composition of microbiomes; ANOSIM: Analysis of similarities; ASCA: ANOVA-simultaneous component analysis; FDR: False discovery rate; FFMANOVA: Fifty-fifty multivariate ANOVA; LDA: Linear discriminant analysis; LEfSe: Linear discriminant analysis effect size; OTU: Operational taxonomic unit; PERMANOVA: Permutational multivariate analysis of variance; PLS-DA: Partial least squares discriminant analysis; SIMPER: Similarity percentage.

Declarations

Ethics approval and consent to participate

Not applicable. The study is based on already published data sets listed in Table S1.

Consent for publication

Not applicable.

Availability of data and materials

All data analysed during this study are included as additional files. Custom R and MATLAB scripts are available from the corresponding authors upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was funded by Nofima – Norwegian Institute of Food, Fisheries and Aquaculture Research and Foundation for Research Levy on Agricultural Products (Projects RCN 262306 and 262308). The funding sponsors had no role in the design of the study, the collection, analyses and interpretation of data, in the writing of the manuscript or in the decision to distribute the results.

Author's contributions

IM, IB, and IR conceived the project. MK, IM, IB, and IR designed the project. MK, IM, and IB analysed the data. MK, IM, IB, and IR interpreted the results. MK generated the figures and tables. MK was a major contributor in writing the manuscript, with help from IM and IB. All authors reviewed and edited the manuscript and approved the final version.

Acknowledgements

We thank all the authors who were involved in data generation used in the present study. All data is properly cited and referred to in the Table S1.

References

1. Li H. Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annu Rev Stat Appl.* 2015;2:73–94.
2. Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. *Nat Rev Microbiol.* 2018;16:410.
3. Blanco-Míguez A, Fdez-Riverola F, Sánchez B, Lourenço A. Resources and tools for the high-throughput, multi-omic study of intestinal microbiota. *Brief Bioinform.* 2017;20:1032–56.
4. Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol.* 2016;62:692–703.
5. Lê Cao K-A, Costello M-E, Lakis VA, Bartolo F, Chua X-Y, Brazeilles R, et al. MixMC: a multivariate statistical framework to gain insight into microbial communities. *PLoS One.* 2016;11:e0160169.
6. Mallick H, Ma S, Franzosa EA, Vatanen T, Morgan XC, Huttenhower C. Experimental design and quantitative analysis of microbial community multiomics. *Genome Biol.* 2017;18:228.

7. Waldron L. Data and Statistical Methods To Analyze the Human Microbiome. *mSystems*. 2018;3:e00194-17.
8. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *MBio*. 2018;9:e00525-18.
9. Paliy O, Shankar V. Application of multivariate statistical techniques in microbial ecology. *Mol Ecol*. 2016;25:1032–57.
10. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol*. 2001;26:32–46.
11. Anderson MJ, Walsh DC. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecol Monogr*. 2013;83:557–74.
12. Clarke KR. Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol*. 1993;18:117–43.
13. Schroeder PJ, Jenkins DG. How robust are popular beta diversity indices to sampling error? *Ecosphere*. 2018;9:e02100.
14. Wong RG, Wu JR, Gloor GB. Expanding the UniFrac toolbox. *PLoS One*. 2016;11:e0161196.
15. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71:8228–35.
16. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *ISME J*. 2011;5:169.
17. Roldán Ahumada JA, Avendaño Garrido ML. A Commentary on Diversity Measures UniFrac in Very Small Sample Size. *Evol Bioinform*. 2019;15:1176934319843515.
18. Schloss PD. Evaluating different approaches that test whether microbial communities have the same structure. *ISME J*. 2008;2:265–75.
19. Warton DI, Wright ST, Wang Y. Distance-based multivariate analyses confound location and dispersion effects. *Methods Ecol Evol*. 2012;3:89–101.
20. Randolph TW, Zhao S, Copeland W, Hullar M, Shojaie A. Kernel-penalized regression for analysis of microbiome data. *Ann Appl Stat*. 2018;12:540–66.
21. Tang Z-Z, Chen G, Alekseyenko AV. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. *Bioinformatics*. 2016;32:2618–25.
22. Wu C, Chen J, Kim J, Pan W. An adaptive association test for microbiome data. *Genome Med*. 2016;8:56.
23. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*. 2014;2:15.
24. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:2224.

25. Aitchison J. The Analysis of Compositional Data. London: Chapman and Hall; 1986.
26. Mandal S, Van Treuren W, White RA, Eggesbo M, Knight R, Peddada SD. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis*. 2015;26:27663.
27. Xia Y, Sun J. Hypothesis testing and statistical analysis of microbiome. *Genes Diseases*. 2017;4:138–48.
28. Saccenti E, Hoefsloot HC, Smilde AK, Westerhuis JA, Hendriks MM. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics*. 2014;10:361–74.
29. Stahle L, Wold S. Multivariate analysis of variance (MANOVA). *Chemometr Intell Lab*. 1990;9:127–41.
30. Smilde AK, Jansen JJ, Hoefsloot HC, Lamers R-JA, Van Der Greef J, Timmerman ME. ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data. *Bioinformatics*. 2005;21:3043–8.
31. Langsrud Ø. 50–50 multivariate analysis of variance for collinear responses. *J R Stat Soc Ser D Stat*. 2002;51:305–17.
32. Combrink M, Du Preez I, Ronacher K, Walzl G, Loots DT. Time-Dependent Changes in Urinary Metabolome Before and After Intensive Phase Tuberculosis Therapy: A Pharmacometabolomics Study. *OMICS*. 2019;23:560–72.
33. Gómez-Canela C, Prats E, Lacorte S, Raldúa D, Piña B, Tauler R. Metabolomic changes induced by nicotine in adult zebrafish skeletal muscle. *Ecotox Environ Safe*. 2018;164:388–97.
34. Trimigno A, Khakimov B, Savorani F, Tenori L, Hendrixson V, Čivilis A, et al. Investigation of Variations in the Human Urine Metabolome amongst European Populations: An Exploratory Search for Biomarkers of People at Risk-of-Poverty. *Mol Nutr Food Res*. 2019;63:1800216.
35. Bjerke GA, Rudi K, Avershina E, Moen B, Blom H, Axelsson L. Exploring the brine microbiota of a traditional norwegian fermented fish product (rakfisk) from six different producers during two consecutive seasonal productions. *Foods*. 2019;8:72.
36. Moen B, Røssvoll E, Måge I, Møretrø T, Langsrud S. Microbiota formed on attached stainless steel coupons correlates with the natural biofilm of the sink surface in domestic kitchens. *Can J Microbiol*. 2016;62:148–60.
37. Måge I, Steppeler C, Berget I, Paulsen JE, Rud I. Multi-way methods for understanding longitudinal intervention effects on bacterial communities. *bioRxiv*. 2018:363630.
38. Moen B, Henjum K, Måge I, Knutsen SH, Rud I, Hetland RB, et al. Effect of dietary fibers on cecal microbiota and intestinal tumorigenesis in azoxymethane treated A/J Min/+ mice. *PLoS One*. 2016;11:e0155402.
39. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60.

40. Chi C, Xue Y, Lv N, Hao Y, Liu R, Wang Y, et al. Longitudinal gut bacterial colonization and its influencing factors of low birth weight infants during the first 3 months of life. *Front Microbiol.* 2019;10:1105.
41. Girard C, Tromas N, Amyot M, Shapiro BJ. Gut microbiome of the canadian arctic Inuit. *mSphere.* 2017;2:e00297-16.
42. Langsrud Ø. Rotation tests. *Stat Comput.* 2005;15:53–60.
43. Lai Z-L, Tseng C-H, Ho HJ, Cheung CK, Lin J-Y, Chen Y-J, et al. Fecal microbiota transplantation confers beneficial metabolic effects of diet and exercise on diet-induced obese mice. *Sci Rep.* 2018;8:15625.
44. Le Sciellour M, Labussiere E, Zemb O, Renaudeau D. Effect of dietary fiber content on nutrient digestibility and fecal microbiota composition in growing-finishing pigs. *PLoS One.* 2018;13:e0206159.
45. Wang Y, Ames NP, Tun HM, Tosh SM, Jones PJ, Khafipour E. High molecular weight barley β -glucan alters gut microbiota toward reduced cardiovascular disease risk. *Frontiers in microbiology.* 2016;7:129.
46. Birkeland E, Gharagozlian S, Birkeland KI, Valeur J, Måge I, Rud I, et al. Prebiotic effect of inulin-type fructans on fecal microbiota and short-chain fatty acids in type 2 diabetes: A randomized controlled trial. *Accepted in Eur J Nutr.* 2020.
47. Vangay P, Johnson AJ, Ward TL, Al-Ghalith GA, Shields-Cutler RR, Hillmann BM, et al. US immigration westernizes the human gut microbiome. *Cell.* 2018;175:962–72. e10.
48. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, et al. Daily sampling reveals personalized diet-microbiome associations in humans. *Cell Host Microbe.* 2019;25:789–802.
49. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature.* 2018;555(7695):210–5.
50. Hughes RL, Kable ME, Marco M, Keim NL. The role of the gut microbiome in predicting response to diet and the development of precision nutrition models. Part II: results. *Adv Nutr.* 2019;10:979–98.
51. So D, Whelan K, Rossi M, Morrison M, Holtmann G, Kelly JT, et al. Dietary fiber intervention on gut microbiota composition in healthy adults: a systematic review and meta-analysis. *Am J Clin Nutr.* 2018;107:965–83.
52. Nagpal R, Wang S, Solberg Woods LC, Seshie O, Chung ST, Shively CA, et al. Comparative microbiome signatures and short-chain fatty acids in mouse, rat, non-human primate, and human feces. *Front Microbiol.* 2018;9:2897.
53. Baker DH. Animal models in nutrition research. *Nutr J.* 2008;138:391–6.
54. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genom.* 2006;7:142.

55. Hughes RL, Marco ML, Hughes JP, Keim NL, Kable ME. The Role of the Gut Microbiome in Predicting Response to Diet and the Development of Precision Nutrition Models—Part I: Overview of Current Methods. *Adv Nutr.* 2019;10:953–78.
56. Boulesteix AL, Binder H, Abrahamowicz M, Sauerbrei W. On the necessity and design of studies comparing statistical methods. *Biometrical J.* 2018;60:216–8.
57. Hawinkel S, Mattiello F, Bijnsens L, Thas O. A broken promise: microbiome differential abundance methods do not control the false discovery rate. *Brief Bioinform.* 2017;20:210–21.
58. Thorsen J, Brejnrod A, Mortensen M, Rasmussen MA, Stokholm J, Al-Soud WA, et al. Large-scale benchmarking reveals false discoveries and count transformation sensitivity in 16S rRNA gene amplicon data analysis methods used in microbiome studies. *Microbiome.* 2016;4:62.
59. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017;5:27.
60. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. 2019.
61. Palarea-Albaladejo J, Martin-Fernandez JA. zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometr Intell Lab.* 2015;143:85–96.
62. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-5. <https://CRAN.R-project.org/package=vegan>. 2019.
63. Eriksson L, Byrne T, Johansson E, Trygg J, Vikström C. Multi-and megavariable data analysis basic principles and applications. Malmö: Umetrics Academy; 2013.

Tables

Due to technical limitations, the tables are only available as a download in the supplemental files section.

Table 1. An overview of statistical methods and their properties.

Table 2. Community-level method comparison across five microbial data sets. Distance-based ANOSIM and PERMANOVA and compositional ASCA and FFMANOVA were compared with respect to effect sizes (expressed as percentage of explained variance) and corresponding p-values.

Figures

Figure 1

A diagram of statistical methods used in this study.

Figure 2

Spearman's correlation (x axis) calculated for pairwise comparison of statistical methods (y axis) for five microbial data sets. Each point represents a Spearman's rank correlation coefficient between OTU ranking metrics from Method 1 and Method 2. Point color reflects data set: Moen (red), Lai (blue), Le Sciellour (green), Wang (yellow) and Birkeland (black). LEfSe was evaluated only for data sets Moen and Lai.

Figure 3

Mean relative abundance (log-scale) plotted versus SIMPER average (left) and ANCOM W-stat (right) for the (A) Moen and (B) Lai microbial data sets.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.pdf](#)
- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [FigureS2.pdf](#)
- [Appendix1.xlsx](#)
- [Supplementaryinformation.pdf](#)
- [Appendix2.xlsx](#)
- [TableS1.xlsx](#)