



OPEN

Integrative web cloud computing and analytics using MiPair for design-based comparative analysis with paired microbiome data

Hyojung Jang^{1,3}, Hyunwook Koh^{1,3}✉, Won Gu¹ & Byungkon Kang²

Pairing (or blocking) is a design technique that is widely used in comparative microbiome studies to efficiently control for the effects of potential confounders (e.g., genetic, environmental, or behavioral factors). Some typical paired (block) designs for human microbiome studies are repeated measures designs that profile each subject's microbiome twice (or more than twice) (1) for pre and post treatments to see the effects of a treatment on microbiome, or (2) for different organs of the body (e.g., gut, mouth, skin) to see the disparity in microbiome between (or across) body sites. Researchers have developed a sheer number of web-based tools for user-friendly microbiome data processing and analytics, though there is no web-based tool currently available for such paired microbiome studies. In this paper, we thus introduce an integrative web-based tool, named MiPair, for design-based comparative analysis with paired microbiome data. MiPair is a user-friendly web cloud service that is built with step-by-step data processing and analytic procedures for comparative analysis between (or across) groups or between baseline and other groups. MiPair employs parametric and non-parametric tests for complete or incomplete block designs to perform comparative analyses with respect to microbial ecology (alpha- and beta-diversity) and taxonomy (e.g., phylum, class, order, family, genus, species). We demonstrate its usage through an example clinical trial on the effects of antibiotics on gut microbiome. MiPair is an open-source software that can be run on our web server (<http://mipair.micloud.kr>) or on user's computer (<https://github.com/yj7599/mipairgit>).

The human microbiome is the entire community of all microbes that inhabit different organs (e.g., gut, mouth, nose, skin, etc.) of the human body. The recent advance in next generation sequencing has enabled a faster, cheaper, and more precise quantification of the human microbiome. Then, the human microbiome field has rapidly emerged in both academia and industry. Researchers have found numerous significant discoveries on the effect of a treatment on the human microbiome^{1–5}, the effect of an environmental/behavioral factor on the human microbiome^{6,7}, and/or the effect of the human microbiome on human health or disease^{3,8–14}. However, this would also indicate in contradiction that there can exist many potential confounders that lead to spurious discoveries.

One of the most efficient and practical ways to control for potential confounders is to use pairs (or blocks) at a design stage. Researchers can, for example, profile the human microbiome repeatedly per subject (1) before and after a treatment to see the effects of the treatment on microbiome^{3,15–19} or (2) for different organs of the body to see the disparity in microbiome between (or across) body sites^{20–22}. Then, a study subject forms a pair/block for such repeatedly profiled microbiomes, in which potential confounders (e.g., genetic, environmental, or behavioral factors) are equally retained. Then, the use of appropriate statistical methods for such paired (block) designs can lead to valid and objective conclusions, not distorting the effects of a treatment on microbiome or the disparity in microbiome between (or across) body sites due to confounders.

Researchers have recently developed a sheer number of web-based data processing and analytic tools such as QIIME²³, PUMAA²⁴, MicrobiomeAnalyst²⁵, METAGENassist²⁶, EzBioCloud²⁷ and MiCloud²⁸ for user-friendly

¹Department of Applied Mathematics and Statistics, The State University of New York, Korea, Incheon, South Korea. ²Department of Computer Science, The State University of New York, Korea, Incheon, South Korea. ³These authors contributed equally: Hyojung Jang (co-first author) and Hyunwook Koh (co-first author). ✉email: hyunwook.koh@stonybrook.edu

microbiome data processing and analytics. These web-based tools have greatly accelerated the human microbiome studies with the facilities for cloud computing service and streamlined web environments that are easy-to-use for many people in a variety of disciplines (e.g., medicine, public health, biology, etc.). However, unfortunately, there is no web-based analytic tool currently available for paired microbiome studies. MiCloud²⁸ is the web-based analytic tool that we developed for cross-sectional or longitudinal studies, yet even MiCloud²⁸ can handle confounding effects only through covariate adjustments. Of course, covariate-adjusted analyses are important, though in practice, numerous potential confounders (e.g., genetic, environmental, or behavioral factors) can exist and they are usually lurking (i.e., nuisance variables that are unknown or not available in the data). Hence, it is often very hard to adjust them sufficiently in later statistical modeling.

Therefore, in this paper, we introduce an integrative web-based tool, named MiPair, for design-based comparative analysis with paired microbiome data. MiPair is a user-friendly web cloud service that enables comprehensive data processing and analysis sequentially for comparative analysis between (or across) groups or between baseline and other groups. MiPair employs parametric and non-parametric tests for complete (in which every block contains all possible levels of treatments or body sites) or incomplete (in which not every block contains all possible levels of treatments or body sites) block designs to perform comparative analyses with respect to microbial ecology (alpha- and beta-diversity) and taxonomy (e.g., phylum, class, order, family, genus, species) (Fig. 1). Thus, users can easily deal with comprehensive design-based data analyses with paired microbiome data. MiPair is an open-source software that can be run on our web server (<http://mipair.micloud.kr>) or alternatively on user's computer (<https://github.com/yj7599/mipairgit>).

We organized the rest of the paper as follows. In “Results”, we delineate all individual data processing and analytic components (Fig. 1) with an example clinical trial on the effects of antibiotics on gut microbiome³. To brief, Zhang et al. collected fecal samples from non-obese diabetic mice and profiled their microbiomes using 16S rRNA amplicon sequencing³ and constructed microbiome data using QIIME²⁹, whereas more details on this example study can be found in the original article³. The data were huge and motivated various study orientations, though for demonstration purposes, we reanalyzed a small portion of the data to see if the gut microbiome recovers from the time of a pulsed (macrolide) antibiotic administration (say, baseline) to 2 weeks and 4 weeks afterwards, respectively³ (see “Example”). In “Discussion”, we summarize the results, and importantly, discuss numerous potential applications of MiPair to other microbiome studies based on family/twin or matched designs. Finally, in “Materials and methods”, we described our web server, GitHub repository and the software packages that we used.

Results

Data processing: data input and quality control. We applied most parts of the Data Processing: Data Input and Quality Control component in MiCloud²⁸ to MiPair. Yet, we additionally uploaded three new example datasets for a clinical trial on the effects of antibiotics on gut microbiome³ for users to easily catch up on. These three new example datasets are the ones for (1) a two-group comparison (a baseline group at the time of antibiotic administration and 2 weeks afterwards), (2) a three-group comparison (a baseline group at the time of antibiotic administration and 2 weeks and 4 weeks afterwards) based on a complete block design, where every subject contains all possible three levels of baseline, 2 weeks and 4 weeks afterwards, and (3) a three-group comparison (a baseline group at the time of antibiotic administration and 2 weeks and 4 weeks afterwards) based on an incomplete block design, where not every subject contains all possible three levels of baseline, 2 weeks and 4 weeks afterwards³. In the following sections, we describe the machinery of MiPair using the third example dataset for a three-group comparison based on an incomplete block design.

As in MiCloud²⁸, users first need to upload four requisite data components: (1) feature table [i.e., count data for microbial features such as operational taxonomic units (OTUs) or amplicon sequence variants (ASVs)], (2) taxonomic table (i.e., taxonomic annotations on seven taxonomic ranks, kingdom/domain, phylum, class, order, family, genus, species), (3) metadata/sample information (e.g., treatment status, body sites, pair/block IDs) and (4) phylogenetic tree (i.e., rooted phylogenetic tree) using a unified phyloseq³⁰ format or four individual files (Fig. 1).

Then, the data go through quality controls with respect to (1) a kingdom of interest ['Bacteria' (default) for 16S data, 'Fungi' for ITS data, or any other kingdom of interest for shotgun metagenomic data], (2) a library size for the samples to be removed [i.e., the samples that have a library size/total read count lower than 2000 (default) are removed], (3) a mean proportion for the features (OTUs or ASVs) to be removed [i.e., the microbial features that have a mean proportion lower than 0.002% (default) are removed] and (4) erroneous taxonomic names to be removed (Fig. 1).

MiPair displays summary data [sample size, numbers of features (OTUs, ASVs), phyla, classes, orders, families, genera, and species] using boxes, and data distributions using interactive histograms and box plots before and after quality controls.

Example. We uploaded the data for a three-group comparison based on an incomplete block design and applied the default quality control settings. Then, we rescued 151 features, 6 phyla, 12 classes, 15 orders, 17 families, 22 genera and 8 species for 128 samples (Fig. 2).

Ecological analysis: diversity calculation. As in MiCloud²⁸, MiPair considers a breadth of alpha- and beta-diversity indices that properly modulate the richness and evenness in diversity while reflecting phylogenetic tree information or not^{31–34}. The alpha-diversity indices that MiPair calculates are Observed, Shannon³⁵, Simpson³⁶, Inverse Simpson³⁶, Fisher³⁷, Chao1³⁸, abundance-based coverage estimator (ACE)³⁹, incidence-based coverage estimator (ICE)⁴⁰ and phylogenetic diversity (PD)⁴¹ indices. The beta-diversity indices that MiPair

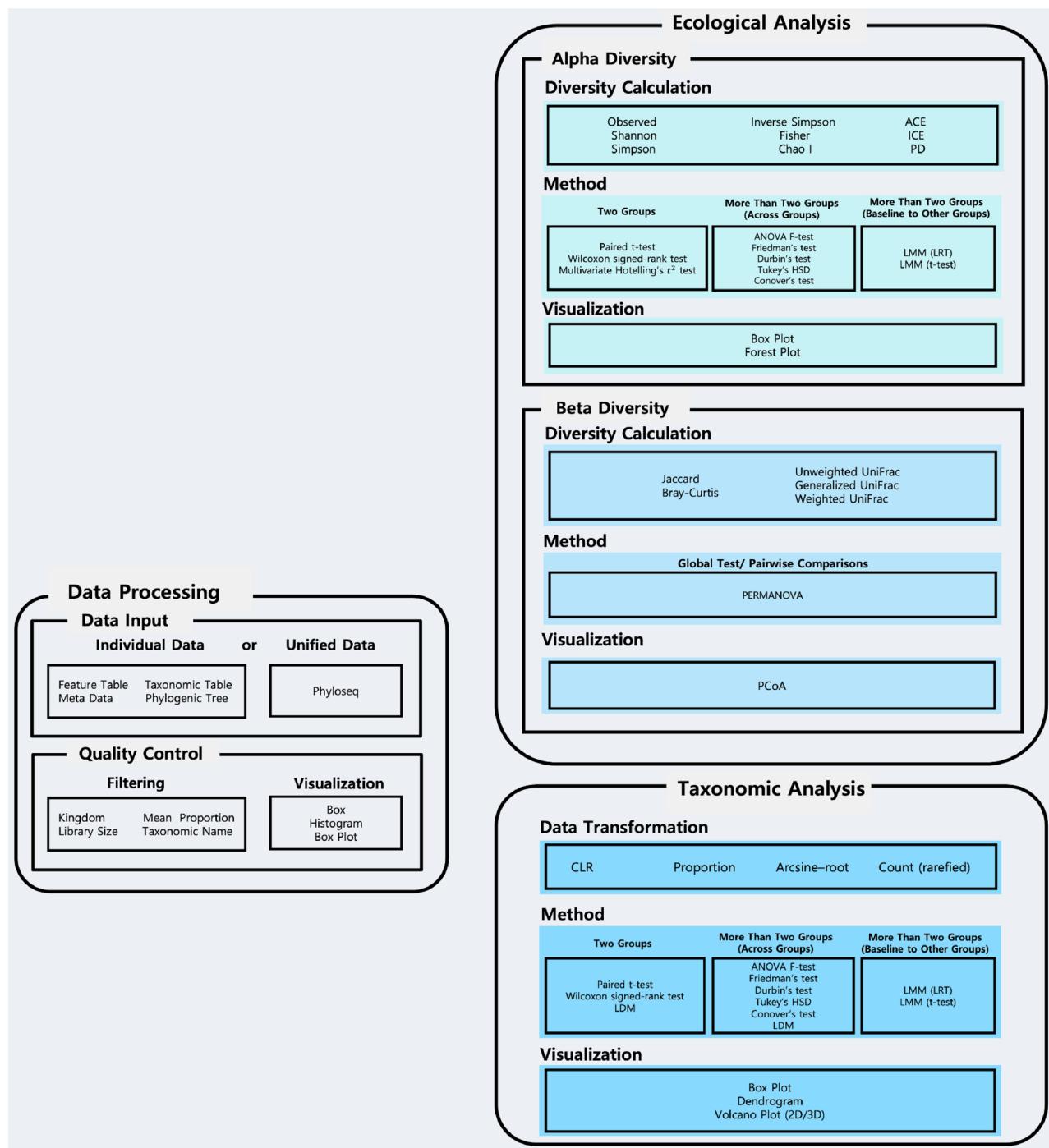


Figure 1. Overall workflow for MiPair. MiPair starts with a data processing component: data processing and then moves to two data analytic components: ecological analysis and taxonomic analysis.

calculates are Jaccard dissimilarity⁴², Bray–Curtis dissimilarity⁴³, Unweighted UniFrac distance⁴⁴, Generalized UniFrac distance⁴⁵ and Weighted UniFrac distance⁴⁶ (Fig. 1) indices. Users can download those alpha- and beta-diversity indices for reference.

Ecological analysis: alpha diversity. MiPair performs comparative analysis in alpha-diversity between (or across) groups (i.e., pre-treatment and post-treatment group(s), different body sites). Users first need to choose a primary variable of interest (i.e., a factor variable that contains multiple groups/levels of treatments or body sites). Then, MiPair lists groups/levels in a chosen primary variable and ask to choose at least two groups/levels to be compared. Then, users need to choose a variable for pair/block IDs (e.g., subjects IDs for pre and post treatments or body sites). Then, MiPair compares two groups or more than two groups (across groups or a baseline group to each of the other groups) in alpha-diversity (Fig. 1) as follows.

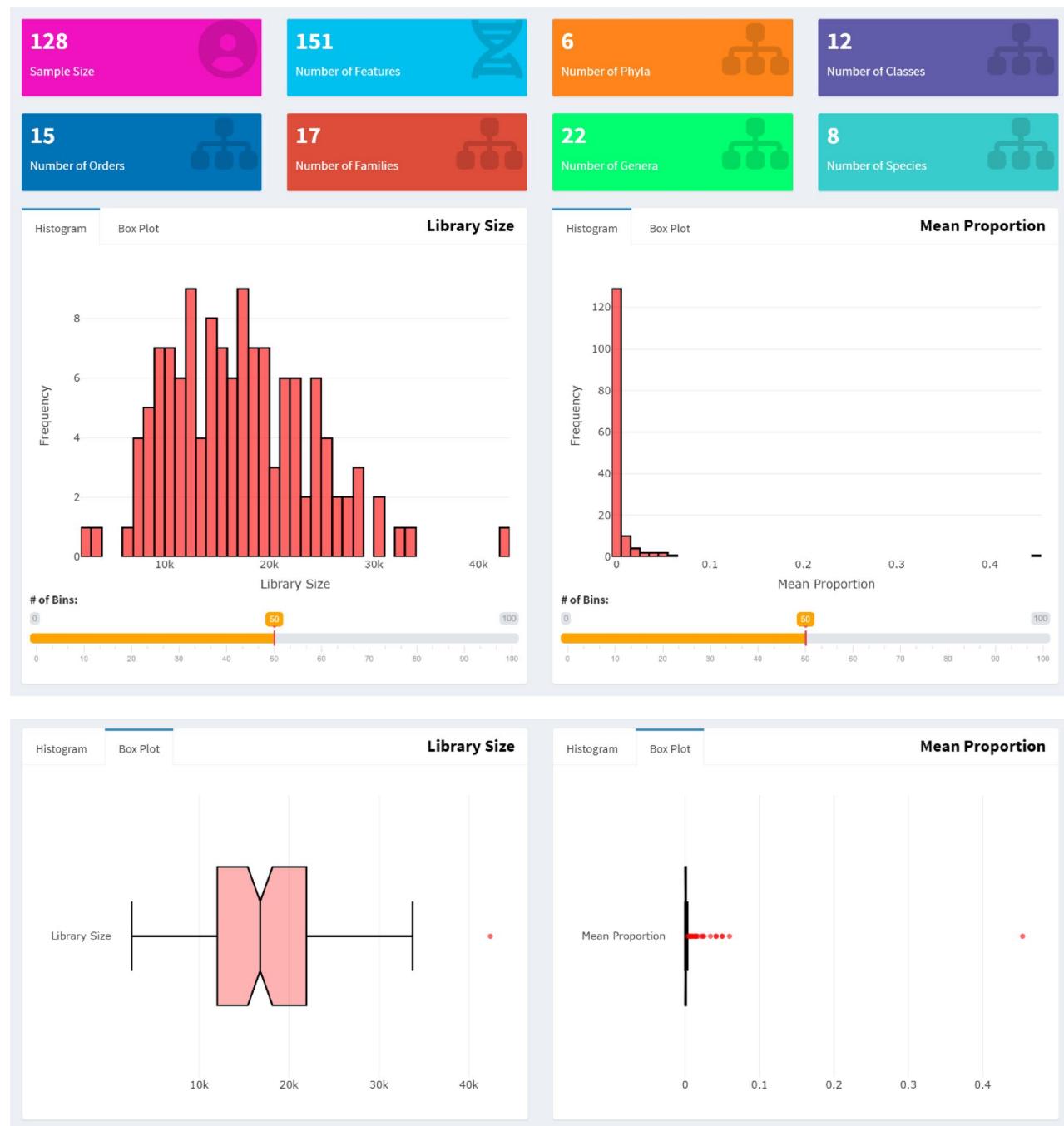


Figure 2. The results after the quality controls of MiPair. MiPair displays summary data (sample size, numbers of features (OTUs, ASVs), phyla, classes, orders, families, genera, and species) using boxes and visualizes the distributions of library sizes across samples and mean proportions across microbial features using histograms and box plots.

Two-group comparison. The parametric Paired t-test or the non-parametric Wilcoxon signed-rank test (default)⁴⁷ can be employed to see if two groups have the same distribution for each alpha-diversity index (H_0) or if they have different distributions (H_1). For omnibus testing to see if the two groups have the same distribution across all alpha-diversity indices (H_0) or if they have different distributions for at least one alpha-diversity index (H_1), the multivariate Hotelling's t-squared test⁴⁸ can also be employed. MiPair visualizes the results using box plots and/or forest plots.

More than two-group comparison (across groups). For the parametric inference, the repeated measures analysis of variance (ANOVA) F-test for global testing (to see if all groups have the same distribution for each alpha-diversity index (H_0) or if at least one group has a different distribution (H_1)) with the Tukey's honestly significant difference (HSD) test⁴⁹ for post-hoc comparisons (to test all possible pairs of groups, individually) can be

employed. For the non-parametric inference in complete block designs, the Friedman's test⁵⁰ for global testing with the Conover's test⁵¹ for post-hoc comparisons can be employed. For the non-parametric inference in incomplete block designs, the Durbin's test for global testing with the Conover's test⁵¹ for post-hoc comparisons can be employed. MiPair visualizes the results using box plots.

More than two-group comparison (baseline to other groups). The likelihood ratio test (LRT) for global testing with the t-test for pairwise comparisons from a baseline group to each of the other groups based on the parametric linear mixed model (LMM)⁵² can be employed. MiPair visualizes the results using box plots.

Example. We performed comparative analysis in alpha-diversity from the baseline group at the time of antibiotic administration to 2 weeks and 4 weeks afterwards³ using LMM for global testing (Fig. 3) and pairwise comparisons (Table 1). We found significant disparity in alpha-diversity for at least one group across the three groups with respect to Shannon, Simpson, Inverse Simpson, Chao 1, ACE, ICE and PD at the significance level of 5% (Fig. 3). We further observed that the alpha-diversity was significantly enriched 2 weeks afterwards with respect to Shannon and PD and 4 weeks afterwards with respect to Shannon, Simpson, Inverse Simpson, Chao 1, ACE, ICE and PD at the significance level of 5% (Table 1).

Ecological analysis: beta diversity. MiPair performs comparative analysis in beta-diversity between (or across) groups (i.e., pre-treatment and post-treatment group(s), different body sites). As in Alpha Diversity, users first need to choose a primary variable of interest (i.e., a factor variable that contains multiple groups/levels of treatments or body sites). Then, MiPair lists groups/levels in a chosen primary variable and ask to choose at least two groups/levels to be compared. Then, users need to choose a variable for pair/block IDs (e.g., subjects IDs for pre and post treatments or body sites). Then, MiPair compares two groups or more than two groups (across groups or a baseline group to each of the other groups) in beta-diversity (Fig. 1) as follows.

Two-group comparison. The nonparametric permutational multivariate analysis of variance (PERMANOVA)^{53,54} for paired microbiome designs can be employed to see if two groups have the same microbiome composition for each beta-diversity index (H_0) or if they have different microbiome compositions (H_1). MiPair visualizes the results using principal coordinate analysis (PCoA) plots⁵⁵.

More than two-group comparison (across groups). MiPair employs PERMANOVA^{53,54} for global testing to see if all groups have the same microbiome composition for each beta-diversity index (H_0) or if at least one group has a different microbiome composition (H_1), and also for pairwise comparisons for all possible pairs of groups individually applying the Benjamini–Hochberg (BH) procedures⁵⁶ to control for false discovery rate (FDR). MiPair visualizes the results using PCoA plots⁵⁵.

More than two-group comparison (baseline to other groups). MiPair employs PERMANOVA^{53,54} for global testing, and also for pairwise comparisons for all possible pairs of a baseline and each of the other groups individually applying the BH procedures⁵⁶ to control for FDR. MiPair visualizes the results using PCoA plots⁵⁵.

Example. We performed comparative analysis in beta-diversity from the baseline group at the time of antibiotic administration to 2 weeks and 4 weeks afterwards³. We found significant disparity in beta-diversity for at least one group across the three groups with respect to all the surveyed beta-diversity indices at the significance level of 5% (Fig. 4). We further observed significant disparity in beta-diversity for all possible pairs of the baseline group and each of the other two groups (2 weeks and 4 weeks afterwards) with respect to all the surveyed beta-diversity indices at the significance level of 5% (Table 2).

Taxonomic analysis: data transformation. For taxonomic analyses at each of the seven taxonomic ranks (phylum, class, order, family, genus and species), MiPair first transforms the original count data into four different data forms, (1) centered log ratio (CLR)⁵⁷ to normalize the data and relax the compositional constraint, (2) proportion to control for varying library sizes across samples, (3) arcsine-root to control for varying library sizes across samples and stabilize the variability across samples (4) count (rarefied)⁵⁸ to control for varying library sizes across samples and use counts as the data form. These data forms have all been widely used, and each of them has both advantages and disadvantages. Hence, it is hard to conclude which data form is superior to the other data forms in all contexts. We set up all such data forms as user options with no default setting. Users can download the original and transformed datasets for reference.

Taxonomic analysis: differential abundance analysis. MiPair performs comparative analysis in each microbial taxon at each of the seven taxonomic ranks (phylum, class, order, family, genus and species). Users first need to choose a data format among CLR⁵⁷, proportion, arcsine-root and count (rarefied)⁵⁸ (Fig. 1). Then, as in Alpha Diversity and Beta Diversity, users need to choose a primary variable of interest (i.e., a factor variable that contains multiple groups/levels of treatments or body sites). Then, MiPair lists groups/levels in a chosen primary variable and ask to choose at least two groups/levels to be compared. Then, users need to choose a variable for pair/block IDs (e.g., subjects IDs for pre and post treatments or body sites). Then, users need to choose to analyze from phylum to genus (default) for 16S rRNA data^{29,59} or from phylum to species for shotgun metagenomic data⁶⁰. Then, MiPair compares two groups or more than two groups (across groups or a baseline group to each of the other groups) in each taxon (Fig. 1) as follows.

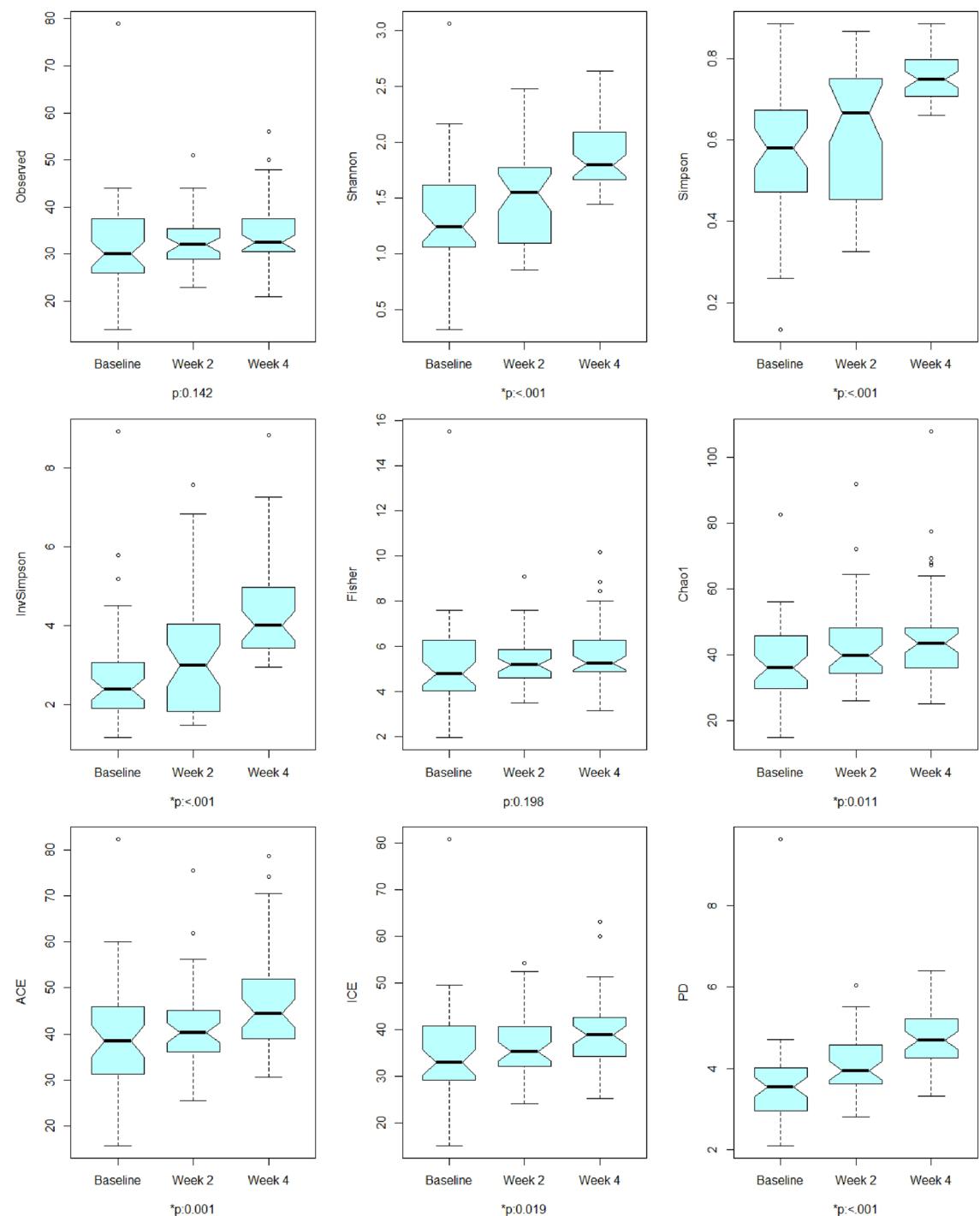


Figure 3. The results for comparative analysis in alpha-diversity (global test). The p-values were calculated using LRT based on LMM for global testing to see if all groups have the same distribution in each alpha-diversity index (H_0) or if at least one group has a different distribution in each alpha-diversity index (H_0). *p represents statistical significance at the level of 5%.

Two-group comparison.

- (1) For CLR: The parametric Paired t-test or the non-parametric Wilcoxon signed-rank test (default)⁴⁷ can be employed to see if two groups have the same distribution for each taxon (H_0) or if they have different distributions (H_1). MiPair applies the BH procedures⁵⁶ to each taxonomic rank to control for FDR. MiPair visualizes the results using box plots and dendograms.
- (2) For Proportion, Arcsine-root or Count (rarefied): The parametric Paired t-test, the non-parametric Wilcoxon signed-rank test⁴⁷, or the non-parametric linear decomposition model (LDM) (default)⁶¹ can be

Alpha-diversity index	Ref	Com	Est	SE	t	Adj. P-value
Observed	Baseline	Week 2	1.568	1.164	0.972	0.334
	Baseline	Week 4	3.229	1.636	1.973	0.103
Shannon	Baseline	Week 2	0.167	0.076	2.195	0.031
	Baseline	Week 4	0.545	0.077	7.086	<0.001
Simpson	Baseline	Week 2	0.051	0.027	1.888	0.063
	Baseline	Week 4	0.190	0.027	6.964	<0.001
InvSimpson	Baseline	Week 2	0.462	0.264	1.745	0.085
	Baseline	Week 4	1.650	0.268	6.148	<0.001
Fisher	Baseline	Week 2	0.249	0.327	0.764	0.447
	Baseline	Week 4	0.592	0.331	1.787	0.155
Chao1	Baseline	Week 2	4.882	2.874	1.699	0.092
	Baseline	Week 4	8.778	2.909	3.017	0.006
ACE	Baseline	Week 2	3.197	2.396	1.334	0.186
	Baseline	Week 4	8.935	2.426	3.683	0.001
ICE	Baseline	Week 2	2.091	1.738	1.203	0.232
	Baseline	Week 4	5.004	1.762	2.840	0.011
PD	Baseline	Week 2	0.484	0.163	2.967	0.004
	Baseline	Week 4	1.098	0.166	6.629	<0.001

Table 1. The results for comparative analysis in alpha-diversity (pairwise comparisons). *Ref represents the reference/baseline group, Com represents the comparison group, Est and SE represent the estimated regression coefficient and its standard error, t represents the t statistic value, and Adj. P-value represents the FDR adjusted P-value.

employed. MiPair applies the BH procedures⁵⁶ to each taxonomic rank to control for FDR. MiPair visualizes the results using box plots and dendograms.

More than two-group comparison (across groups).

- (1) For CLR: For the parametric inference, the repeated measures ANOVA F-test for global testing (to see if all groups have the same distribution for each taxon (H_0) or if at least one group has a different distribution (H_1)) with the Tukey's HSD test⁴⁹ for post-hoc comparisons (to test all possible pairs of groups individually) can be employed. For the non-parametric inference in complete block designs, the Friedman's test⁵⁰ for global testing with the Conover's test⁵¹ for post-hoc comparisons (default) can be employed. For the non-parametric inference in incomplete block designs, the Durbin's test for global testing with the Conover's test⁵¹ for post-hoc comparisons (default) can be employed. MiPair applies the BH procedures⁵⁶ to each taxonomic rank to control for FDR. MiPair visualizes the results using box plots and interactive volcano plots.
- (2) For Proportion, Arcsine-root or Count (rarefied): For the parametric inference, the repeated measures ANOVA F-test for global testing (to see if all groups have the same distribution for each taxon (H_0) or if at least one group has a different distribution (H_1)) with the Tukey's HSD test⁴⁹ for post-hoc comparisons (to test all possible pairs of groups individually) can be employed. For the non-parametric inference in complete block designs, the Friedman's test⁵⁰ for global testing with the Conover's test⁵¹ for post-hoc comparisons can be employed. For the non-parametric inference in incomplete block designs, the Durbin's test for global testing with the Conover's test⁵¹ for post-hoc comparisons can be employed. For the non-parametric inference in either incomplete or complete block designs, LDM (default)⁶¹ can be employed for both global testing and pairwise comparisons. MiPair applies the BH procedures⁵⁶ to each taxonomic rank to control for FDR. MiPair visualizes the results using box plots and interactive volcano plots.

More than two-group comparison (baseline to other groups). For either CLR, Proportion, Arcsine-root or Count (rarefied), the likelihood ratio test (LRT) for global testing with the t-test for pairwise comparisons from a baseline group to each of the other groups based on LMM⁵² can be employed. MiPair applies the BH procedures⁵⁶ to each taxonomic rank to control for FDR. MiPair visualizes the results using box plots and interactive volcano plots.

Example. We chose CLR (default) as the data format to use and performed comparative analysis in each genus from the baseline group at the time of antibiotic administration to 2 weeks and 4 weeks afterwards³ using LMM for both global testing (Fig. 5) and pairwise comparisons (Table 3, Fig. 6). We found significant disparity in CLR transformed relative abundance for at least one group across the three groups for 15 genera at the significance level of 5% (Figs. 5, 6). Table 3 reports the results for those 15 genera in the context of pairwise comparisons between the baseline group and 2 weeks afterwards, and between the baseline group and 4 weeks afterwards, respectively.

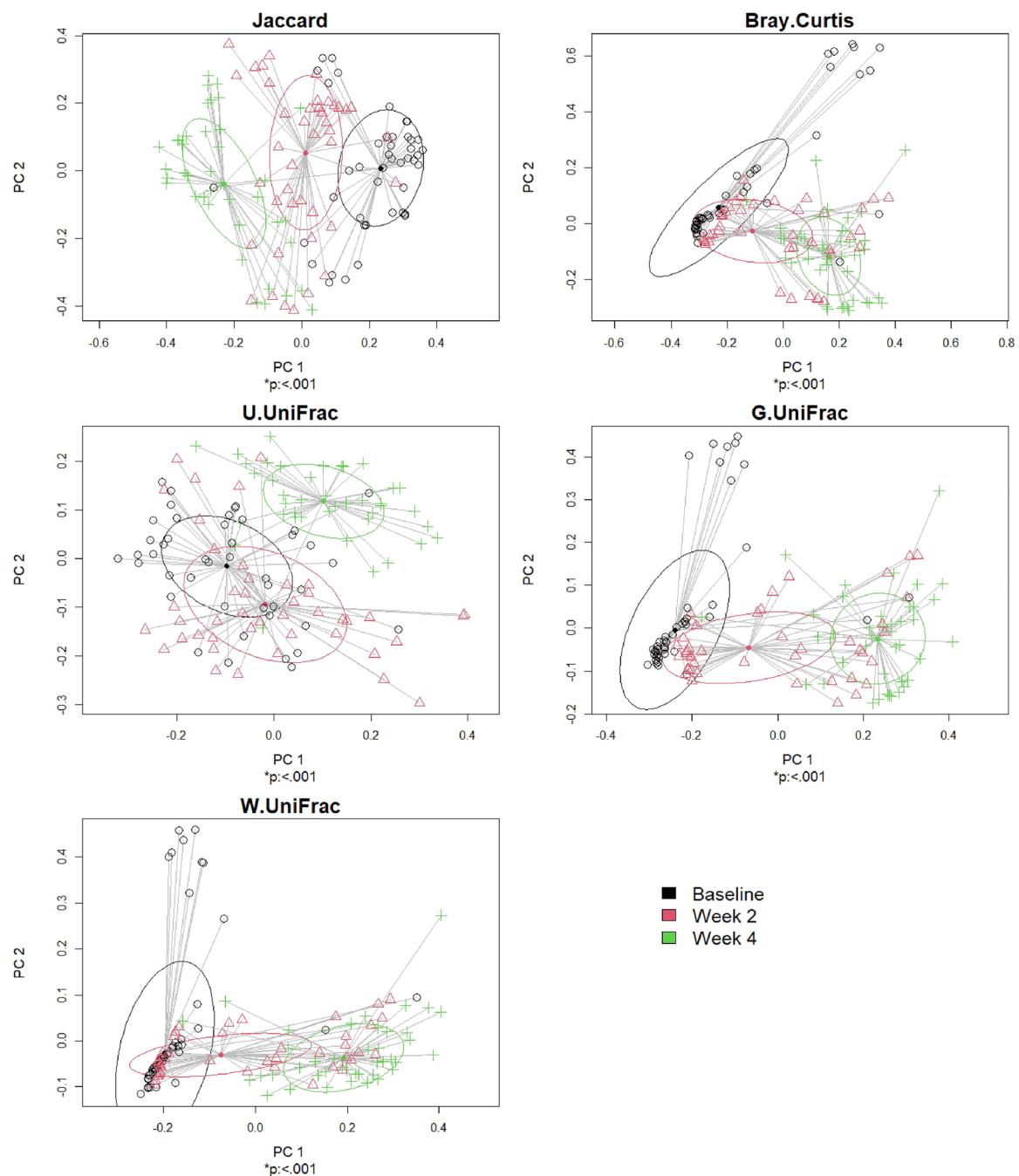


Figure 4. The results for comparative analysis in beta-diversity (global test). The p-values were calculated using PERMANOVA for global testing if all groups have the same microbiome composition in each beta-diversity index (H_0) or if at least one group has a different microbiome composition in each beta-diversity index (H_1). $*p$ represents statistical significance at the level of 5%.

Discussion

In this paper, we introduced an open-source web-based analytic tool, MiPair, for design-based comparative analysis with paired microbiome data. We described that MiPair can handle comprehensive microbiome data processing and analytic procedures using parametric or non-parametric tests for complete (in which every block contains all possible levels of treatments or body sites) or incomplete (in which not every block contains all possible levels of treatments or body sites) block designs to perform comparative analyses with respect to microbial ecology (alpha- and beta-diversity) and taxonomy (e.g., phylum, class, order, family, genus, species). We also described all the detailed widgets, methodologies and visualizations for the two-group comparison, more than two-group comparison (across groups) and more than two-group comparison (baseline to other groups), respectively.

Beta-diversity index	Ref	Com	F	Adj. P-value
Jaccard	Baseline	Week 2	11.828	< 0.001
	Baseline	Week 4	19.136	< 0.001
Bray.Curtis	Baseline	Week 2	9.468	< 0.001
	Baseline	Week 4	21.565	< 0.001
U.UniFrac	Baseline	Week 2	8.226	< 0.001
	Baseline	Week 4	14.584	< 0.001
G.UniFrac	Baseline	Week 2	15.690	< 0.001
	Baseline	Week 4	44.064	< 0.001
W.UniFrac	Baseline	Week 2	12.951	< 0.001
	Baseline	Week 4	53.649	< 0.001

Table 2. The results for comparative analysis in beta-diversity (pairwise comparisons). *Ref represents the reference/baseline group, Com represents the comparison group, F represents the F statistic value, and Adj. P-value represents the FDR adjusted P-value.

We demonstrated the application of MiPair using an example clinical trial to see if the gut microbiome recovers from the time of a pulsed (macrolide) antibiotic administration to 2 weeks and 4 weeks afterwards, respectively³. However, the application of MiPair can be much broader. MiPair can be, in general, applied to any paired (block) designs, in which each pair/block contains different groups or levels of treatments. In the main text, we described subjects as example pairs or blocks for repeated measurements for different groups or levels of treatments or different body sites, yet twins or families can also be example pairs or blocks to control for genetic and/or environmental factors as in Refs.^{9,12,62,63}. Besides, any groups of subjects that are matched in selected nuisance variables (e.g., age, sex) in an observational or quasi-experimental study can be pairs or blocks to control for such matched nuisance variables (e.g., age, sex) as in Refs.^{64,65}. MiPair can substantially contribute to the rapidly growing human microbiome field as a useful and user-friendly data analytic tool for numerous potential applications.

Materials and methods

Web server, GitHub, URLs and pre-requisites. As in Ref.²⁸, we constructed all the user interfaces and server functions of our app using R Shiny (<https://shiny.rstudio.com>), and distributed our app to web environments using ShinyProxy (<https://www.shinyproxy.io>) and Apache2 (<https://httpd.apache.org>). Our web server currently runs on Ubuntu 20.04 (<https://ubuntu.com>) on the computing device with Intel Core i7-12700T (12-core) processor and 36 GB DDR4 memory allowing up to ten concurrent connections. We also set up a GitHub repository to allow users to run MiPair using their local computers in case that our web server is busy. We are the host that is responsible for maintaining our web server and GitHub repository stable. Users can report any issues that they have to us through the GitHub page (<https://github.com/yj7599/mipairgit/issues>).

URLs. MiPair is an open-source software, and can be reached through our web server (<http://mipair.micloud.kr>) or our GitHub repository (<https://github.com/yj7599/mipairgit>) locally on user's computer.

Pre-requisites. MiPair depends on many other existing R packages, and thus it seems to require many pre-installations. However, users do not need to install them all individually because they are already installed on our web server. For the local device, they can also be installed and imported automatically using a simple command, `library(shiny); shiny::runGitHub("mipairgit", "yj7599", ref = "main")`, using the 'shiny' package on R Studio (<https://www.rstudio.com>). We have run unit tests using our web server with the specifications of Intel Core i7-12700T (12-core) processor and 36 GB DDR4 memory on Ubuntu 20.04 with R version 4.2.0, and also using two different local computers with the specifications of AMD Ryzen 7 5800U (8-core) processor and 8 GB DDR4 memory on Windows 11 Home (Version: 21H2, Build: 22000.1098) with R version 4.1.0 and the specifications of Apple M1 Ultra (20-core) processor and 64 GB memory on macOS Monterey 12.4 with R version 4.2.0, respectively. We have checked up each possible combination of the computing devices, datasets, and functionalities. For the datasets, we used the three example datasets³ and a huge synthetic dataset. The synthetic dataset was the one generated based on the Dirichlet-multinomial model⁶⁶ using the estimated proportions and dispersion parameter of the gut microbiome data for the monozygotic twins in Ref.⁹. We generated the feature table for 6671 features and 3000 subjects, and created the metadata to have blocks with size three arbitrarily for the three-group comparison. Of course, the use of this synthetic dataset does not provide any biological or medical meanings at all. We used it just to check the running times for using such a huge dataset to provide some guideline on the upper limit of the data size that can be handled by MiPair. We organized the results from our unit tests in (Online resource, Supplementary Table 1). To summarize, we found no error for any procedure (Online resource, Supplementary Table 1). We also observed only small running times for any procedure for any of the three example datasets, yet we observed much greater running times for the huge synthetic dataset (Online resource, Supplementary Table 1). However, we would say that MiPair can still handle a huge dataset like the synthetic dataset with 6671 features and 3000 subjects in a manageable time. For the local device, we would also set up the minimum requirements as the one with 8-core processor and 8 GB memory on Windows

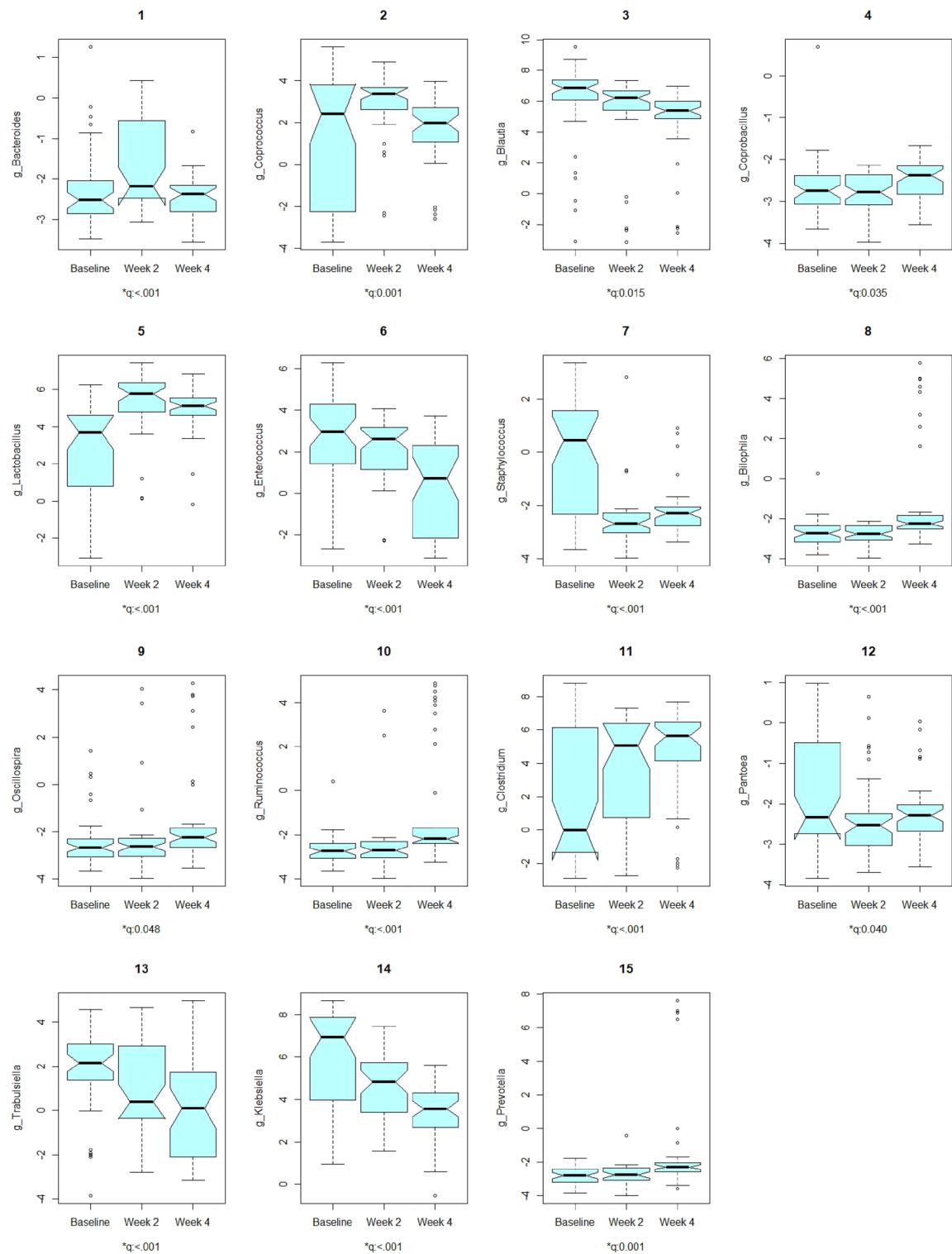


Figure 5. The 15 significant discoveries for comparative analysis on genera (global test). The Q-values are the FDR adjusted P-values for global testing using LRT based on LMM to see if all groups have the same distribution in each genus (H_0) or if at least one group has a different distribution in each genus index (H_0).

or Macintosh with R ($\geq 4.1.0$). We monitor the capacity and functionality of our web server periodically. Users can also report any issues for using MiPair on our GitHub Issues page (<https://github.com/yj7599/mipairgit/issues>). We also plan to provide troubleshooting tips on our GitHub page (<https://github.com/yj7599/mipairgit>).

Software packages. We wrote MiPair using R language, and MiPair is based on many R packages as follows.

Genus	Ref	Com	Est	SE	t	Adj. P-value
g_Bacteroides	Baseline	Week 2	0.772	0.198	3.898	0.001
	Baseline	Week 4	-0.201	0.200	-1.004	0.389
g_Coprococcus	Baseline	Week 2	1.876	0.468	4.006	0.001
	Baseline	Week 4	0.597	0.474	1.258	0.290
g_Blautia	Baseline	Week 2	-0.938	0.483	-1.942	0.111
	Baseline	Week 4	-1.519	0.490	-3.099	0.012
g_Coprobacillus	Baseline	Week 2	-0.127	0.100	-1.270	0.285
	Baseline	Week 4	0.161	0.101	1.586	0.285
g_Lactobacillus	Baseline	Week 2	2.339	0.334	6.995	<0.001
	Baseline	Week 4	1.892	0.339	5.579	<0.001
g_Enterococcus	Baseline	Week 2	-0.765	0.375	-2.038	0.098
	Baseline	Week 4	-2.705	0.380	-7.109	<0.001
g_Staphylococcus	Baseline	Week 2	-2.504	0.314	-7.973	<0.001
	Baseline	Week 4	-2.157	0.318	-6.779	<0.001
g_Bilophila	Baseline	Week 2	-0.092	0.340	-0.272	0.864
	Baseline	Week 4	1.552	0.344	4.514	<0.001
g_Oscillospira	Baseline	Week 2	0.104	0.320	0.325	0.863
	Baseline	Week 4	0.793	0.325	2.442	0.067
g_Ruminococcus	Baseline	Week 2	0.195	0.351	0.555	0.751
	Baseline	Week 4	1.768	0.356	4.968	<0.001
g_Clustridium	Baseline	Week 2	2.142	0.596	3.591	0.002
	Baseline	Week 4	3.057	0.605	5.050	<0.001
g_Pantoea	Baseline	Week 2	-0.568	0.221	-2.573	0.062
	Baseline	Week 4	-0.453	0.224	-2.025	0.076
g_Trabulsiella	Baseline	Week 2	-0.916	0.307	-2.986	0.012
	Baseline	Week 4	-1.663	0.312	-5.332	<0.001
g_Klebsiella	Baseline	Week 2	-1.379	0.348	-3.968	0.001
	Baseline	Week 4	-2.654	0.353	-7.529	<0.001
g_Prevotella	Baseline	Week 2	0.048	0.362	0.131	0.896
	Baseline	Week 4	1.349	0.366	3.683	0.002

Table 3. The results for comparative analysis on genera (pairwise comparisons). *Ref represents the reference/baseline group, Com represents the comparison group, Est and SE represent the estimated regression coefficient and its standard error, t represents the t statistic value, and Adj. P-value represents the FDR adjusted P-value.

Diversity calculation and data transformation. The alpha- and beta-diversity indices^{35–46} are calculated using the ‘phyloseq’, ‘picante’, ‘dist’, ‘ecodist’ and ‘GUniFrac’ packages. The CLR⁵⁷ transformation and rarefaction⁵⁸ are performed using the ‘compositions’ and ‘phyloseq’ packages.

Data analytic methods. The Paired t-test, Wilcoxon signed rank test⁴⁷, and multivariate Hotelling’s t-squared test⁴⁸ are performed using the ‘stats’ and ‘ICSNP’ packages. The ANOVA F-test, Friedman’s test⁵⁰, Durbin test, Tukey’s HSD⁴⁹ and Conover’s test⁵¹ are performed using the ‘stats’ and ‘PMCMRplus’ packages. The LMM⁵² is fitted using the ‘lme4’ package. The LDM⁶¹ is fitted using the ‘LDM’ package. PERMANOVA^{53,54} is performed using the ‘vegan’ package. The BH procedures⁵⁶ are applied using the ‘stats’ package.

Visualizations. The box plots, histograms and forest plots are drawn using the ‘graphics’ and ‘forestplot’ packages. The PCoA plots⁵⁵ are drawn using the ‘vegan’ package. The volcano plots are drawn using ‘plotly’ and ‘volcano3D’ packages.

Data availability

The raw sequence data for our example demonstration are publicly available in the database QIITA with the identifier 10508 (<https://qiita.ucsd.edu/study/description/10508>), and all the processed data components can be found on the app (see example datasets on Data Processing: Data Input). MiPair is an open-source software under the General Public License (GPL-1, GPL-2), which can be run on our web server (<http://mipair.micloud.kr>) or on user’s computer (<https://github.com/yj7599/mipairgt>).

Received: 15 September 2022; Accepted: 24 November 2022

Published online: 28 November 2022

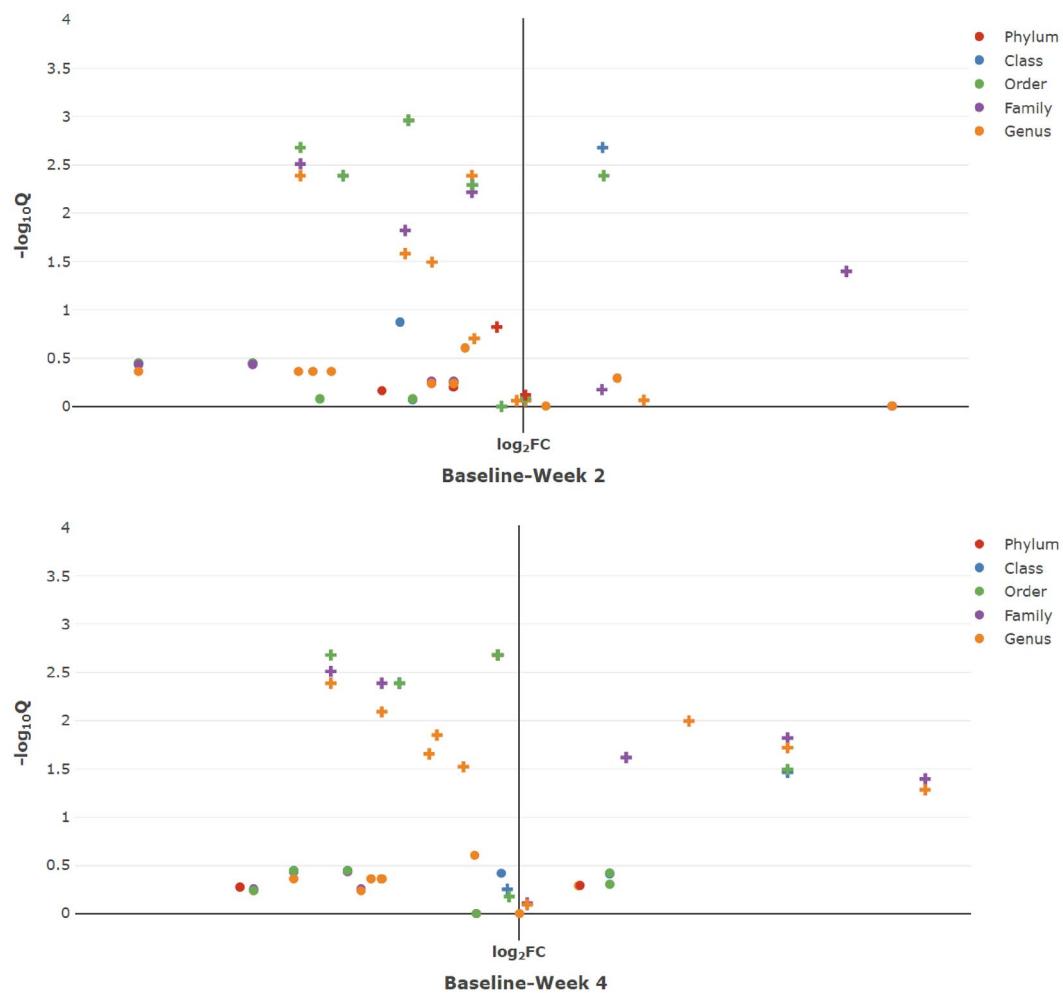


Figure 6. The volcano plot to overview the taxonomic differential abundances. Left: between the baseline group at the time of antibiotic administration and 2 weeks afterwards. Right: between the baseline group at the time of antibiotic administration and 4 weeks afterwards (right). x represents significantly differential taxa.

References

1. Han, M. K. *et al.* Association between lung microbiome and disease progression in IPF; A prospective cohort study. *Lancet Respir. Med.* **2**, 548–556. [https://doi.org/10.1016/S2213-2600\(14\)70069-4](https://doi.org/10.1016/S2213-2600(14)70069-4) (2014).
2. Livanos, A. E. *et al.* Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nat. Microbiol.* **1**, 6140. <https://doi.org/10.1038/nmicrobiol.2016.140> (2016).
3. Zhang, X. S. *et al.* Antibiotic-induced acceleration of type 1 diabetes alters maturation of innate intestinal immunity. *Elife* **7**, e37816. <https://doi.org/10.7554/elife.37816> (2018).
4. Vich, V. A. *et al.* Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. *Nat. Commun.* **11**, 362. <https://doi.org/10.1038/s41467-019-14177-z> (2020).
5. Gui, X., Yang, Z. & Li, M. D. Effect of cigarette smoke on gut microbiota: State of knowledge. *Front. Physiol.* **12**, 673341. <https://doi.org/10.3389/fphys.2021.673341> (2021).
6. Singh, R. K. *et al.* Influence of diet on the gut microbiome and implications for human health. *J. Transl. Med.* **15**, 73. <https://doi.org/10.1186/s12967-017-1175-y> (2017).
7. Liu, R. *et al.* Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nat. Med.* **23**, 859–868. <https://doi.org/10.1038/nm.4358> (2017).
8. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031. <https://doi.org/10.1038/nature05414> (2006).
9. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799. <https://doi.org/10.1016/j.cell.2014.09.053> (2014).
10. Frankel, A. E. *et al.* Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific human gut microbiota and metabolites associated with immune checkpoint therapy efficacy in melanoma patients. *Neoplasia* **19**, 848–855. <https://doi.org/10.1016/j.neo.2017.08.004> (2017).
11. Gopalakrishnan, V. *et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science* **359**, 97–103. <https://doi.org/10.1126/science.aan4236> (2018).
12. Matson, V. *et al.* The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science* **359**, 104–108. <https://doi.org/10.1126/science.aoa3290> (2018).
13. Sharma, S. & Tripathi, P. Gut microbiome and type 2 diabetes: Where we are and where to go?. *J. Nutr. Biochem.* **63**, 101–108. <https://doi.org/10.1016/j.jnutbio.2018.10.003> (2019).

14. Glassner, K. L., Abraham, B. P. & Quigley, E. M. The microbiome and inflammatory bowel disease. *J. Allergy Clin. Immunol.* **145**, 16–27. <https://doi.org/10.1016/j.jaci.2019.11.003> (2020).
15. Joffe, H. *et al.* Low-dose estradiol and the serotonin-norepinephrine reuptake inhibitor venlafaxine for vasomotor symptoms: a randomized clinical trial. *JAMA Intern. Med.* **174**, 1058–1066. <https://doi.org/10.1001/jamainternmed.2014.1891> (2014).
16. Hall, A. B. *et al.* A novel *Ruminococcus gnatus* clade enriched in inflammatory bowel disease patients. *Genome Med.* **9**, 103. <https://doi.org/10.1186/s13073-017-0490-5> (2014).
17. Mitchell, C. M. *et al.* Vaginal microbiota and genitourinary menopausal symptoms: A cross-sectional analysis. *Menopause* **24**, 1160–1166. <https://doi.org/10.1097/GME.0000000000000904> (2017).
18. Kusakabe, S. *et al.* Pre-and post-serial metagenomic analysis of gut microbiota as a prognostic factor in patients undergoing haematopoietic stem cell transplantation. *Br. J. Haematol.* **188**, 438–449. <https://doi.org/10.1111/bjh.16205> (2020).
19. Izhak, M. B. *et al.* Projection of gut microbiome pre- and post- bariatric surgery to predict surgery outcome. *mSystems* **6**, 3. <https://doi.org/10.1128/mSystems.01367-20> (2021).
20. Charlson, E. S. *et al.* Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS One* **5**, 12. <https://doi.org/10.1371/journal.pone.0015216> (2010).
21. Jiang, Y. *et al.* Comparison of red-complex bacteria between saliva and subgingival plaque of periodontitis patients: A systematic review and meta-analysis. *Front. Cell Infect. Microbiol.* **11**, 727732. <https://doi.org/10.3389/fcimb.2021.727732> (2021).
22. Reyman, M. *et al.* Microbial community networks across body sites are associated with susceptibility to respiratory infections in infants. *Commun. Biol.* **4**, 1233. <https://doi.org/10.1038/s42003-021-02755-1> (2021).
23. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME2. *Nat. Biotechnol.* **37**, 852–857. <https://doi.org/10.1038/s41587-019-0209-9> (2019).
24. Mitchell, K. *et al.* PUMAA: A platform for accessible microbiome analysis in the undergraduate classroom. *Front. Microbiol.* **11**, 584699. <https://doi.org/10.3389/fm.0000000000000904> (2020).
25. Dhariwal, A. *et al.* MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* **45**, W1. <https://doi.org/10.1093/nar/gkx295> (2017).
26. Arndt, D. *et al.* METAGENassist: A comprehensive web server for comparative metagenomics. *Nucleic Acids Res.* **40**, W1. <https://doi.org/10.1093/nar/gks497> (2012).
27. Yoon, S. H. *et al.* Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* **67**, 1613–1617. <https://doi.org/10.1099/ijsem.0.001755> (2017).
28. Gu, W. *et al.* MiCloud: A unified web platform for comprehensive microbiome data analysis. *PLoS ONE* **17**, 8. <https://doi.org/10.1371/journal.pone.0272354> (2022).
29. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336. <https://doi.org/10.1038/nmeth.f.303> (2010).
30. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, 4. <https://doi.org/10.1371/journal.pone.0061217> (2013).
31. Koh, H. An adaptive microbiome α -diversity-based association analysis method. *Sci. Rep.* **8**, 1. <https://doi.org/10.1038/s41598-018-36355-7> (2018).
32. Zhao, N. *et al.* Testing in microbiome-profiling studies with MiRKAT, the microbiome regression-based kernel association test. *Am. J. Hum. Genet.* **96**, 797–807. <https://doi.org/10.1016/j.ajhg.2015.04.003> (2015).
33. Koh, H., Li, Y., Zhan, X., Chen, J. & Zhao, N. A distance-based kernel association test based on the generalized linear mixed model for correlated microbiome studies. *Front. Genet.* **10**, 458. <https://doi.org/10.3389/fgene.2019.00458> (2019).
34. Wilson, N. *et al.* MiRKAT: Kernel machine regression-based global association tests for the microbiome. *Bioinformatics* **37**, 1595–1597. <https://doi.org/10.1093/bioinformatics/btaa951> (2021).
35. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
36. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688. <https://doi.org/10.1038/163688a0> (1949).
37. Fisher, R. A., Corbet, A. S. & Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58. <https://doi.org/10.2307/1411> (1943).
38. Chao, A. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* **11**, 265–270 (1984).
39. Chao, A. & Lee, S. M. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* **87**, 210–217. <https://doi.org/10.2307/2290471> (1992).
40. Lee, S. M. & Chao, A. Estimating population size via sample coverage for closed capture-recapture models. *Biometrics* **50**, 88–97. <https://doi.org/10.2307/2533199> (1994).
41. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3) (1992).
42. Jaccard, P. The distribution of the flora in the alpine zone. *New Phytol.* **11**, 37–50 (1912).
43. Bray, J. R. & Curtis, J. T. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* **27**, 325–349. <https://doi.org/10.2307/1942268> (1957).
44. Lozupone, C. & Knight, R. UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235. <https://doi.org/10.1128/AEM.71.12.8228-8235.2005> (2005).
45. Chen, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**, 2106–2113. <https://doi.org/10.1093/bioinformatics/bts342> (2012).
46. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* **73**, 1576–1585. <https://doi.org/10.1128/AEM.01996-06> (2007).
47. Wilcoxon, F. Individual comparisons by ranking methods. *Biometr. Bull.* **1**, 80–83. <https://doi.org/10.2307/3001968> (1945).
48. Hotelling, H. The generalization of Student's ratio. *Ann. Math. Stat.* **2**, 360–378 (1931).
49. Tukey, J. Comparing individual means in the analysis of variance. *Biometrics* **5**, 99–114. <https://doi.org/10.2307/3001913> (1949).
50. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **32**, 675–701. <https://doi.org/10.2307/2279372> (1937).
51. Conover, W. J. *Practical Nonparametric Statistics*, 3rd ed. 428–433 (Wiley, 1999).
52. Laird, N. M. & Ware, J. H. Random-effects models for longitudinal data. *Biometrics* **38**, 963–974. <https://doi.org/10.2307/2529876> (1982).
53. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* **26**, 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.ppx> (2001).
54. McArdle, B. H. & Anderson, M. J. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* **82**, 290–297. <https://doi.org/10.1126/science.aoa3290> (2001).
55. Torgerson, W. S. Multidimensional scaling: I. Theory and method. *Psychometrika* **17**, 401–419. <https://doi.org/10.1007/BF02288916> (1952).
56. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57**, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x> (1995).
57. Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **44**, 139–160 (1982).
58. Sanders, H. L. Marine benthic diversity: A comparative study. *Am. Nat.* **102**, 243–282 (1968).

59. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152. <https://doi.org/10.1101/gr.085464.108> (2009).
60. Thomas, T., Gilbert, J. & Meyer, F. Metagenomics—A guide from sampling to data analysis. *Microb. Inform. Exp.* **2**, 3. <https://doi.org/10.1186/2042-5783-2-3> (2012).
61. Zhu, Z., Satten, G. A., Mitchell, C. & Hu, Y. Constraining PERMANOVA and LDM to within-set comparisons by projection improves the efficiency of analyses of matched sets of microbiome data. *Microbiome* **9**, 133. <https://doi.org/10.1186/s40168-021-01034-9> (2021).
62. Coelho, L. P. *et al.* Similarity of the dog and human gut microbiomes in gene content and response to diet. *Microbiome* **6**, 72. <https://doi.org/10.1186/s40168-018-0450-3> (2018).
63. Van, D. E., Knol, J. & Belzer, C. Microbial transmission from mother to child: Improving infant intestinal microbiota development by identifying the obstacles. *Crit. Rev. Microbiol.* **45**, 613–648. <https://doi.org/10.1080/1040841X.2019.168060> (2019).
64. Vogt, N. M. *et al.* Gut microbiome alterations in Alzheimer’s disease. *Sci. Rep.* **7**, 13537. <https://doi.org/10.1038/s41598-017-13601-y> (2017).
65. Zhao, N. *et al.* Low diversity in nasal microbiome associated with staphylococcus aureus colonization and bloodstream infections in hospitalized neonates. *Open Forum Infect. Dis.* **8**, 10. <https://doi.org/10.1093/ofid/ofab475> (2021).
66. Mosimann, J. E. On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika* **49**, 65–82 (1962).

Acknowledgements

The authors are grateful for anonymous reviewers for their careful observations and insightful comments.

Author contributions

H.K. conceived the concept and methods. H.J. and H.K. wrote the manuscript. H.J., H.K. and W.G., wrote the programs. H.J., W.G. and B.K. constructed the web server and GitHub repository. H.J. and H.K. contributed equally as co-first authors. H.K. is the corresponding author. All authors reviewed the manuscript.

Funding

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2021R1C1C1013861).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-25093-6>.

Correspondence and requests for materials should be addressed to H.K.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2022