A maximum-type microbial differential abundance test with application to high-dimensional microbiome data analyses

Zhengbang Li¹, Xiaochen Yu¹, Hongping Guo², TingFang Lee³ and Jiyuan Hu³*

¹School of Mathematics and Statistics, Central China Normal University, Wuhan, China, ²School of Mathematics and Statistics, Hubei Normal University, Huangshi, China, ³Division of Biostatistics, Department of Population Health, New York University (NYU) Grossman School of Medicine, New York, NY, United States

Background: High-throughput metagenomic sequencing technologies have shown prominent advantages over traditional pathogen detection methods, bringing great potential in clinical pathogen diagnosis and treatment of infectious diseases. Nevertheless, how to accurately detect the difference in microbiome profiles between treatment or disease conditions remains computationally challenging.

Results: In this study, we propose a novel test for identifying the difference between two high-dimensional microbiome abundance data matrices based on the centered log-ratio transformation of the microbiome compositions. The test p-value can be calculated directly with a closed-form solution from the derived asymptotic null distribution. We also investigate the asymptotic statistical power against sparse alternatives that are typically encountered in microbiome studies. The proposed test is maximum-type equal-covariance-assumption-free (MECAF), making it widely applicable to studies that compare microbiome compositions between conditions. Our simulation studies demonstrated that the proposed MECAF test achieves more desirable power than competing methods while having the type I error rate well controlled under various scenarios. The usefulness of the proposed test is further illustrated with two real microbiome data analyses. The source code of the proposed method is freely available at https://github.com/Jiyuan-NYU-Langone/MECAF.

Conclusions: MECAF is a flexible differential abundance test and achieves statistical efficiency in analyzing high-throughput microbiome data. The proposed new method will allow us to efficiently discover shifts in microbiome abundances between disease and treatment conditions, broadening our understanding of the disease and ultimately improving clinical diagnosis and treatment.

KEYWORDS
microbiome data, relative abundances, high-dimensional compositional, differential abundance analysis, sparse alternatives
1 Introduction

The human microbiota, a collection of microbes living on or inside human bodies, has been shown to play a fundamental role in human health and diseases, including diabetes, cancer, and obesity (Turnbaugh et al. 2007; Ursell et al. 2012). Recently, the metagenomic next-generation sequencing (mNGS) technique has been introduced in the clinical diagnosis of infectious diseases (Gu et al. 2019; Dulanto Chiang and Dekker 2020; Govender et al. 2021) and emerged as a revolutionary technique to replace/supplement traditional culture-based and molecular microbiologic techniques: i) mNGS allows the parallel sequencing of hundreds of samples per run; ii) it provides an unbiased detection of bacteria, viruses, fungi, and parasites collectively; iii) this culture-free technology enables the identification of new species and others.

In microbiome studies, it is of general research interest to study the microbiome profiles/features between different disease treatments or conditions. Various statistical methods have been proposed recently for examining differential abundances (DAs) (Anderson 2014; Cao et al. 2018; Zhao et al. 2018; Banerjee et al. 2019; Lin and Peddada 2020). These methods can be categorized into univariate and multivariate approaches depending on whether microbial features are analyzed individually or in a set-based fashion. For example, Lin and Peddada (2020) proposed ANCOM-BC under a linear regression framework to conduct DA analysis for the assessed taxa individually. DESeq2 (Love et al. 2014) and edgeR (Robinson et al. 2010), two popular differential expression gene analysis methods, are commonly used for differential abundance analysis. However, multiple comparison procedures need to be conducted afterward for these univariate methods, which largely hinders the statistical power (Hu et al. 2018). Alternatively, we can assess the microbial features as a set in order to enhance the statistical power. Typically, the microbial abundances are normalized toward the total counts to make the microbial proportions [or called relative abundances (RAs)] comparable between samples. The normalized data have a summation of the features equal to one, termed compositional in microbiome studies (Mandal et al. 2015; Gloor et al. 2017). Directly applying standard multivariate statistical methods developed for unconstrained data to compositional data may result in inappropriate or misleading inferences. Cao et al. (2018) proposed a two-sample test for assessing the difference between two high-dimensional microbial composition matrices and treating all microbial features (the microbiome profile) as a set. Banerjee et al. (2019) proposed an adaptive test for comparing microbiome compositions from two independent groups. Zhao et al. (2018) developed a generalized Hotelling test for paired microbiome composition data comparison. These methods can be applied to the full microbiome profiles and also microbial features that belong to the same upper-level taxonomic rank, gene family, or functional pathway.

Nevertheless, they either need a strong assumption that the covariance matrices of compared compositions are equal (Cao et al. 2018) or require time-consuming permutations to determine the statistical significance (Zhao et al. 2018; Banerjee et al. 2019).

To address this challenge, we propose a two-sample maximum-type equal-covariance-assumption-free (MECAF) test. This multivariate differential abundance test statistics relaxes the equal covariance assumption required by the test proposed by Cao et al. (2018). The closed-form formula of the asymptotic null distribution largely resolves the computational burden in microbiome analysis. The method can be applied to analyze both taxonomic and functional profiles including microbial taxa (operational taxonomic units (OTUs), strains, etc., from either shotgun metagenomic or 16S rRNA amplicon sequencing technique), functional pathways, and gene families. The performance of the proposed MECAF test is demonstrated through simulation studies and applications to the shotgun metagenome sequencing study of Clostridium difficile infection (CDI) (Vincent et al. 2016) and the 16S rRNA amplicon murine microbiome study of type I diabetes (T1D) (Livanos et al. 2016).

The rest of this article is as follows. In Section 2, we briefly introduce the novel test statistics MECAF for conducting a two-group comparison of microbiome compositions, carry out extensive simulations to estimate the empirical type I error rate and statistical power for the proposed test in comparison with competing methods, and further conduct two real data applications. We conclude with a discussion in Section 3. Notation, test hypothesis, and the asymptotic properties of the MECAF test are given in the last section. All the theoretical derivations are detailed in the Supplementary Material.

2 Results

2.1 The MECAF test

We consider the comparison of high-dimensional microbiome compositions from two independent groups. We propose an independent two-sample test named MECAF, which 1) is derived based on the centered log-ratio (CLR) transformed compositions, 2) has the aim to test the null hypothesis of equal mean vectors for the microbial features against unequal mean vectors, and 3) does not require the assumption of equal covariance matrices between groups. The equation of the test statistics and corresponding asymptotic null distribution is given in Section 4.

2.2 Simulation studies

2.2.1 Simulation setup

We conducted extensive simulations to evaluate the numerical performance of the proposed MECAF test compared with
that covariance matrices for the CLR of compositions are equal, i.e., \( \Sigma^1 = \Sigma^2 \). Cao et al. (2018) proposed a test for hypothesis (1) as \( t_{\text{cao}} = \max \{ \frac{\Xi^{i} - \bar{X}}{\sigma_{\Xi^{i}}} \} \), where \( \sigma_{\Xi^{i}} = \frac{1}{N} \sum_{i=1}^{N} (\Xi^{i} - \bar{X})^2; \) henceforth, this is a maximum-type test with an equal covariance assumption, we denote it as the MEC test in this article. In addition, we also assessed the performance of the MEC statistics applied to the raw RA, the log transformation of RA, and the original AA. These three tests are obtained by replacing the CLR data used by MEC, i.e., \( \frac{a_{X^i}}{n^i} \), with \( R^a \), \( \log \left( R^a \right) \), and \( A^a \), respectively. MEC-Oracle is considered the benchmark method in the simulation study (under equal covariance matrices assumption), as the true difference is simulated for the log-absolute abundances. 

**Permutational multivariate analysis of variance (PERMANOVA)** is a popular multivariate analysis method widely adopted in community-level microbiome data analysis (Anderson (2014)). We therefore included PERMANOVA, which tests the null hypothesis that the centroid and the spread of the microbiome profiles are equivalent for the compared groups.

We set the sample size in the first group as \( n_1 = 100 \) and increased the sample size in the second group \( n_2 \) from 200 to 300. We increased the number of components (taxa) \( p \) from 100 to 300 to demonstrate different relationships between \( n = n_1 + n_2 \) and \( p \). We set the significance level as \( \alpha = 0.05 \) in the simulation, and 1,000 replications were conducted to evaluate the empirical type I error rate and statistical power of the assessed methods under various settings.

### 2.2.3 Simulation results

Figure 1 shows the numerical performance of assessed methods under Scenario 1, where unequal covariance matrices are considered. All competing methods that require equal covariance matrix assumption, i.e., MEC-Oracle, MEC-Log, MEC-Raw, and MEC, have inflated type I error rates. The type I error rate of MEC approaches 0.25 when \( p = 150 \) and \( p = 200 \). This indicates that MEC-type tests are not applicable to data with unequal matrices. In comparison, the proposed MECAF test can control the empirical type I error rate around the nominal level of 0.05. The statistical power of MECAF increases with the proportion of signal taxa. The results of PERMANOVA are not shown in Figure 1, since the corresponding type I error rate and statistical power are all equal to 1 under this scenario. This is because the abundance data were generated from unequal covariance matrices and therefore violate the null hypothesis tested by PERMANOVA.

Simulation results for equal banded covariance and equal sparse covariance scenarios are depicted in Figures 2 and 3, respectively. As expected, all assessed methods have a well-controlled type I error rate under these simulation settings. MECAF and MEC achieved statistical power comparable to

**2.2.2 Competing methods**

In this article, we mainly focus on the comparisons of multivariate differential abundance approaches. By assuming

• Specification of location parameters \( \mu^1 \) and \( \mu^2 \). Following Cao et al. (2018), the components of \( \mu^i \) were drawn from the uniform distribution Uniform \((0,10)\). Each component of \( \mu^i \) was set by \( \mu^i_{\text{loc}} = \mu^i_{\text{loc}} - \delta \sigma_{\text{loc}}^2 \). Here, \( \delta \) represents the signal, i.e., the difference in CLR means for component \( j \) between two groups. \( \sigma_{\text{loc}}^2 = 0.05p_{\text{loc}} \), where \( p_{\text{loc}} = 0.1p \), \( 0.2p \), and \( 0.5p \). The other \( \sigma_{\text{loc}}^2 \)’s were set as 0. We can see that \( s = 0.05p_{\text{loc}} \) corresponds to the null hypothesis setting and \( s = 0.1p_{\text{loc}} \), \( 0.2p_{\text{loc}} \), and \( 0.5p_{\text{loc}} \) represent three alternative hypothesis settings. When \( s \) becomes larger, there are more signal taxa in the microbiome compositions. \( \sigma_{\text{loc}}^{\text{loc}} \) is the \( j \)-th diagonal component of the covariance matrix \( \Sigma^i_{\text{loc}} \) with specifications as follows.

• Specification of covariance matrices \( \Sigma^1_{\text{loc}} \) and \( \Sigma^2_{\text{loc}} \). We included two types of covariance matrices, i.e., a banded covariance matrix \( \Sigma_{\text{b}} \) and sparse covariance matrix \( \Sigma_{\text{s}} \) with the same parameters as those in Cao et al. (2018). Three scenarios were considered to assess the impact of equal vs. unequal covariance matrices between two groups in the comparison of compositional mean vectors. Specifically, in Scenario 1, the covariance matrices of groups 1 and 2 are set as \( \Sigma^1_{\text{loc}} = \Sigma_{\text{b}} \) and \( \Sigma^2_{\text{loc}} = \Sigma_{\text{s}} \) to represent the setting with unequal covariance matrices between groups; equal banded covariance matrices were considered in Scenario 2, i.e., \( \Sigma^1_{\text{loc}} = \Sigma^2_{\text{loc}} = \Sigma_{\text{b}} \); and equal sparse covariance matrices were considered in Scenario 3, i.e., \( \Sigma^1_{\text{loc}} = \Sigma^2_{\text{loc}} = \Sigma_{\text{s}} \).
that of MEC-Oracle, with sparsity measure $s$ ranging from 0.05$p$ to 0.2$p$. This indicates the statistical efficiency of the MECAF test. In comparison, MEC-Log, MEC-Raw, and PERMANOVA have evidently smaller power than MEC-Oracle for all settings of the two scenarios.

In summary, MECAF has a well-controlled type I error rate for two group comparisons of mean composition vectors either with equal or unequal covariance matrices. The statistical power is desirable under all scenarios with various sparsity measures.

### 2.3 Applications to two microbiome studies

Here, we first apply the proposed MECAF test to the shotgun metagenomic sequencing data from the CDI study (Vincent et al. (2016)). Since the test is also applicable to microbiome abundance data generated from the 16S rRNA amplicon sequencing technology, we further illustrate our proposed method in a murine microbiome study of T1D [Livanos et al. (2016)].
2.3.1 Analysis of the *Clostridium difficile* infection metagenomic dataset

Vincent et al. (2016) conducted a prospective study to investigate the intestinal microbiota dynamics over time among 98 hospitalized patients at risk for CDI, a leading infectious cause of nosocomial diarrhea. Patients were followed up to 60 days, and a total of \( N = 229 \) fecal samples (averaging 2.34 samples per subject) were examined by the shotgun metagenomics sequencing platform. The bioinformatics pre-processing steps were detailed by Vincent et al., 2016, and the processed microbial counts and metadata are available in the R package "curatedMetagenomicsData" (Version 1.16.1) from the Bioconductor by running the function curatedMetagenomicsData(VincentC_2016.metaphlan_bugs_list.stool,dryrun = FALSE) (Pasolli et al. (2017)). Zero counts were imputed with 0.5 before converting counts to relative abundances for taxa from taxonomic ranks of phylum, class, order, family, genus, and species (strains). In this secondary data analysis, we aim to examine whether there are shifts in the microbial relative abundances between patients with CDI or asymptomatic *C. difficile* colonization (CDI group, \( N = 8 \) subjects) and patients without (control group, \( N = 90 \) subjects) i) upon hospitalization (baseline), ii) at 1 week of hospitalization, and iii) over 1 week of hospitalization. The latest sample at each time window was included for each patient.

Figure 4 shows the available samples at each assessed time window (Figure 4A), the number of taxa observed at each taxonomic rank (Figure 4B), the differential abundance test results of the MECAF test, and competing methods (Figures 4C). A p-value of <0.05 was indicated as statistically significant. Of note, MEC-Oracle was not included in the comparison since the absolute abundance data required by...
MEC-Oracle are unobservable in real data. The results of MECAF indicate significant differences in the microbiome compositions between CDI and control patients at baseline and week 1. The significance is consistent for most of the taxonomic ranks, and a stronger signal is depicted at the lower ranks. The difference though seems to disappear after 1 week of hospitalization, where only the test at the species (strain) level is significant. In comparison, MEC does not detect significant differences in microbiome compositions except at the species and strain levels at baseline and over 1 week, with less stringent p-values reported. At week 1, MEC detects microbial composition differences at the family, genus, and species (strain) levels but not at the phylum, class, or order level. MEC-Raw and PERMANOVA have similar results as MEC, and MEC-log reported similar results as those of MECAF with higher p-values. In summary, we observe more consistent findings from the MECAF test over six taxonomic ranks. The corresponding p-values are in general smaller than those of competing methods, indicating statistical efficiency gain over other methods.

We further assessed the type I error rate of competing methods using the baseline data of the control subjects (N = 90). To achieve this, we randomly split the dataset into two groups and conducted DA tests between the mock groups. A total of 1,000 replications were carried out to calculate the empirical type I error rate as shown in Table 1. As expected, all assessed methods are able to control the type I error rate below the nominal level of 0.05.
2.3.2 Analysis of the MICE 16S rRNA amplicon microbiome data

Livanos et al. (2016) carried out a murine microbiome study to investigate the effect of early-life antibiotic exposure on the alteration of gut microbiota composition. Here, we re-examined the 16S microbiome abundance profile from the early-life sub-therapeutic antibiotic treatment (STAT) group and the control group that received no antibiotic exposure. The abundances were compared between the two groups at each of the four assessed time points, i.e., weeks 3, 6, 10, and 13, for female and male mice using the MECAF test and competing methods.

The available samples and number of taxa observed are shown in Figures 5A, B, which illustrate the circumstance of \( n < p \) (number of samples < number of taxa) most often encountered in microbiome data analysis. The differential abundance analysis results from the phylum to genus rank are depicted in Figures 5C, 6 for female and male mice, respectively.

| Method       | Taxonomic rank | Strains |
|--------------|----------------|---------|
|              | Phylum | Class | Order | Family | Genus | Strains |
| MECAF        | 0.033  | 0.045 | 0.038 | 0.032  | 0.023 | 0.016   |
| MEC          | 0.032  | 0.044 | 0.038 | 0.032  | 0.023 | 0.016   |
| MEC-Log      | 0.029  | 0.041 | 0.031 | 0.035  | 0.021 | 0.013   |
| MEC-Raw      | 0.011  | 0.013 | 0.002 | 0.003  | 0.003 | 0.000   |
| PERMANOVA    | 0.049  | 0.056 | 0.046 | 0.051  | 0.052 | 0.048   |

CDI, *Clostridium difficile* infection.
result from MECAF indicated that in female mice, the abundance profile is significantly different in the STAT group from the control group from week 3 to 6 for almost all taxonomic ranks. The significance is weaker at weeks 10 and 13, indicating the recovery of gut microbiota in the STAT mice upon maturation. In comparison, a significant difference is detected by MECAF in male mice over the four assessed time points. This result is consistent with Livanos et al. (2016) in which the alpha- and beta-diversity measures were compared between groups over time. MEC has similar results to MECAF. MEC-Log did not detect significance in female mice from weeks 6 to 10 for most of the taxonomic ranks. MEC-Raw and PERMANOVA either did not detect significant differences (female mice from weeks 3 to 10) or reported weaker signals (male mice, weeks 3, 10, and 13).

3 Discussion

In this article, we propose a novel test named MECAF for the two-sample test of high-dimensional compositions. The test statistics is developed based on the centered log-ratio transformation of the compositions following Aitchison (1982) and Cao et al. (2018). The asymptotic null distribution of the test
statistic is derived, and the power against sparse alternatives is investigated. The derived null distribution allows for the closed-form solution of statistical significance and largely resolves the computational burden. Simulation results show that the proposed method is evidently more powerful than competing methods when the covariance matrices differ between groups, and comparable performance is achieved when the groups have equal covariance. Two real data applications have illustrated the usefulness of the proposed method. The comparisons of competing methods have focused on multivariate approaches only since they are not directly comparable with univariate approaches (such as ANCOM-BC, DESeq2, and edgeR). We admit the limitation of the MECAF test that it is used as the first screening step of microbiome analysis for the examination of the global shift of microbiome profiles. Other regression models that are built upon Dirichlet distribution or generalized Dirichlet distribution [Tang and Chen (2019); Liu et al. (2020)] have distinct features from differential abundance methods discussed herein. For example, they allow for covariate adjustment, feature selections, repeated sampling, etc., which is beyond the scope of this article.

The MECAF test extends the MEC proposed by Cao et al. (2018) by relaxing the assumption of equal covariance matrix structure between groups. Therefore, MECAF can be applied to a wider set of circumstances. In the real data applications, we applied MECAF to compare the microbiome abundances aggregated to each taxonomic rank. In practice, we can also apply MECAF to assess the composition of a given sub-tree or a subset of the microbiome taxa (Shi and Li (2017)). As a future direction, we will aim to extend the MECAF test to accommodate repeated measures from each individual for group comparisons.

4 Methods

4.1 Notation and specification of test hypothesis

In this article, we consider microbiome compositions from two independent groups. The notation used in this manuscript is summarized in Table 2. Specifically, for subject i
from group $g (g = 1, 2)$, denote the $n_g$ independently observed composition vectors as $\{ R_g^i = (R_g^{i,1}, \ldots, R_g^{i,n_g})^T, i = 1, 2, \ldots, n_g \}$ with length of $p$, and the $j$th component (taxon) of the vector $R_g^i$ as $R_{g_{ij}}$, where $R_{g_{ij}} \in (0, 1)$. Of note, zero proportions are imputed by a pseudo-positive proportion prior to conducting the analysis.

Then the compositional constraints can be expressed as $\sum_{p=1}^{R_{g_{ij}} = 1, i = 1, \ldots, n_g; g = 1, 2}$. Obviously, $R_g^i$ represents compositions that lie in the $p - 1$ dimensional simplex $\{ S_{p-1} = \{ r_1, \ldots, r_p \}; r_j > 0, j = 1, \ldots, p, \sum_{p=1}^{\sqrt{p}} r_j = 1 \}$. $R_1$ and $R_2$ are the observed data matrices of dimension $n_1 \times p$ and $n_2 \times p$, respectively, from the two groups. Let $\{ A_g^i = (A_g^{i,1}, \ldots, A_g^{i,n_g})^T, i = 1, \ldots, n_g \}$, $g = 1, 2 \}$ denote the the $n_g$ unobserved absolute abundances of the microbiome. The numerical relationship between the absolute abundance matrix and composition matrix is as follows:

$$
R_g^i = \sum_{p=1}^{A_g^i} A_g^{i,j}, i = 1, \ldots, n_g; j = 1, \ldots, p; g = 1, 2,
$$

where $A_g^{i,j}$ is the $j$th component of $A_g^i$. Suppose the log transformations of $A_g^i$, denoted by $L_g^i$, are i.i.d. from distributions with mean vectors $\mu_g^i = (\mu_g^{i,1}, \ldots, \mu_g^{i,p})^T = E[L_g^i]$ and covariance matrices $\Sigma_g^i = (\sigma_g^{i,ij})_{k=1 \ldots p} = cov(L_g^i, L_g^j)$, $g = 1, 2$. Cao et al. (2018) introduced a testable hypothesis to compare the log-mean absolute abundance vectors through the observed compositional data $R_1$ and $R_2$ by exploiting the CLR transformation of the compositions. Denote the CLR transformation of $R_g^i$ by

$$
X_g^i = \log \left( \frac{R_g^i}{(I_{p-1} \cdot R_g^i)^{1/2}} \right), i = 1, \ldots, n_g; j = 1, \ldots, p; g = 1, 2.
$$

Assume that the CLR vectors $\{ X_g^i = (X_g^{i,1}, \ldots, X_g^{i,p})^T, i = 1, 2, \ldots, n_g \}$ are i.i.d. from distributions with mean vectors $\mu_g^i = (\mu_g^{i,1}, \ldots, \mu_g^{i,p})^T = E[X_g^i]$, and covariance matrices $\Sigma_g^i = (\sigma_g^{i,ij})_{k=1 \ldots p} = cov(X_g^i, X_g^j)$, $g = 1, 2$. Then the testable hypothesis under the definition of compositional equivalence as follows. It is straightforward that $\sum_{p=1}^{\sqrt{p}} X_g^{i,j} = 0$, and $\sum_{p=1}^{\sqrt{p}} \mu_g^{i,j} = 0$ holds if and only if for $j \in \{ 1, \ldots, p-1 \}$, as $\mu_g^{i,j} = 0$. Therefore, an equivalent hypothesis that only considers the first $p - 1$ components is

$$
H_0(1): \mu_g^i = \mu_g^j \quad \text{versus} \quad H_1(1): \mu_g^i \neq \mu_g^j.
$$

In this work, we consider another testable hypothesis of compositional equivalence as follows. It is straightforward that $\sum_{p=1}^{\sqrt{p}} X_g^{i,j} = 0$, and $\sum_{p=1}^{\sqrt{p}} \mu_g^{i,j} = 0$ holds if and only if for $j \in \{ 1, \ldots, p-1 \}$, as $\mu_g^{i,j} = 0$. Therefore, an equivalent hypothesis that only considers the first $p - 1$ components is

$$
H_0(2): \mu_g^{i,j} = \mu_g^{j,j} \quad \text{for any} \ j \quad \text{versus} \quad H_1(2): \mu_g^{i,j} \neq \mu_g^{j,j} \quad \text{for at least one} \ j \in \{ 1, \ldots, p-1 \}.
$$

In the following, we will introduce our proposed test specifically for hypothesis 2. We also investigate the theoretical properties of the test statistics.

4.2 The proposed MECAF test

Cao et al. (2018) proposed one maximum-type two-sample test for high-dimensional compositions by assuming that the covariance matrices of two groups are equal (see equation 9 of Cao et al. (2018)). In practice, it is unable to assess the compositional equivalence as follows. It is straightforward that $\sum_{p=1}^{\sqrt{p}} X_g^{i,j} = 0$, and $\sum_{p=1}^{\sqrt{p}} \mu_g^{i,j} = 0$ holds if and only if for $j \in \{ 1, \ldots, p-1 \}$, as $\mu_g^{i,j} = 0$. Therefore, an equivalent hypothesis that only considers the first $p - 1$ components is

$$
H_0(2): \mu_g^{i,j} = \mu_g^{j,j} \quad \text{for any} \ j \quad \text{versus} \quad H_1(2): \mu_g^{i,j} \neq \mu_g^{j,j} \quad \text{for at least one} \ j \in \{ 1, \ldots, p-1 \}.
$$

In the following, we will introduce our proposed test specifically for hypothesis 2. We also investigate the theoretical properties of the test statistics.

### TABLE 2 Notation summary.

| Notation | Description |
|----------|-------------|
| $g$      | Group indicator, $g = 1, 2$ |
| $n_g$    | Sample size for group $g$ |
| $i$      | $i$th sample, $i = 1, 2, \ldots, n_g$ for group 1, and $i = 1, 2, \ldots, n_g$ for group 2 |
| $p$      | Number of taxa in the microbiome data matrix |
| $R_g$    | Observed microbial relative abundances (RAs) for group $g$ with dimension $n_g \times p$ |
| $R_g^i$  | Observed RA for subject $i$ for group $g$ |
| $A_g^i$  | Unobserved microbial absolute abundances (AAs) for group $g$ with dimension $n_g \times p$ |
| $L_g^i$  | Unobserved log-AA for subject $i$ for group $g$ with dimension $n_g \times p$ |
| $\Sigma_g^i$ | Unobserved centered log-ratio transformation of absolute abundances of group $g$ with dimension $n_g \times p$ |
| $X_g^i$  | Observed centered log-ratio (CLR) transformation of RA for subject $i$ from group $g$, and $X_{g_{ij}}$ i.d. from distribution with mean $\mu_g^{i,j}$ and covariance matrix $\Sigma_g^{i,j}$ |
| $X_g^{i,j}$, and $\Sigma_g^{i,j}$ | Observed centered log-ratio (CLR) transformation of RA for subject $i$ from group $g$, and $X_{g_{ij}}$ i.d. from distribution with mean $\mu_g^{i,j}$ and covariance matrix $\Sigma_g^{i,j}$ |
and $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ is the average of the CLR transformation of relative abundances. Our proposed test is given as follows:

$$T_{MECAF} = \frac{T_{CLL}}{\sqrt{\frac{\sum_{j=1}^{k} \bar{X}_{ij} - \bar{X}_i}{\hat{\sigma}_{\bar{X}_{ij}}}}},$$

where $\hat{\sigma}_{\bar{X}_{ij}} = \frac{m + n_i}{m} \bar{\sigma}_{\bar{X}_{ij}} + \frac{n_i + n_j}{m} \bar{\sigma}_{\bar{X}_{ij}}$ and $\hat{\sigma}_{\bar{X}_{ij}} = \frac{n_i}{m} \sum (X_{ij} - \bar{X}_{ij})^2$, $g = 1, 2$. We name it the MECAF test. As a maximum-type test statistic, it is in general better than sum-of-squares type statistics under sparse alternatives (Tony Cai et al. (2014)). The test statistic, it is in general better than sum-of-squares type statistics under sparse alternatives (Tony Cai et al. (2014)). The assumption of the equal high-dimensional covariance matrices for two groups is relaxed to allow for wider applicable conditions.

We successfully derived the asymptotic null distribution of $T_{MECAF}$ given by

$$\Pr\left(2 - (\log (p - 1))^{1/2} \left| T_{MECAF} - \left( h_0 + \log 4 + \frac{\log 4}{2 \log (p - 1)} \right) \right| < t \right) \to \exp(-\exp(-t)),$$

for any real number $t$ as $n_1, n_2 \to \infty$, where $h_0 = 2 \log (p - 1) - \left[ \log (\log (p - 1)) + \log (4\pi) \right] + \log (\log (p - 1)) + \log (4\pi)$. Denote $q_\alpha$ as the $(1 - \alpha)$-quantile of the derived distribution function $\exp(-\exp(-t))$. Namely, $q_\alpha = -\log (1 - \alpha)^{1/2}$. We can define an asymptotic $\alpha$-level test denoted by

$$\Phi_{1, \alpha} = 1\{T_{MECAF} > \left( 2 - \frac{1}{\log (p - 1)} \right) q_\alpha + h_0 + \log 4 - \frac{\log 4}{2 \log (p - 1)} \}.$$

The null hypothesis $H_0^{(2)}$ is rejected whenever $\Phi_{1, \alpha} = 1$. We also prove that the power of test $\Pr(\Phi_{1, \alpha} = 1)$ converges to 1 under some settings and $H_1^{(2)}$ as $n_1, n_2 \to \infty$. All the detailed proof is given in the Supplementary Material.

**Data availability statement**

The metagenomics abundance data of the CDI study is readily available through the R package “curatedMetagenomicData” (Version 1.16.1) from the Bioconductor (https://bioconductor.org/packages/release/data/experiment/html/curatedMetagenomicData.html). The 16S rRNA amplicon sequencing data from the murine T1D study is publicly available at EBI with accession number ERP016357.

**Ethics statement**

No ethics approval or consent to participate was required for this study.

**Author contributions**

ZL developed the proposed method and performed theoretical proof and simulation studies and manuscript writing. XY performed simulation and real data analyses and manuscript writing. HG performed theoretical proof and manuscript writing. TL performed simulation analyses. JH conceptualized the ideas for the proposed method, simulations, and real data analyses and performed manuscript writing. All authors contributed to the article and approved the submitted version.

**Funding**

HG is funded by the Young Talents Project of Scientific Research Plan of the Hubei Provincial Department of Education (Grant No. Q20212506). JH is partly supported by NIH National Institute on Minority Health and Health Disparities under Award Number U54MD000538, and NIH National Institute on Aging under Award Number R33AG057382.

**Acknowledgments**

The authors would like to thank the reviewers and editors for their valuable comments and suggestions.

**Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s note**

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Supplementary material**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2022.988717/full#supplementary-material
References

Aitchison, J. (1982). The statistical analysis of compositional data. J. R. Stat. Soc. Ser. B Methodol. 44, 139–160. doi: 10.1111/j.2517-6161.1982.tb01195.x

Anderson, M. J. (2014). Permutational multivariate analysis of variance (permanova) (Wiley statsref: statistics reference online). I–15.

Banerjee, K., Zhao, N., Srinivasan, A., Xue, I., Hicks, S. D., Middleton, F. A., et al. (2019). An adaptive multivariate two-sample test with application to microbiome differential abundance analysis. Front. Genet. 10, 350. doi: 10.3389/fgene.2019.00350

Cao, Y., Lin, W., and Li, H. (2018). Two-sample tests of high-dimensional means for compositional data. Biometrika 105, 115–132. doi: 10.1093/biomet/asy060

Dulanto Chiang, A., and Dekker, J. P. (2020). From the pipeline to the bedside: advances and challenges in clinical metagenomics. J. Infect. Dis. 221, S331–S340. doi: 10.1093/infdis/jiu151

Gloor, G. B., Macklaim, J. M., Pavolowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. Front. Microbiol. 8, 2224. doi: 10.3389/fmicb.2017.02224

Govender, K. N., Street, T. L., Sanderson, N. D., and Eyre, D. W. (2021). Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. J. Clin. Microbiol. 59, e02916–e02920. doi: 10.1128/JCM.02916-20

Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. Annu. Rev. Pathol. 14, 319. doi: 10.1146/annurev-pathmechdis-021418-012751

Hu, J., Koh, H., He, L., Liu, M., Blaser, M. J., and Li, H. (2018). A two-stage microbial association mapping framework with advanced fdr control. Microbiome 6, 1–16. doi: 10.1186/s40168-018-0517-1

Lin, H., and Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. Nat. Commun. 11, 1–11. doi: 10.1038/s41467-020-17041-7

Liu, T., Zhao, H., and Wang, T. (2020). An empirical bayes approach to normalization and differential abundance testing for microbiome data. BMC Bioinf. 21, 1–18. doi: 10.1186/s12859-020-03352-z

Livanos, A. E., Greiner, T. U., Vangay, P., Pathmasiri, W., Stewart, D., McRitchie, S., et al. (2016). Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. Nat. Microbiol. 1, 1–13. doi: 10.1038/nmicrobiol.2016.140

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome Biol. 15, 1–21. doi: 10.1186/s13059-014-0550-8

Mandal, S., Van Treuren, W., White, R. A., Eggesbo, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. Microbial Ecol. Health Dis. 26, 27663. doi: 10.3402/mehd.v26.27663

Pesoli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D. T., et al. (2017). Accessible, curated metagenomic data through experimenthub. Nat. Methods 14, 1023–1024. doi: 10.1038/nmeth.4468

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). Edger: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. doi: 10.1093/bioinformatics/btp16

Shi, P., and Li, H. (2017). A model for paired-multinomial data and its application to analysis of data on a taxonomic tree. Biometrics 73, 1266–1278. doi: 10.1111/biom.12661

Tang, Z.-Z., and Chen, G. (2019). Zero-inflated generalized dirichlet multinomial regression model for microbiome compositional data analysis. Biostatistics 20, 698–713. doi: 10.1093/biostatistics/kxy025

Tony Cai, T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. J. R. Stat. Soc. Ser. B Stat Methodol 76, 349–372. doi: 10.1002/jrsm.12034

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. Nature 449, 804–810. doi: 10.1038/nature06244

Ursell, L. K., Metcalfe, J. L., Parfrey, L. W., and Knight, R. (2012). Defining the human microbiome. Nutr. Rev. 70, 538–544. doi: 10.1111/j.1753-4887.2012.00493.x

Vincent, C., Miller, M. A., Edens, T. J., Mehrrota, S., Dewar, K., and Manges, A. R. (2016). Bloom and bust: intestinal microbiota dynamics in response to hospital exposures and coidtruum difficult colonization or infection. Microbiome 4, 1–11. doi: 10.1186/s40168-016-0156-3

Zhao, N., Zhan, X., Guthrie, K. A., Mitchell, C. M., and Larson, J. (2018). Generalized hotelling’s test for paired compositional data with application to human microbiome studies. Genet. Epidemiol. 42, 459–469. doi: 10.1002/gepi.22127