BMC Genomics

METHODOLOGY ARTICLE   **Open Access**

Microbial trend analysis for common dynamic trend, group comparison, and classification in longitudinal microbiome study

Chan Wang[1], Jiyuan Hu[1], Martin J. Blaser[2] and Huilin Li[1]*

Abstract

Background: The human microbiome is inherently dynamic and its dynamic nature plays a critical role in maintaining health and driving disease. With an increasing number of longitudinal microbiome studies, scientists are eager to learn the comprehensive characterization of microbial dynamics and their implications to the health and disease-related phenotypes. However, due to the challenging structure of longitudinal microbiome data, few analytic methods are available to characterize the microbial dynamics over time.

Results: We propose a microbial trend analysis (MTA) framework for the high-dimensional and phylogenetically-based longitudinal microbiome data. In particular, MTA can perform three tasks: 1) capture the common microbial dynamic trends for a group of subjects at the community level and identify the dominant taxa; 2) examine whether or not the microbial overall dynamic trends are significantly different between groups; 3) classify an individual subject based on its longitudinal microbial profiling. Our extensive simulations demonstrate that the proposed MTA framework is robust and powerful in hypothesis testing, taxon identification, and subject classification. Our real data analyses further illustrate the utility of MTA through a longitudinal study in mice.

Conclusions: The proposed MTA framework is an attractive and effective tool in investigating dynamic microbial pattern from longitudinal microbiome studies.

Keywords: Composition, Classification, Dynamic, High dimensionality, Hypothesis testing, Longitudinal microbiome, Phylogenetic tree, Variable selection

Background

The human microbiota represents a complex and rich ecosystem of over 100 trillion microbial cells, playing a fundamental role in maintaining health and driving disease [1, 2]. Considering that the human microbiota is inherently dynamic and can be substantially altered by many factors at any time point either temporally or permanently, recent microbiome studies [2–8] have shifted their research design from cross-sectional or case-control studies to longitudinal analyses. Rather than untangling the associations between microbes and a wide range of diseases at a fixed time point [2, 9–13], longitudinal microbiome studies target understanding how the dynamic changes in microbiome are linked with disease susceptibility. For example, the Integrative Human Microbiome Project [2, 3] was designed to comprehensively characterize the dynamic changes in human microbiome in three disease-specific cohorts: pregnancy and preterm birth, onset of inflammatory bowel disease (IBD), and onset of type 2 diabetes. Thaiss (2018) [5] found that

*Correspondence: Huilin.Li@nyulangone.org
[1]Division of Biostatistics, Department of Population Health, New York University School of Medicine, 10016 New York, NY, USA
Full list of author information is available at the end of the article

the microbiome underwent oscillations in composition, functional activity, and localization over the course of a day and these dynamic patterns were aberrant in obesity. Lloyd-Price et al. (2019) [6] reported that periods of IBD disease activity were distinguished by increases in temporal variability of gut microbiome, with taxonomic, functional, and biochemical shifts of microbiota. Such scientific results provide insights into the characterization of the microbial dynamics and raise further questions about understanding these underlying microbial dynamics as well. Do the microbial dynamics significantly relate or contribute to the group differentials? If so, which specific microbes dominate them? Can subjects be classified based on their microbial dynamics?

To address such questions, recent efforts have focused on analyzing the longitudinal microbiome data. Several parametric methods have been proposed to elucidate the microbial dynamic changes, including mixed-effects model [14–16], generalized Lotka-Volterra equations [17, 18], time series models [19–21], and state-space models [22, 23]. While those methods provide capabilities to capture the microbial dynamics and identify the time-dependent taxa, they assume that the microbial abundance changes at a fixed rate, in the autoregressive or autoregressive integrated moving average pattern, or following the Markov process, which cannot always be justified and implemented because of the limited sample size. In contrast, spline-based approaches [24, 25] are proposed to examine whether one taxon's changing pattern over time or within a time interval is significantly different between two groups or not with permutation test. MetaLonDA [24] models the mapped read counts by a negative binomial (NB) distribution and fits the longitudinal profiles in each phenotypic group with NB smoothing splines. While a nonparametric approach Permuspliner [25] uses the loess spline for the relative abundance. However, since these methods need to do the modeling or testing taxon by taxon, the large number of taxa can inevitably affect the statistical power after the multiple testing corrections, even at high taxonomic ranks.

To support the comprehensive analyses in the longitudinal microbiome data, we propose a microbial trend analysis (MTA) as a complementary tool which can capture the overall community level microbial dynamic pattern for a group of subjects. In particular, following the path of the principal trend analysis (PTA) [26], MTA integrates the spline-based method for time-course data analysis and principal component analysis for dimension reduction to extract the dynamic patterns from a group of subjects. Matrix decomposition and lasso technique are used to address the high-dimensionality feature as in PTA, and the graph Laplacian penalty [27–30] is additionally used to incorpo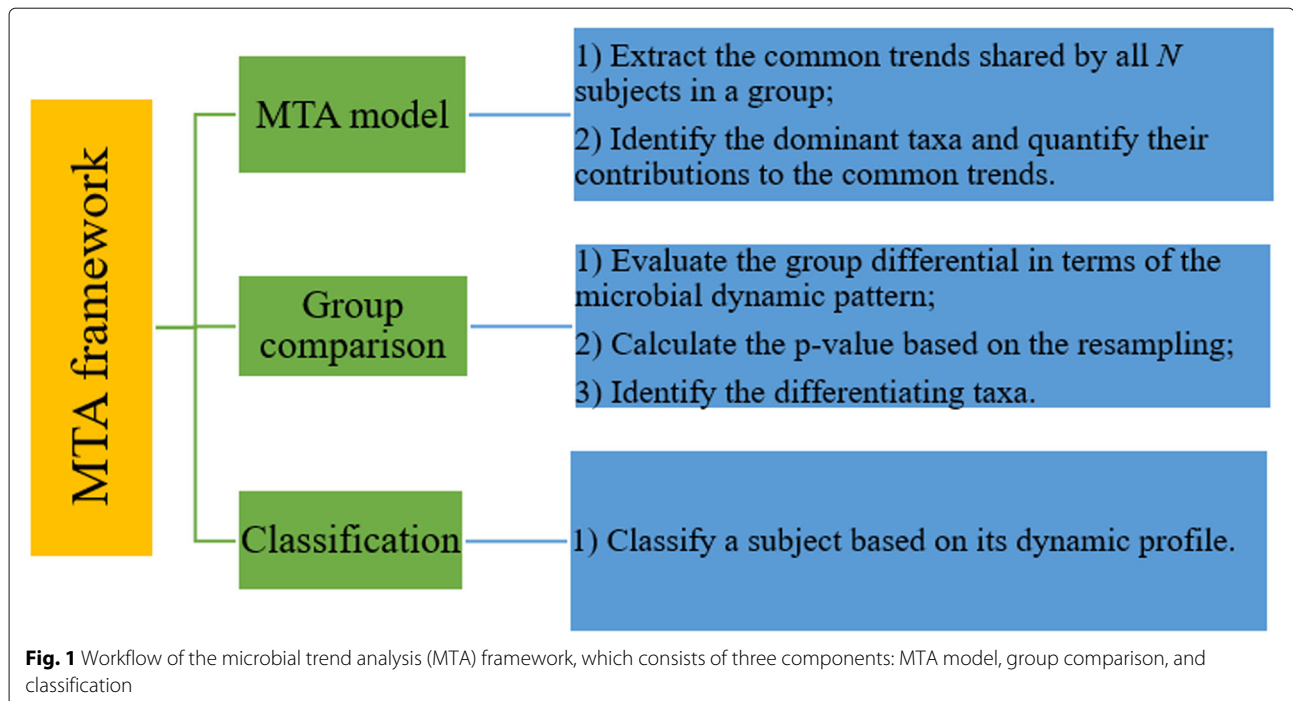rate phylogenetic tree structure, the unique feature of microbiome data, into the analysis. In combination, these help MTA to identify the dominant taxa that contribute to the common trends simultaneously. To make MTA hold practical value in the microbiome research, we further propose: 1) a microbial trend group differential test to confirm the statistical significance of the group comparison and identify the key taxa contributing to the group differential trend, and 2) a distance-based classification algorithm to assign a group label to a given subject.

Results

There are three components of the proposed MTA framework: MTA model, group comparison, and classification algorithm, based on the longitudinal microbiome data. Its workflow is illustrated in Fig. 1 and the detailed method is provided in the Methods section. Briefly, for N subjects in a given group with the relative abundance Y_{npt} for the nth subject of the pth taxon at the tth time point, $n = 1, \ldots, N$, $t = 1, \ldots, T$, and $p = 1, \ldots, P$, we first propose a nonlinear iterative algorithm (Algorithm 1, MTA model) to extract the common trends shared by all N subjects, identify the dominant taxa, and estimate their contributions to the common trends. Next, we propose a group differential test (Algorithm 2) to evaluate how different two groups are and quantify the differentiating taxa if the group difference is significant. At last, we propose a distance-based classification algorithm (Algorithm 3) to classify subjects based on their longitudinal microbial profiling. In the following, we explore the performance of MTA through the simulation study and illustrate the utility of MTA to investigate the relationship between antibiotic usage and the microbial dynamics using a longitudinal murine study.

Simulation results

We constructed extensive simulations to evaluate the performance of the proposed MTA method in both group comparison (Algorithm 2) and classification (Algorithm 3). Specifically, the performance of group comparison is evaluated in terms of the empirical type I error rate and statistical power at the group level, as well as in terms of sensitivity and specificity at the taxon level, compared to the competing methods Permuspliner [25], MetaLonDA [24], and LonGP [16]. Note that Permuspliner, MetaLonDA, and LonGP all do the analysis taxon by taxon. Permuspliner and MetaLonDA are strategically similar in terms of testing the group differential, but MetaLonDA has a poorer performance. Meanwhile, LonGP, as a Baysian method, provides the component relevance and corresponding selection with a given threshold based on the linear mixed model, rather than a testing p-value, and its sampling procedure to determine the prior parameters is time-consuming. Comparisons in Additional file 1:

Wang *et al. BMC Genomics* (2021) 22:667

Page 3 of 16

Fig. 1 Workflow of the microbial trend analysis (MTA) framework, which consists of three components: MTA model, group comparison, and classification

Section S6 show that LonGP does not have the competitive sensitivities either. We include only Permuspliner in this section and report all the other comparisons in Additional file 1: Section S6. Permuspliner tests the differences in microbial abundance between two groups over the longitudinal time course based on the loess splines [31, 32]. We evaluated the performance of classification in terms of receiver operating characteristic (ROC) curve and area under the curve (AUC) [33].

Simulation results for group comparison

Since the common trends extracted by Algorithm 2 manifest the difference between case and control groups and have a hierarchical structure as well, the overall empirical type I error rate and power for the proposed MTA method are defined as the proportion of the adjusted p-values for the first common trend less than the given significance level (usually 5%) with 1000 independent replications under the null and alternative hypotheses respectively. Since the competing method Permuspliner works at the taxon level and only provides the individual p-value for each taxon separately, its overall type I error rate and power are calculated as the proportion of at least one taxon that has a significant adjusted p-value with the Benjamini-Hochberg (BH) correction for multiple comparisons [34]. The number of resamplings in Algorithm 2 is set as $R = 50$.

Type I error rate and power. The first group of bars in each subfigure of Fig. 2 reports the empirical type I error rates of the proposed MTA method and the competing method Permuspliner. They are all around the nominal significance level 5%, except that when sample size $N = 20$ and the number of time points $T = 20$, both methods, especially Permuspliner, have slightly conservative type I error rates. Thus, both are statistically valid tests to differenciate microbial dynamic patterns between groups in the longitudinal study.

The scenarios 1-3 in each subfigure of Fig. 2 report the estimated powers of both methods with sample size $N = 20, 30$ and the number of time points $T = 10, 20$, respectively. The relative performance of these two methods is similar across scenarios 1-3 in all combinations of N and T. MTA is more powerful than Permuspliner in most scenarios, especially when the number of time points is large at $T = 20$. In addition, the power of MTA increases as N or T increases. Taking scenario 2 as an example, the estimated power of MTA increases from 37.1% with $N = 20$ and $T = 10$ to 66.0% with $N = 30$ and $T = 10$, and to 90.5% with $N = 30$ and $T = 20$. The number of time points T has an effect on the power of the competing method Permuspliner, since increasing time points from 10 to 20 diminishes its power dramatically in all scenarios. This is because Permuspliner is based on the area under the microbial splines across all time points to evaluate dynamic patterns. The area under the curve method is most effective when the temporal trend difference between the compared groups is relatively consistent along time, i.e. one trajectory is always above the other. However, when the microbial trends from two groups cross over, which is more likely to happen

Algorithm 1 Nonlinear iterative algorithm for MTA

Input: the observed data Y_n ($n = 1, \ldots, N$), the number of common trends M, and the fixed tuning parameters λ_1, λ_2 and λ_3

For $m = 1, 2, \ldots, M$

> **Initialize**: $Y_n^1 = Y_n, n = 1, \ldots, N$
>
> **For** $i = 1, 2, \ldots$
>
> > **Initialize**: $\hat{c}_m^{(1)} = 0$, and $\hat{f}_m^{(1)}$ with $\|\hat{f}_m^{(1)}\|_2 = 1$ as the first estimated factor score vector from the mean matrix $\frac{1}{N}\sum_{n=1}^{N} Y_n^m$.
> >
> > **Step 1**: $\hat{c}_m^{(i+1)} = \left(NB'B + \lambda_1\Omega\right)^{-1} B'\left(\sum_{n=1}^{N} Y_n^m\right)' \hat{f}_m^{(i)}$
> >
> > **Step 2**:
> > $$\hat{f}_m^{(i+1)} = \left[\text{Soft}\left(\hat{c}_m^{(i+1)'} B'\left(\sum_{n=1}^{N} Y_n^m\right)', \frac{\lambda_2}{2}\right) \right.$$
> > $$\left. (NhI_P + \lambda_3 L)^{-1}\right]',$$
> >
> > where $h = \text{scalar}\left(\hat{c}_m^{(i+1)'} B'B\hat{c}_m^{(i+1)}\right)$, $\text{Soft}(a, b) = \text{sign}(a)(|a| - b)_+$ is the soft-thresholding operator.
> >
> > **Step 3**: $\hat{f}_m^{(i+1)} = \hat{f}_m^{(i+1)}/\|\hat{f}_m^{(i+1)}\|_2$
> >
> > **If** both $\hat{f}_m$ and $\hat{c}_m$ are convergent with
> > $$\max\left(\frac{|\hat{f}_{m1}^{(i+1)} - \hat{f}_{m1}^{(i)}|}{\hat{f}_{m1}^{(i)}}, \ldots, \frac{|\hat{f}_{mP}^{(i+1)} - \hat{f}_{mP}^{(i)}|}{\hat{f}_{mP}^{(i)}}, \frac{|\hat{c}_{m1}^{(i+1)} - \hat{c}_{m1}^{(i)}|}{\hat{c}_{m1}^{(i)}}, \right.$$
> > $$\left. \ldots, \frac{|\hat{c}_{m(T+2)}^{(i+1)} - \hat{c}_{m(T+2)}^{(i)}|}{\hat{c}_{m(T+2)}^{(i)}}\right) \leq 10^{-5}, \textbf{ then } \text{break};$$
>
> $\hat{Q}_m = B\hat{c}_m$ is the estimated mth common trend and $\hat{f}_m$ is the corresponding estimated factor scores.

$Y_n^{m+1} = Y_n^m - \hat{f}_m\hat{c}_mB'$

Note: Instead of giving a predetermined number of top common trends, M, to extract, the percentage of variance, denoted as V, can be specified to be explained by the top common trends. Then $M = \arg\min_M\left[\frac{N\|\hat{F}\hat{C}B^T\|_F^2}{\sum_{n=1}^{N}\|Y_n\|_F^2} \geq V\right]$, where V is set as 85% by default in the MTA package.

Algorithm 2 Hypothesis testing for group comparison

Input: the observed data $X = \{X_1, \ldots, X_{N_1}\}$ and $Z = \{Z_1, \ldots, Z_{N_2}\}$, and the number of resamplings R

For $r = 1, \ldots, R$

> **For** the difference array between X and Z
>
> > ∗ **Construct** the difference array $G^{*(r)} = Z^{*(r)} - X^{*(r)}$, where $Z^{*(r)} = \left\{Z_1^{*(r)}, \ldots, Z_{N^*}^{*(r)}\right\}$ and $X^{*(r)} = \left\{X_1^{*(r)}, \ldots, X_{N^*}^{*(r)}\right\}$ are randomly selected from cases and controls without replacement, respectively, $N^* = \min(N_1, N_2)$.
> > ∗ **Apply** MTA on $G^{*(r)}$ to extract the common trends $\hat{Q}_1^{(r)}, \ldots, \hat{Q}_{M^{(r)}}^{(r)}$, and obtain the corresponding estimated factor scores $\hat{f}_1^{(r)}, \ldots, \hat{f}_{M^{(r)}}^{(r)}$, where the number of common trends $M^{(r)}$ is related to the explained variance captured by estimates $\hat{F}^{(r)} = \left(\hat{f}_1^{(r)}, \ldots, \hat{f}_{M^{(r)}}^{(r)}\right)$ and $\hat{C}^{(r)} = \left(\hat{c}_1^{(r)}, \ldots, \hat{c}_{M^{(r)}}^{(r)}\right)'$ and is determined by $\arg\min_{M^{(r)}}\left(\|\hat{F}_{+1}^{(r)}\hat{C}_{+1}^{(r)}B'\|_F^2 - \|\hat{F}^{(r)}\hat{C}^{(r)}B'\|_F^2\right)/\|\hat{F}^{(r)}\hat{C}^{(r)}B'\|_F^2 \leq 10^{-3}$, where $\hat{F}_{+1}^{(r)} = \left(\hat{f}_1^{(r)}, \ldots, \hat{f}_{M^{(r)}}^{(r)}, \hat{f}_{M^{(r)}+1}^{(r)}\right)$ and $\hat{C}_{+1}^{(r)} = \left(\hat{c}_1^{(r)}, \ldots, \hat{c}_{M^{(r)}}^{(r)}, \hat{c}_{M^{(r)}+1}^{(r)}\right)'$.
> > ∗ **Calculate** the distance of $\hat{Q}_m^{(r)}$ from 0 by $D_m^{(r)} = \|\hat{Q}_m^{(r)}\|_2, m = 1, \ldots, M^{(r)}$.
>
> **For** the noise array
>
> > ∗ **Construct** the noise array $G_{Noise}^{*(r)} = X^{1(r)} - X^{2(r)}$, where $X^{1(r)} = \left\{X_1^{1(r)}, \ldots, X_{N^*}^{1(r)}\right\}$ and $X^{2(r)} = \left\{X_1^{2(r)}, \ldots, X_{N^*}^{2(r)}\right\}$ are exclusive and randomly selected from controls, $N^* = \lfloor N_1/2 \rfloor$.
> > ∗ **Apply** MTA on $G_{Noise}^{*(r)}$ to extract the primary random tend $\hat{Q}_{Noise}^{(r)}$.
> > ∗ **Calculate** the distance of $\hat{Q}_{Noise}^{(r)}$ from 0 by $D_{Noise}^{(r)} = \|\hat{Q}_{Noise}^{(r)}\|_2$.

Use the Wilcoxon rank-sum test [57] to obtain the p-value p_m for testing whether distances $D_m = \left\{D_m^{(1)}, \ldots, D_m^{(R)}\right\}$ are greater than $D_{Noise} = \left\{D_{Noise}^{(1)}, \ldots, D_{Noise}^{(R)}\right\}, m = 1, \ldots, M^*$, where $M^* = \max\left(M^{(1)}, \ldots, M^{(R)}\right)$ and $D_m^{(r)}$ is a missing value if $m > M^{(r)}$.

Obtain the adjusted p-values based on a gatekeeping procedure [58, 59]. Specifically, the adjusted p-values are $p_1^{adj} = p_1$ and $p_m^{adj} = \max\left(p_{m-1}^{adj}, p_m\right), m = 2, \ldots, M^*$.

Output: (i) the adjusted p-values for all the extracted common trends; (ii) the dominant taxa which contribute to the significant common trends; (iii) the discover rates for the dominant taxa among R resamplings; (iv) the mean and standard error of the estimated factor scores for the dominant taxa.

---

when the observed time period is longer, the differences between the areas under two curves balance out. Accordingly Permuspliner's power is reduced in differentiating dynamic patterns. Since in practice, microbial dynamic patterns are nonlinear and the group changing trends are more likely to intersect, as supported by our real data in Fig. 6, the proposed MTA method has superior power.

Sensitivity and specificity. The estimated powers above illustrate that the proposed MTA method has superior performance in testing the differences between two groups than the competing method Permuspliner at the

group level. We further evaluate the performance of both methods at the taxon level in terms of sensitivity and specificity. Specifically, sensitivity is defined as the proportion of taxa correctly identified from 5 dominant taxa that truly contribute to the difference between two groups, and specificity is the proportion of uninvolved taxa that are correctly identified as such. Furthermore, the proposed MTA method determines whether a taxon

---

Algorithm 3 Classification

Input: a training set involving controls X and cases Z and a new subject Y_{new} with an unknown group label
For $Y = X$ and Z, respectively

>   **Apply** MTA on Y to extract M common trends $\hat{C}B'$ and obtain the corresponding estimated factor scores $\hat{F}$, where M is determined by the proportion of explained variance as indicated in Algorithm 1. Then the predicted microbial matrix is $\hat{Y} = \hat{F}\hat{C}B'$.

>   **Calculate** the distances of Y_{new} from $\hat{X}$ and $\hat{Z}$ and define them as $D^{con} = \|Y_{new} - \hat{X}\|_2$ and $D^{ca} = \|Y_{new} - \hat{Z}\|_2$, respectively.
>   **If** $D^{con} > D^{ca}$, then this new subject Y_{new} is classified as a case, otherwise a control.

---

is significantly contributing or not by the estimated confidence interval of its factor score with the number of resamplings $R = 50$ in Algorithm 2. For the competing method Permuspliner, the adjusted p-value with BH multiple comparison correction determines whether a taxon is significantly different between two groups or not.

Figure 3 presents the estimated sensitivity and specificity of the MTA and Permuspliner methods for identifying the dominant taxa that contribute to the microbial dynamic differences between control and case groups, with sample size $N = 20, 30$ and the number of time points $T = 10, 20$ in scenarios 1-3, respectively. The proposed MTA method has much higher sensitivity and sufficiently high specificity in all scenarios compared to Permuspliner, although Permuspliner has slightly higher specificity. This indicates that MTA has the ability to identify not only dominant taxa but also those inactive ones. But Permuspliner suffers from the limited sensitivity in identifying dominant taxa reliably, especially with $T = 20$, which agrees with its conservative performance in the power section. Thus, the proposed MTA method has better performance in identifying whether or not taxa are dominant than Permuspliner in the taxonomic analysis.

Furthermore, the proposed MTA method has the distinct advantage that it not only identifies the dominant taxa that contribute to the extracted common trends, but also quantifies their contributions by the estimated factor scores. This suggests their individual effects on the difference between two groups and provides additional insights for followup studies. Additional file 2: Figures S1–S3 report the extracted group difference trends and the estimated factor scores for the corresponding dominant taxa in scenarios 1-3 with $N = 30$ and $T = 10$ respectively, produced by our MTA R package.

Simulation results based on the simulation design used in Permuspliner method. To assess the robustness of the proposed MTA method, we further evaluated both methods using the simulation design in the original Permuspliner report [25], which is shown in detail in Additional file 1: Section S1. When the effect directions of the signal taxa at different time points are mixed, the relative performance of MTA and Permuspliner is quite similar to that reported above (the "Simulation results for group comparison" section). MTA exhibits general superior performance in both group and taxon levels evaluation (Additional file 2: Figures S4(A) and S5(A)). When the effects of the signal taxon across all the time points are in the same direction, which becomes less likely as the number of time points increases, Permuspliner has higher power than MTA, since Permuspliner directly measures the area between the microbial splines of control and case groups for each taxon. However, as the effect size gets larger, the estimated powers of MTA and Permuspliner become similar (Additional file 2: Figure S4(B)). Furthermore, MTA has similar performance with Permuspliner in terms of sensitivity and specificity (Additional file 2: Figure S5(B)). Therefore, considering that the same effect direction across a long period is less likely, the proposed MTA method is robust and has superior performance to Permuspliner in terms of statistical power at the group level, as well as in terms of sensitivity and specificity at the taxonomic level.

Simulation results for classification
In this subsection, we evaluate the performance of the proposed classification Algorithm 3 in terms of ROC curve and AUC using the 10-fold cross-validation (CV) method based on scenario 2 with different effect sizes. Specifically, we simulated the microbial relative abundances for 300 cases and 300 controls, where $\boldsymbol{\beta}_{.t}^1$ and $\boldsymbol{\beta}_{.t}^2$ were assigned as below,

$$\boldsymbol{\beta}_{.t}^1 = (-1, 2, -1)' \times \min(|e \times \sin(t)|, \beta_{0p}/2) \times \text{sign}(\sin(t)), \; p \in \wedge_1; \tag{1}$$
$$\boldsymbol{\beta}_{.t}^2 = (-1, 1)' \times \min(|e \times \sin(2t)|, \beta_{0p}) \times \text{sign}(\sin(2t)), \; p \in \wedge_2, \tag{2}$$

where e is the magnitude of effect size and was assigned as $1, 2, 4, 6$ and 8, respectively and $\text{sign}(\cdot)$ is the sign function. We then randomly divided the 600 subjects into 10 groups of equal size, each of which was in turn used as a testing set to classify a new subject, while the rest were used as a training set to establish the predicted microbial composition matrix for the control and case groups. The ROC curve and AUC were estimated as the average performance over 10 testing sets.

Additional file 2: Figure S6 and Fig. 4 present the overall ROC curves and AUCs with effect size $e = 1, 2, 4, 6, 8$,
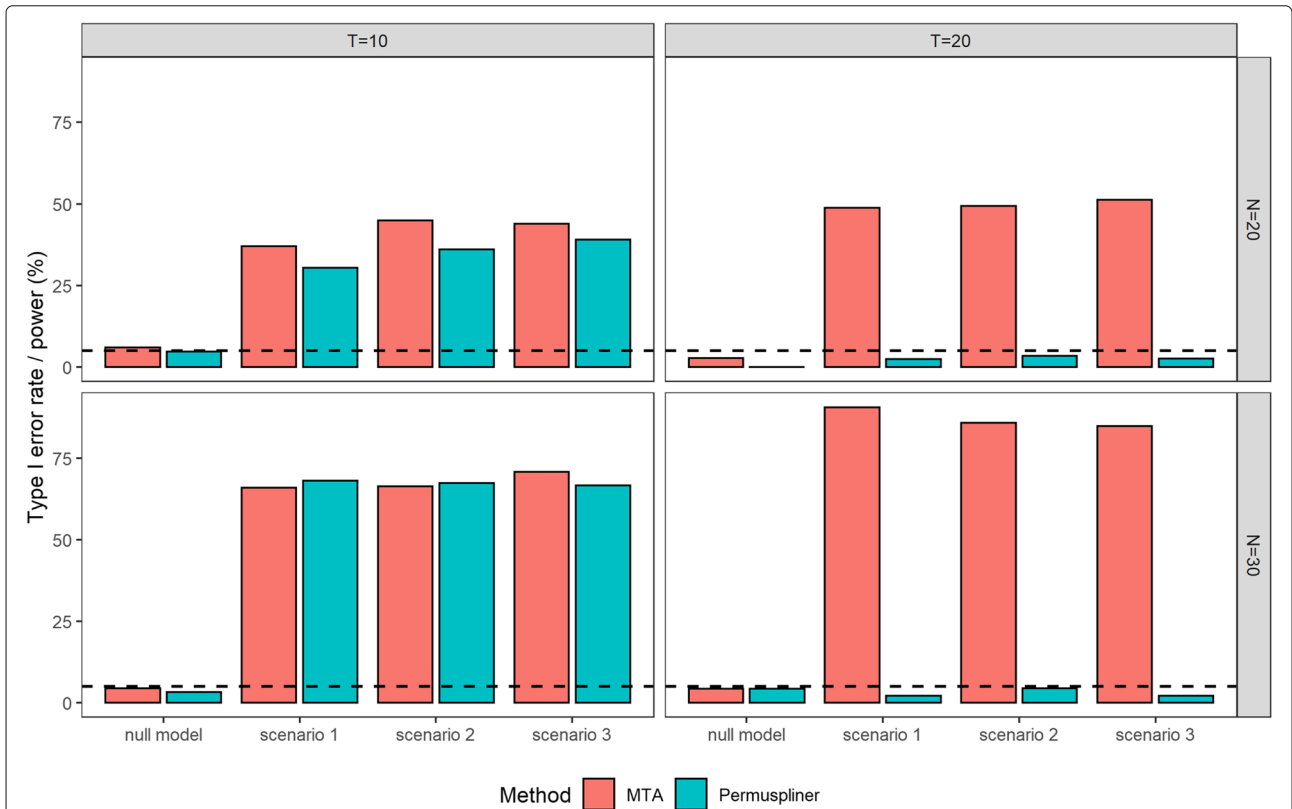
Fig. 2 Empirical type I error rate and power (%) for testing the difference between case and control groups with sample size $N = 20, 30$ and the number of time points $T = 10, 20$ in null model and scenarios 1-3, respectively. The dashed line represents the given significance level 5%
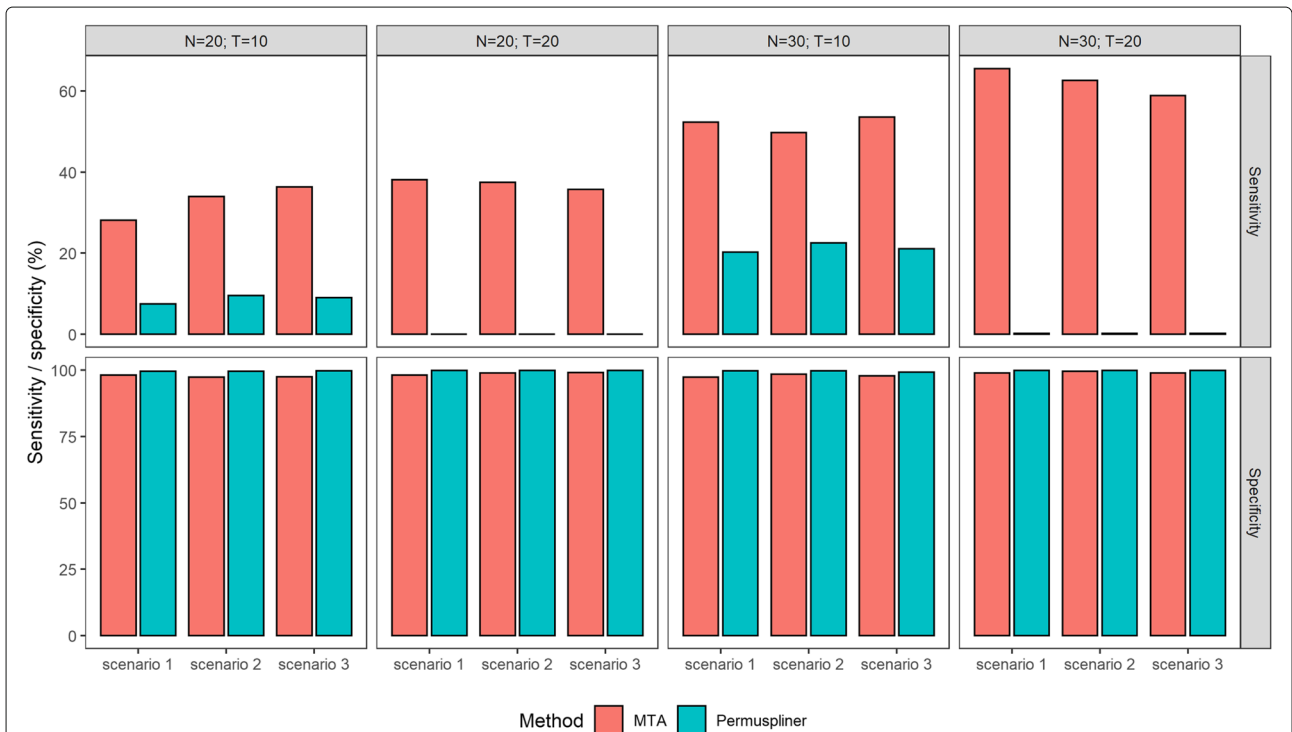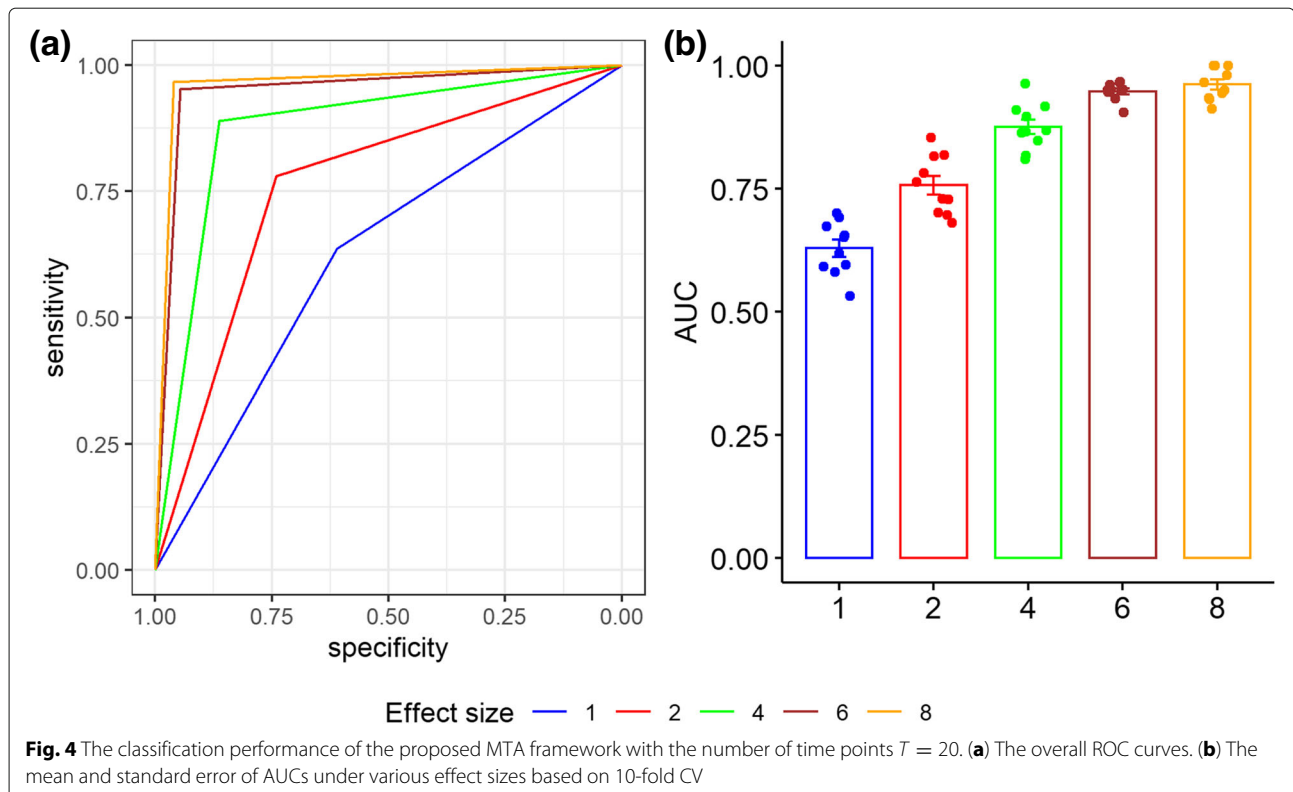

Fig. 3 The estimated sensitivity and specificity for identifying the dominant taxa which contribute to the extracted trends with sample size $N = 20, 30$ and the number of time points $T = 10, 20$ in scenarios 1-3, respectively

Fig. 4 The classification performance of the proposed MTA framework with the number of time points $T = 20$. (**a**) The overall ROC curves. (**b**) The mean and standard error of AUCs under various effect sizes based on 10-fold CV

and the number of time points $T = 10, 20$, respectively. As expected, AUC increases as the effect size e increases, with the average AUC being 0.63, 0.76, 0.88, 0.95 and 0.96 for $e = 1, 2, 4, 6, 8$ respectively. These results show that the proposed MTA framework has satisfactory performance to classify.
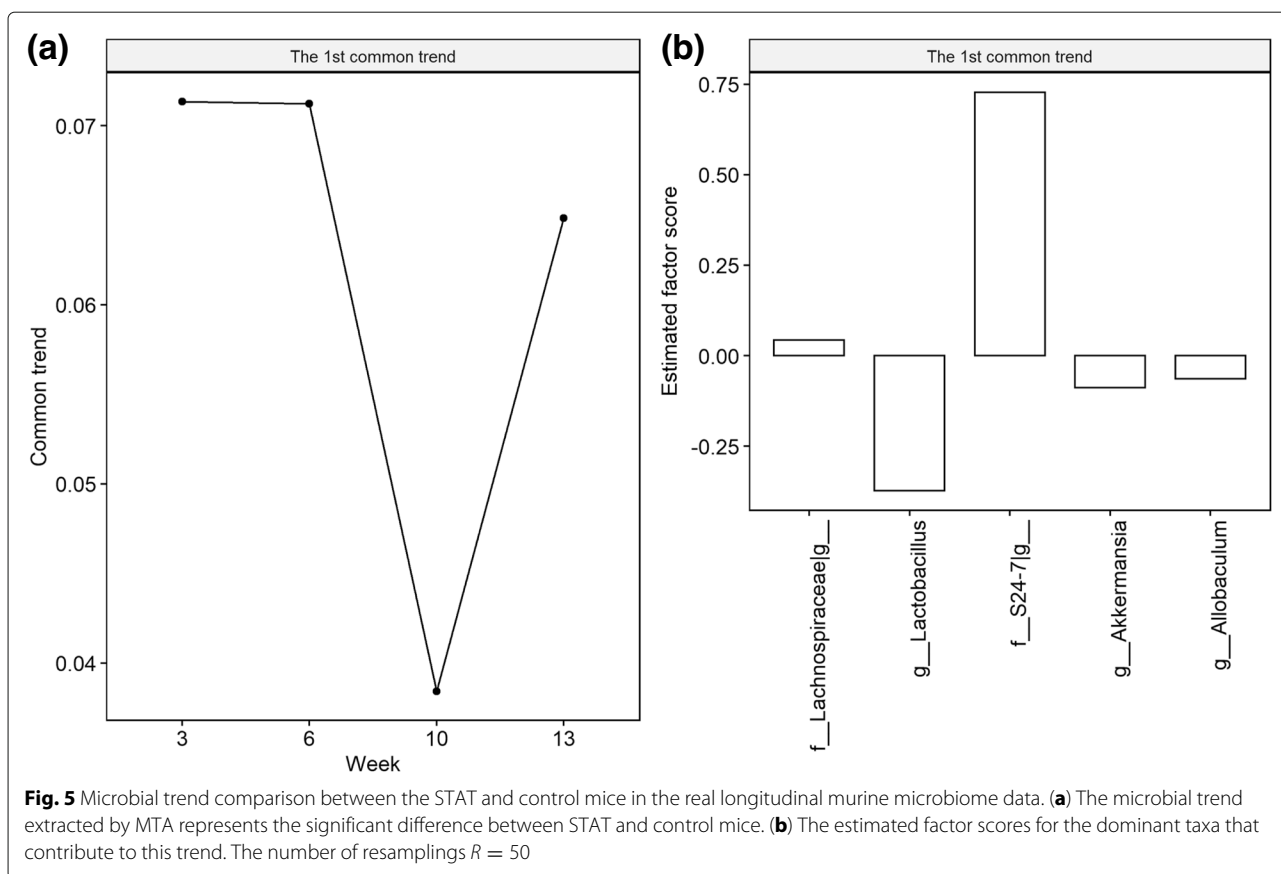
Real data analysis

Livanos et al. (2016) [35] conducted a longitudinal microbiome experiment in mice at risk for developing Type I diabetes (T1D). They found that early exposure to antibiotics altered the gut microbiota and this shift accelerated T1D onset compared to control mice. Here we apply the proposed MTA method to the longitudinal microbiome data collected in this experiment to examine the changing microbial trend differences between the STAT (antibiotic) and control groups. Specifically, MTA determines whether the overall dynamic patterns present in the STAT and control groups are significantly different or not, identifies which taxa contribute to the identified overall group difference, and subsequently provides the corresponding estimated factor scores for those identified taxa. Algorithm 3 in the MTA framework then is applied to classify the mice based on their longitudinal microbial profiles.

In this study, the fecal samples collected from antibiotic and control mice at 3, 6, 10, and 13 weeks were sequenced to examine their 16S rRNA genes, and their median

sequencing depths were $12,524 \pm 3,295$ sequences. The OTU table and taxonomy were determined using the QIIME pipeline [36], and 106 genera were originally observed. As indicated in the "Simulation design" section, we analyzed 35 genera in 17 antibiotic and 20 control male mice at 3, 6, 10 and 13 weeks.

Figure 5a reports the primary significant microbial trend determined by the MTA method, with p-value 9.8×10^{-10}, representing the major difference in the microbial dynamic pattern between antibiotic and control male mice. With MTA, based on the 35 genera, we identify 5 dominant genera (*Lactobacillus*, *Lachnospiraceae_Other*, *S24-7_Other*, *Akkermansia*, and *Allobaculum*) that contribute to the microbial trend difference. In Fig. 5b, we report the nonzero estimated factor scores of these 5 dominant genera and they reveal the extent of their contributions to the microbial trend group difference. In contrast, the competing test Permuspliner only identifies 3 significant genera (*Lactobacillus*, *Coprobacillus*, *Coriobacteriaceae_Other*) associated with STAT treatment (adjusted p-values $<5\%$).

Figure 6 presents the group average relative abundances over time for each identified genus by either method. Specifically, genus *Lactobacillus* identified by both methods has consistently higher relative abundances at all 4 time points in the control group than in the case group. The genera identified by MTA tend to have different changing directions between two groups, such as genera
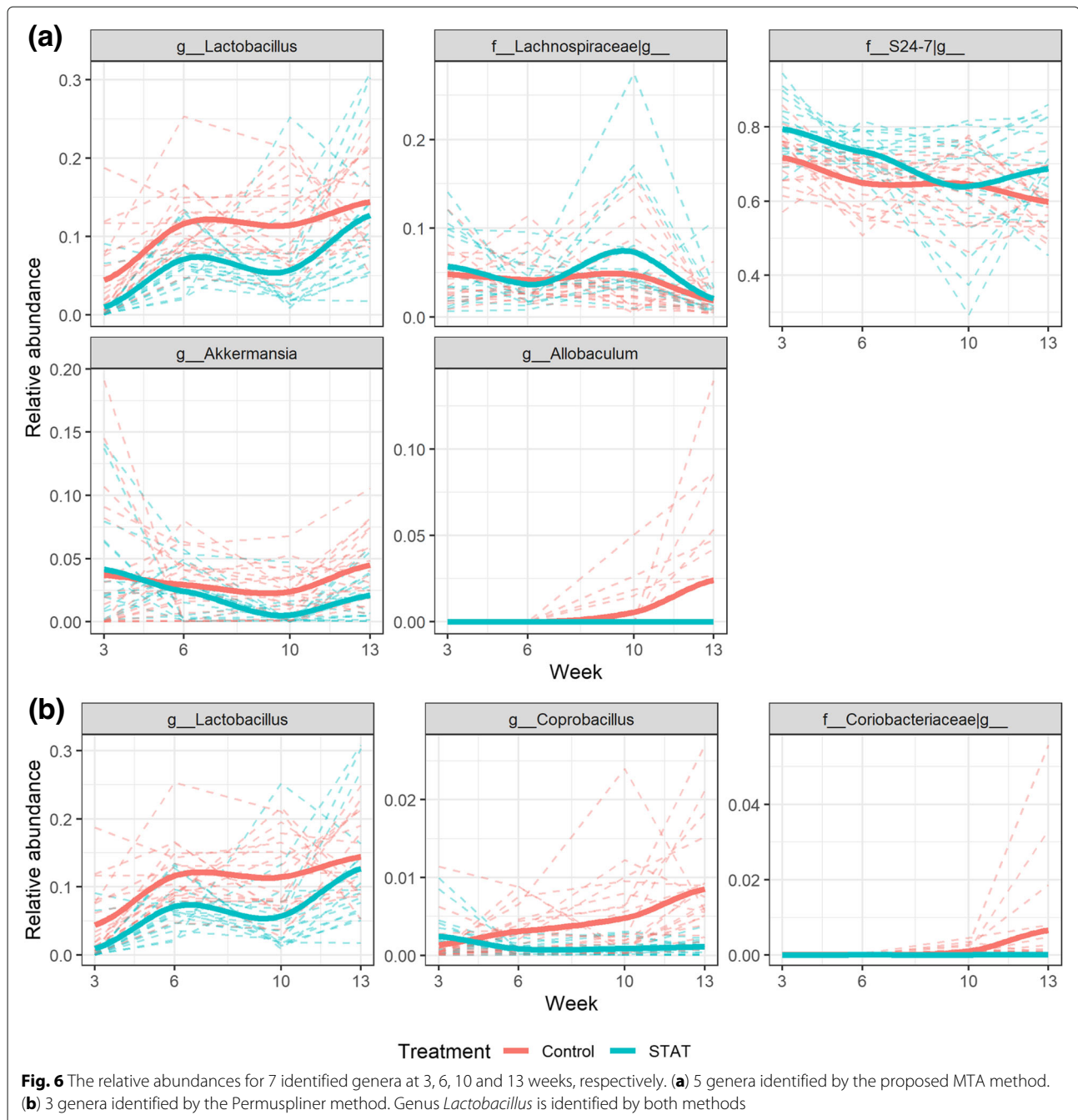
Fig. 5 Microbial trend comparison between the STAT and control mice in the real longitudinal murine microbiome data. (**a**) The microbial trend extracted by MTA represents the significant difference between STAT and control mice. (**b**) The estimated factor scores for the dominant taxa that contribute to this trend. The number of resamplings $R = 50$

Lachnospiraceae_Other, *S24-7_Other*, and *Akkermansia*, while Permuspliner fails to identify them, which agrees with the simulation results. The 5 genera identified by MTA are relatively common with their cumulative relative abundances ranging from 46.4% to 97.6% across all 4 time points. In contrast, two of three genera identified by Permuspliner are rarer with their cumulative relative abundances ranging from 0.6% to 31.0% (Additional file 2: Figure S7). Additional file 2: Figure S8 presents a principal coordinate analysis (PCoA) visualization based on the Bray-Curtis dissimilarity index using all 35 original genera; the 5 genera identified by MTA, and the 3 genera identified by Permuspliner at 3, 6, 10 and 13 weeks, respectively. As expected, the PCoA plot based on the genera identified by MTA closely resembles the original plot, while the plot based on the genera identified by Permuspliner presents a completely different pattern. These results illustrate that 5 genera identified by MTA well represent the original microbial diversities among all samples, which demonstrates that MTA is capable of capturing the differential microbial dynamic signals, and representing taxa in the longitudinal microbiome analysis.

Figure 7 presents the distances of all mice from the predicted microbial matrices of control and STAT mice which are estimated from the training set based on

Algorithm 3. For a given mouse, the training set consists of all other mice. We observe that 18 of 20 control mice (specificity=90.0%) and 13 of 17 antibiotic mice (sensitivity=76.4%) are classified correctly by the proposed distance-based classification algorithm (Algorithm 3), which is superior to the clustering results based on beta diversity measurements at a single time point (Additional file 2: Figure S8) as well as the results in Livanos et al. (2016) [35] based on the hierarchical clustering using the samples at week 6. These results demonstrate the satisfactory performance of the proposed MTA method in classification. Furthermore, the variation of distances shows that STAT mice have more diverse microbial profiles than control mice, which is consistent with their beta diversity measurements shown in Additional file 2: Figure S8.
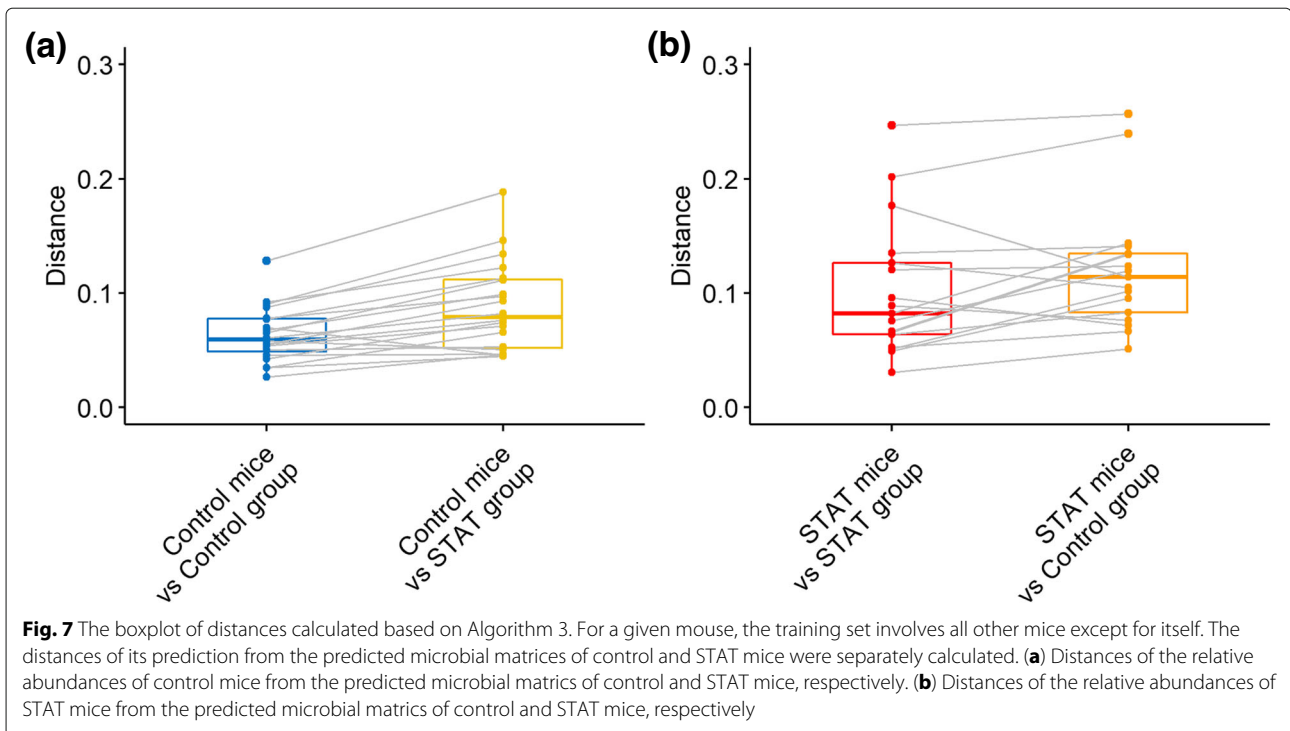
Compared to the significant signals found in Livanos et al. (2016) [35], the proposed MTA method identifies more significant results in terms of dominant taxa identification and mouse classification, which provides insight into the comprehensive characterization of the gut microbiome. This further shows the importance of investigating microbial dynamic changes and their characterization in human diseases based on the longitudinal microbial profiles.

Wang *et al. BMC Genomics*   (2021) 22:667

Page 9 of 16

Fig. 6 The relative abundances for 7 identified genera at 3, 6, 10 and 13 weeks, respectively. (**a**) 5 genera identified by the proposed MTA method. (**b**) 3 genera identified by the Permuspliner method. Genus *Lactobacillus* is identified by both methods

Discussion

Compared to the PTA method [26], which was proposed to specifically extract the common gene expression trend patterns shared by the subjects of a group and identify the dominant genes, the proposed MTA framework is a more comprehensive analytical pipeline on longitudinal microbiome data with three components. First, similar to but extended from the PTA method, the MTA framework extracts the microbial common trends and identifies the key taxa, which employs not only the Lasso penalty and the smoothing technique to deal with the high dimensionality of taxa and the smoothness of the extracted trends, but also the Laplacian penalty to address the distinct feature of microbiome data, the phylogenetic tree structure. The second component is to evaluate the group differential in microbial dynamic trend with permutation strategy, and the third component is to construct a classification algorithm based on microbial dynamic profiling, which are two new add-ons in MTA. The MTA framework assumes that the microbiota act as an integrated microbial

Fig. 7 The boxplot of distances calculated based on Algorithm 3. For a given mouse, the training set involves all other mice except for itself. The distances of its prediction from the predicted microbial matrices of control and STAT mice were separately calculated. (**a**) Distances of the relative abundances of control mice from the predicted microbial matrics of control and STAT mice, respectively. (**b**) Distances of the relative abundances of STAT mice from the predicted microbial matrics of control and STAT mice, respectively

community and there are a group of key taxa that drive the overall microbial dynamics, rather than assuming that each microbe acts alone and needs to be tested individually [2, 37]. This community level integrated analysis provides a systematic view of microbial dynamic responses, which is especially important for understanding complex diseases.

The compositional nature of the microbiome has raised many debates in the microbiome data studies [38–40]. The constant sum constraint can lead traditional statistical methods to produce spurious correlations and errant results. Various log-ratio transformation techniques have been used to deal with the compositionality of microbiome data, such as additive log-ratio transformation (ALR), centered log-ratio transformation (CLR), and isometric log-ratio transformation (ILR) [38–42]. These transformations can maintain the underlying covariance or correlation structure originating from the natural interaction of the components theoretically [41, 42]. However, the utility of log-ratio transformations is not able to ease the statistical analyses entirely from the influence of the compositionality, even in the cross-sectional study [38]. For the longitudinal microbiome studies, there is not much discussion on how to appropriately perform the log-ratio transformations yet. Many questions remain open. For example, how to define the invariant reference taxon, how to interpret the ratio when the denominator varies across time points, whether the transformation retains the inherent temporal pattern, etc. In this paper,

we took a modest action to conduct a small simulation and the results in Additional file 2: Figure S9 exhibit that both ALR and CLR transformations alter the dynamic patterns of the causal taxa that contribute to the group difference between control and case under the simulated scenario 1. Note that we didn't include ILR in the simulation, because it relies on the choice of the orthonormal basis which can be arbitrary. For sure, more comprehensive research are needed to evaluate the effect of log-ratio and other transformations in the longitudinal microbiome study.

On the other hand, we conducted additional simulations to investigate the utility of ALR and CLR transformations in the MTA method. Specifically, we applied the proposed MTA framework to the relative abundances, the corresponding log transformation, ALR, and CLR, respectively (see the details in Additional file 1: Section S2). The estimated powers show that using relative abundances captures more dynamic information and produces the highest power among all the methods in all simulating scenarios. The reason that MTA based on the relative abundance has good performance is twofold. First, MTA evaluates the dynamic difference between cases and controls by comparing the trends extracted from a difference array $\mathbf{G}^*$ to a noise trend which is randomly arranged around 0 across all time points, rather than directly compare the dynamic trends extracted from cases and controls, respectively. Second, MTA employs matrix factorization and regularization techniques to analyze all

taxa together and identify the dominant taxa simultaneously, by assuming that the microbiota act as an integrated microbial community. This strategy addresses the correlations among taxa due to compositionality to some extent [43, 44]. We recognize that all results above are based on our simulation studies, rather than on theoretical consideration. We include log, ALR, and CLR transformations in the MTA R package as options and let the users decide the transformation procedure based on their studies.

Sparsity is another feature of microbiome data, as many taxa are rare and most samples show numerous zero counts. Recent microbiome quality control studies and practical data analyses indicate that many observed rare taxa are subject to sequencing artifacts, contamination, or sequencing error [45–47]. Filtering out the rare taxa is one common approach to deal with this problem. Cao et al. (2021) [48] demonstrates that filtering reduces the complexity of microbiome data while preserving their integrity in downstream analysis and allows researchers to generate more reproducible and comparable results in microbiome data analysis. In Additional file 1, Section S7, we simulated the longitudinal microbiome data from a zero-inflated NB distribution to evaluate whether and how sparsity or zero inflation affects the performance of MTA. Both low vs. high and stable vs. time-varied zero inflations scenarios were considered in the simulation studies. We observe that the sparsity of causal taxa affects the performance of the proposed MTA, while the sparsity of non-causal taxa does not. As the sparsity of causal taxa increases, the estimated sensitivity decreases, and the specificity increases.

The findings of this study have to be seen in light of some challenges. Note that the core of the MTA framework is to identify the common patterns shared by all subjects in a given group. The within-group heterogeneity and outliers potentially weaken MTA's performance in group comparison and classification. Recently, archetypal analysis has been widely used in the clustering and classification, aiming to identify distinct archetypes (extremal points) as the representative, rather than the prototypes (central points or medoids) [49, 50]. The idea on how to extract archetypes gives MTA framework a possible solution to address the issues of heterogeneity and outliers. The proposed MTA framework requires all subjects have microbiome sequencing data at the same time points, and does not allow the missing values, which are not in common in the longitudinal studies. Imputation strategies are suggested to address the missing value issue before applying the MTA framework. Given that the longitudinal relative abundance of one taxon can be considered as a time series, one can employ appropriate imputation methods for univariate time series [51] to impute the missing relative abundance accordingly, then do the normalization

to maintain the compositionality of the microbiome. As a nonparametric method, MTA is also limited for covariate adjustment.

Although this paper focused on binary phenotypes, the proposed MTA framework can be easily generalized for use with more than two categories. For group comparisons, Algorithm 2 can be used to compare the paired difference between any two categories and then determine their microbial differential with correction for multiple testing. For the classification, the predicted microbial matrix for each category can be estimated based on the training set, then the distances of a new subject from its microbial profile can be calculated in relation to the predicted microbial matrices of all categories separately, and finally to determine its label by comparing those distances.

Conclusions

In this paper, we propose a microbial trend analysis (MTA) framework for analyzing the longitudinal microbiome data. MTA can describe microbial dynamics, test the group difference, extract key taxa driving the microbial temporal trend, and classify the subjects. Comparing to the competing method Permuspliner, MTA has superior performance in these tasks based on both extensive simulations and real data analyses. Consequently, with the recent proliferation of microbial longitudinal studies, the proposed MTA framework is an attractive analytical tool to study the comprehensive characterization of microbial dynamics and identify key bacterial species that may affect susceptibility to complex diseases.

Methods

Microbial trend analysis

Suppose there are N subjects in a given group. Let Y_{npt} be the relative abundance of the pth taxon of subject n at time point t, $p = 1, \ldots, P$, and $t = 1, \ldots, T$, and $Y_n = \left\{Y_{npt}\right\}_{p=1,t=1}^{p=P,t=T}$ be a $P \times T$ matrix of the relative abundances of subject n. To extract the common trends shared by all N subjects, we consider the following optimization problem, as illustrated in PTA [26] (Additional file 1: Section S3),

$$\min_{F,C} L(F, C | Y_n) = \sum_{n=1}^{N} \|Y_n - FCB'\|_F^2, \qquad (3)$$

subject to $c_m' \Omega c_m \leq a_1$, $\|F\|_1 \leq a_2$, and $\|F\|_2^2 = 1$,

where $F = (f_1, \ldots, f_M)$ is a $P \times M$ matrix of factor scores, $C = (c_1, \ldots, c_M)'$ is an $M \times (T + 2)$ matrix of spline coefficients, $B = \{B_i(t)\}_{t,i=1}^{t=T,i=T+2}$ is a $T \times (T + 2)$ matrix containing the cubic spline basis and $\|\cdot\|_F$ is the Frobenius norm. Denote $Q = CB'$, which is an $M \times T$ matrix presenting the top M common trends across T time points. $\Omega = \{\Omega_{ij}\}_{i,j=1}^{i,j=T+2}$ is a $(T + 2) \times (T + 2)$ matrix with

$\Omega_{ij} = \int B_i''(t)B_j''(t)dt, i, j = 1, \ldots, T + 2$. M can be either predetermined by the user or determined by specifying the percentage of variance to be explained by the top common trends (see Algorithm 1 for details). The parameter $a_1 (\geq 0)$ controls the smoothness of trends with a smaller value of a_1 producing smoother trends. $a_2 (\geq 0)$ controls the sparsity of taxa to ease the high-dimentionality problem in the microbiome data, that the number of taxa P is usually far larger than the sample size N or the number of time points T, and only a few taxa are assumed to make substantial contributions to the community level dynamic trends [52]. Note that F and C are identical for all subjects respectively, for they integrate information from all subjects and represent the common trends Q shared by all subjects. The elements of f_m are the weights of P taxa on the mth common trend Bc_m, of which a higher absolute value indicates a stronger contribution to this common trend. With this modeling, we can find the top M common trends from a group of subjects, and quantify the contributions of P taxa to each of the common trends.

In addition to the high-dimensionality, another important feature of microbiome data is the phylogenetic tree structure that describes the evolutionary relationships among taxa. Evolutionarily related taxa tend to have similar effects on human phenotypes or diseases [27, 28, 30, 53]. In the existing studies, incorporating the phylogenetic tree information into the analysis improves both statistical power and biological interpretability [27–30]. To take this factor into account, we integrate the following Laplacian penalty in the optimization problem (3),

$$g(\lambda, L) = \lambda F'LF, \tag{4}$$

where $\lambda (\geq 0)$ is the tuning parameter. The Laplacian matrix L is determined by the phylogenetic tree and constructed using a similar approach to that described in Chen et al. (2012) [27]. Detailed construction of L is given in Additional file 1: Section S4. Incorporating this Laplacian penalty L weights the OTUs closely linked on the phylogenetic tree to have similar effects on the extracted common trends. With these two additional penalties, we introduce two Lagrange multipliers to rewrite the optimization problem (3) as

$$\min_{F,C} \left[L(F, C | Y_n) + \lambda_1 C\Omega C' + \lambda_2 \|F\|_1 + \lambda_3 F'LF \right], \tag{5}$$

subject to $\|F\|_2^2 = 1$, where $\lambda_1 (\geq 0)$, $\lambda_2 (\geq 0)$ and $\lambda_3 (\geq 0)$ are tuning parameters that control smoothness of the common trends, sparsity, and smoothness of the estimated factor scores based on the phylogenetic tree distance, respectively. As demonstrated in Zhang and Davis (2013) [26], the optimization problem (5) is biconvex, though not convex. Thus we adapt an iterative nonlinear partial least square algorithm in Algorithm 1 to get estimates $\hat{F}$ and $\hat{C}$ by minimizing the optimization problem

(5) [26, 54], modified from the Algorithm S1. The detailed derivation is provided in Additional file 1: Section S5.

Given the observations $Y_n, n = 1, \ldots, N$, the number of top common trends M or the percentage of variance V for the top common trends to explain, and the tuning parameters λ_1, λ_2 and λ_3, Algorithm 1 solves for the estimated top common trends $\hat{Q}_m = B\hat{c}_m, m = 1, \ldots, M$, shared by all N subjects, which are the linear combinations of the relative abundances of all P taxa across T time points, and $\hat{F}$ indicates the dominant taxa and their contributions to the common trends.

Note that the choices of tuning parameters in Algorithm 1 are crucial for extracting the common trends and we use the K-fold CV [26, 27, 55] to determine the values of λ_1, λ_2 and λ_3. Specifically, we randomly split the indexes of all subjects $1, \ldots, N$ into K exclusive folds with roughly equal sample size $G_1, \ldots, G_K$, and take the subjects $Y_n, n \notin G_k$, as the training set and the remaining subjects as the testing set in the kth cross-validation, $k = 1, \ldots, K$. With the given candidate values of the tuning parameters $\lambda = (\lambda_1, \lambda_2, \lambda_3)$, we obtain estimates $\hat{F}_\lambda^{-k}$ and $\hat{C}_\lambda^{-k}$ from the training set with Algorithm 1 in the kth cross-validation. The predicted group-level microbial composition for this training set therefore is $\hat{Y}_\lambda^{-k} = \hat{F}_\lambda^{-k} \hat{C}_\lambda^{-k} B'$. Then the total squared error on the testing set $Y_n, n \in G_k$, is defined as $e_k(\lambda) = \sum_{n \in G_k} \|Y_n - \hat{Y}_\lambda^{-k}\|_F^2$, and the average mean squared error over all K folds is recorded as

$$CV(\lambda) = \frac{1}{N} \sum_{k=1}^{K} e_k(\lambda) = \frac{1}{N} \sum_{k=1}^{K} \sum_{n \in G_k} \|Y_n - \hat{Y}_\lambda^{-k}\|_F^2.$$

Finally, we select $\lambda^* = \arg\min_\lambda CV(\lambda)$ as the optimal tuning parameters.

Group comparison

In the subsection above, we proposed the MTA method to capture the common dynamic trends shared by all subjects in a given group and identify the dominant taxa contributing to these common trends. We next apply the MTA method to evaluate how different two groups are in terms of their microbial dynamic pattern, and if the group difference is significant, quantify the differentiating taxa. The proposed group test is based on a difference array constructed using pseudo samples and permutation technique. Specifically, let $X = \{X_1, \ldots, X_{N_1}\}$ and $Z = \{Z_1, \ldots, Z_{N_2}\}$ be the arrays of relative abundances for N_1 controls and N_2 cases with P taxa across T time points (here, X_n and Z_n have a similar definition as Y_n in the subsection above), respectively. The difference array then is constructed as

$$G^* = Z^* - X^*,$$

where $G^* = \{G_1^*, \ldots, G_{N^*}^*\}$ and G_n^* represents the relative abundance difference between randomly paired subjects from case and control groups with $\sum_{p=1}^{P} G_{npt}^* = 0$, $n = 1, \ldots, N^*$, $t = 1, \ldots, T$, and $N^* = \min(N_1, N_2)$. $X^* = \{X_1^*, \ldots, X_{N^*}^*\}$ and $Z^* = \{Z_1^*, \ldots, Z_{N^*}^*\}$ are the permuted controls and cases, respectively.

Under the null hypothesis of no difference between cases and controls, the expectation of G_n^* should be a $P \times T$ zero matrix and the difference array G^* thus has only noise signal. On the other hand, under the alternative hypothesis, that control and case groups hold different dynamic patterns, G_n^* is expected to be time-related and to contain some common trends illustrating the underlying differences between cases and controls, which should differ from the random noise signal under the null. Thus, we propose a test to evaluate the dynamic differences between cases and controls by comparing the trends extracted from G^* to the noise trend which is randomly arranged around 0 across all time points. Specifically, the test is evaluated based on the resampling (or bootstrapping) method [56], with details described in Algorithm 2.

Note that Algorithm 2 consists of several components used in constructing the hypothesis test for assessing the microbial dynamic group differences. First, we determine the number of common trends extracted from the difference array $G^{*(r)}$, $M^{(r)}$, by the criterion that the increasing rate of the explained variance is equal to or less than 10^{-3}, to capture most of the information shared by all pseudo subjects. Second, we employ the straightforward and widely used Euclidean distance to measure the differences between the common trends extracted from G^* and the random trend from G_{Noise}^*. Third, we use the Wilcoxon rank-sum test to check whether distances D_m are greater than D_{Noise}, and provide the corresponding p-value p_m for the mth common trend, $m = 1, \ldots, M^*$. Finally, multiple testing correction is needed to adjust these individual p-values to control the overall error rate. Since these M^* hypothesis tests exhibit a hierarchical structure in which the information captured by the common trend Q_m monotonically decreases as m increases, we use a gatekeeping procedure [58, 59] to take care of this hierarchical structure and obtain the adjusted p-values p_m^{adj}, $m = 1, \ldots, M^*$. This implies that the mth common trend is meaningful only if the previous one has a significant difference.

Classification

In addition to the common dynamic trends and the estimated factor scores, the MTA method provides the predicted microbial composition matrix: $\hat{Y} = \hat{F}\hat{C}B'$ at the group level. Since the predicted microbial composition matrix integrates information from all subjects of a given group and represents the unique and underlying pattern defined by this group, it can serve as a prototype of this

group. Therefore, we propose a distance-based classification algorithm (Algorithm 3) to classify subjects based on their distances from the predicted microbial matrices of two different groups in the training set. Specifically, Algorithm 3 employs the k-nearest neighbors strategy [60] and classifies the new subjects in terms of the similarity. Suppose there is a training set involving controls X and cases Z, and a new subject Y_{new} with an unknown group label. First, we apply the MTA method to controls X and cases Z separately to obtain their corresponding predicted matrices $\hat{X}$ and $\hat{Z}$. Here the number of common trends M integrating the predicted microbial matrix is determined by the proportion of explained variance, as illustrated in Algorithm 1. Second, we calculate distances of the new subject Y_{new} from $\hat{X}$ and $\hat{Z}$, which are defined as $D^{\text{con}} = \|Y_{\text{new}} - \hat{X}\|_2$ and $D^{\text{ca}} = \|Y_{\text{new}} - \hat{Z}\|_2$ respectively. Finally, this new subject Y_{new} is classified as a case if $D^{\text{con}} > D^{\text{ca}}$, otherwise as a control.

Simulation design

Following Wang et al. (2020) [61], we designed our simulation at the genus level based on the real data illustrated in the "Real data analysis" section [35]. The real data include 106 genera for 20 control and 17 sub-therapeutic antibiotic treatment (STAT) male mice across 4 time points (weeks 3, 6, 10 and 13). After filtering out those genera that appear in less than 10% of samples or with mean proportions less than 10^{-4}, 35 taxa were included in the analysis. We simulated the microbial relative abundances at each time point for cases and controls respectively from the Dirichlet distribution. Specifically, the mean relative abundance of the pth taxon for a subject at the tth time point O_{pt} is assigned as below for $p = 1, \ldots, P$ and $t = 1, \ldots, T$,

$$E[O_{pt}] = \frac{\gamma_{pt}}{\sum_{p=1}^{P} \gamma_{pt}}, \ \gamma_{pt} = \beta_{0p} + \beta_{pt}, \tag{6}$$

where $\beta_0 = (\beta_{01}, \ldots, \beta_{0P})'$ represents the baseline relative abundances for $P = 35$ taxa, which were set as the corresponding estimates from the control mice using R package dirmult [62]. β_{pt} is the deviation from the baseline relative abundance β_{0p} of the pth taxon at the tth time point, which represents the group difference.

To estimate the empirical type I error rate, the relative abundances of controls and cases were generated under the null hypothesis with no group difference by setting $\beta_{pt} = 0$, $p = 1, \ldots, P$ and $t = 1, \ldots, T$.

To evaluate the statistical power, we designed three scenarios with two different common trends between control and case groups, of which 3 dominant taxa (about 10% of the total) contributed to the first one and 2 dominant taxa (about 5%) contributed to the second one. Denote the set of the indices of these 5 dominant taxa as $\wedge_1$ and $\wedge_2$ and their corresponding deviations as $\beta_{\cdot t}^1 = (\beta_{1t}^1, \beta_{2t}^1, \beta_{3t}^1)'$ and

Wang *et al. BMC Genomics* (2021) 22:667

Page 14 of 16

$\boldsymbol{\beta}_{.t}^2 = (\beta_{1t}^2, \beta_{2t}^2)'$. Following Zhang and Davis (2013) [26], they were assigned as

$$\boldsymbol{\beta}_{.t}^1 = (-1, 2, -1)' \times \sin(t) \times \min(1, \beta_{0p}/2), \ p \in \wedge_1; (7)$$

$$\boldsymbol{\beta}_{.t}^2 = (-1, 1)' \times \sin(2t) \times \min(1, \beta_{0p}), \ p \in \wedge_2. \quad (8)$$

Furthermore, the $\wedge_1$ and $\wedge_2$ were determined by taking the phylogenetic tree structure into account as follows. We first partitioned the 35 taxa into 5 clusters using the partition around medoids algorithm [63] on the patristic distance in the real phylogenetic tree. Second, we randomly selected two clusters and then randomly selected 3 and 2 taxa from these two clusters without replacement as the taxa in $\wedge_1$ and $\wedge_2$, respectively. Consequently, the dominant taxa that contributed to the group differences were phylogenetically related.

Three different simulation scenarios were constructed to thoroughly illustrate the performance of the proposed MTA method and the competing method Permuspliner. Scenario 1: The relative abundances in the control group were generated with $\beta_{pt} = 0$, while the relative abundances in the case group were generated with $\boldsymbol{\beta}_{.t}^1$ for the taxa in $\wedge_1$ and $\beta_{pt} = 0$ for $p \notin \wedge_1$. Scenario 2: The relative abundances in the control group were generated with $\beta_{pt} = 0$, while the relative abundances in the case group were generated with $\boldsymbol{\beta}_{.t}^1$ and $\boldsymbol{\beta}_{.t}^2$ for the taxa in $\wedge_1$ and $\wedge_2$ respectively and $\beta_{pt} = 0$ for $p \notin \wedge_1 \cup \wedge_2$. Scenario 3: The relative abundances in the control group were generated with $\boldsymbol{\beta}_{.t}^2$ for the taxa in $\wedge_2$ and $\beta_{pt} = 0$ for $p \notin \wedge_2$, while the relative abundances in the case group were generated with $\boldsymbol{\beta}_{.t}^1$ and $\boldsymbol{\beta}_{.t}^2$ for the taxa in $\wedge_1$ and $\wedge_2$, respectively, and $\beta_{pt} = 0, p \notin \wedge_1 \cup \wedge_2; p = 1, \dots, P, t = 1, \dots, T$. Hence, there is only one different common trend between control and case groups in both scenario 1 and scenario 3, but two different common trends in scenario 2. Finally, we considered two equal sample sizes $N_{case} = N_{control} = N = 20$ or 30, and two numbers of time points $T = 10$ or 20.

Abbreviations

ALR: Additive log-ratio transformation; AUC: Area under the curve; BH: Benjamini-Hochberg; CLR: Centered log-ratio transformation; CV: Cross-validation; IBD: Inflammatory bowel disease; ILR: Isometric log-ratio transformation; MTA: Microbial trend analysis; OTU: Operational taxonomic unit; PCoA: Principal coordinate analysis; PTA: Principal trend analysis; ROC: Receiver operating characteristic; STAT: Sub-therapeutic antibiotic treatment; T1D: Type 1 diabetes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07948-w.

Additional file 1: Section S1: Simulation study based on the simulation design of the perturbation method. **Section S2:** Comparison between relative abundance and log-ratio transformed relative abundances. **Section S3:** Principal trend analysis. **Section S4:** Laplacian matrix construction. **Section S5:** The derivation of Algorithm 1. **Section S6:** Comparisons among the proposed MTA, Permuspliner, MetaLonDA, and LonGP. **Section S7:** Sparsity evaluation of MTA.

Additional file 2: FIG S1. The proposed MTA method for the comparison between case and control groups in scenario 1 with sample size $N = 30$ and the number of time points $T = 10$. (A) The microbial trend extracted by MTA represents the significant difference between case and control groups. (B) The average and standard error of the estimated factor scores for the dominant taxa that contribute to the extracted trend, respectively. **FIG S2.** The proposed MTA method for the comparison between case and control groups in scenario 2 with sample size $N = 30$ and the number of time points $T = 10$. (A) Two microbial trends extracted by MTA represent the significant difference between case and control groups. (B) The average and standard error of the estimated factor scores for the dominant taxa that contribute to those two trends, respectively. **FIG S3.** The proposed MTA method for the comparison between case and control groups in scenario 3 with sample size $N = 30$ and the number of time points $T = 10$. (A) The microbial trend extracted by MTA represents the significant difference between case and control groups. (B) The average and standard error of the estimated factor scores for the dominant taxa that contribute to the extracted trend, respectively. **FIG S4.** Empirical power for testing the difference between case and control groups with sample size $N = 20, 30$ and the number of time points $T = 10, 20$ under 1X and 2X magnitudes of perturbation, respectively. Here, (A) $\boldsymbol{z} = (0, 0.3, 0.45, 0.2, -0.3, 0.3, -0.3, -0.2, -0.1, 0)'$ and $\boldsymbol{z} = (0, 0.2, 0.5, 0.3, 0.2, -0.2, -0.4, -0.4, -0.2, 0.2, 0.5, 0.2, -0.3, -0.4, -0.2, 0.2, 0.4, 0.2, 0.1, 0)'$, (B) $\boldsymbol{z} = (0, 0.1, 0.2, 0.3, 0.4, 0.3, 0.2, 0.1, 0.1, 0)'$ and $\boldsymbol{z} = (0, 0.05, 0.1, 0.2, 0.2, 0.2, 0.3, 0.3, 0.2, 0.2, 0.2, 0.2, 0.3, 0.3, 0.2, 0.2, 0.2, 0.2, 0.1, 0)'$, for $T = 10$, and 20, respectively. **FIG S5.** The estimated sensitivity and specificity for identifying dominant taxa which contribute to the extracted trends with sample size $N = 20, 30$ and the number of time points $T = 10, 20$ under 1X and 2X magnitudes of perturbation, respectively. Here, (A) $\boldsymbol{z} = (0, 0.3, 0.45, 0.2, -0.3, 0.3, -0.3, -0.2, -0.1, 0)'$ and $\boldsymbol{z} = (0, 0.2, 0.5, 0.3, 0.2, -0.2, -0.4, -0.4, -0.2, 0.2, 0.5, 0.2, -0.3, -0.4, -0.2, 0.2, 0.4, 0.2, 0.1, 0)'$, (B) $\boldsymbol{z} = (0, 0.1, 0.2, 0.3, 0.4, 0.3, 0.2, 0.1, 0.1, 0)'$ and $\boldsymbol{z} = (0, 0.05, 0.1, 0.2, 0.2, 0.2, 0.3, 0.3, 0.2, 0.2, 0.2, 0.2, 0.3, 0.3, 0.2, 0.2, 0.2, 0.2, 0.1, 0)'$, for $T = 10$, and 20, respectively. **FIG S6.** The classification performance of the proposed MTA framework with the number of time points $T = 10$. (A) The overall ROC curves. (B) The mean and standard error of AUCs under various effect sizes based on 10-fold CV. **FIG S7.** The relative abundances for the genera identified by MTA (5 genera) and Permuspliner (3 genera) at 3, 6, 10 and 13 weeks, respectively. **FIG S8.** Beta diversity analysis for all male mice based on the Bray-Curtis dissimilarity index. The PCoA is evaluated based on: (A1)-(A4) 35 original genera; (B1)-(B4) 5 genera identified by the proposed MTA method; and (C1)-(C4) 3 genera identified by the competing method Permuspliner at 3, 6, 10 and 13 weeks, respectively. Points represent samples. **FIG S9.** Evaluation of the transformations affect on the longitudinal trends of the causal taxa that contribute to group difference in scenario 1 with sample size $N = 30$ and the number of time points $T = 10$, respectively. (A) The true relative abundance time series of the causal taxa with equation (1). (B) The mean and standard error of the relative abundances without none transformation. (C) The mean and standard error of the relative abundances after ALR (additive log-ratio transformation), with the last one taxon being the reference. (D) The mean and standard error of the relative abundances after CLR (centered log-ratio transformation).

Authors' contributions

CW developed the microbial trend analysis framework, performed simulation analyses, real data analyses and manuscript writing. JH contributed to real data analyses and manuscript writing. MJB contributed to the biological insights and interpretation, and manuscript writing. HL contributed to the methodological ideas for the proposed framework, simulations, real data analyses, and manuscript writing. All authors read and approved the final manuscript.

Wang *et al. BMC Genomics* (2021) 22:667

Page 15 of 16

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Division of Biostatistics, Department of Population Health, New York University School of Medicine, 10016 New York, NY, USA. [2]Center for Advanced Biotechnology and Medicine, Rutgers University, 08854-8021 Piscataway, NJ, USA.

References

1. Hoffmann AR, Proctor L, Surette M, Suchodolski J. The microbiome: the trillions of microorganisms that maintain health and cause disease in humans and companion animals. Eur J Vet Pathol. 2016;53(1):10–21.
2. Integrative H. The integrative human microbiome project. Nature. 2019;569(7758):641.
3. Integrative H. The integrative human microbiome project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe. 2014;16(3):276.
4. Zhou W, Sailani MR, Contrepois K, Zhou Y, Ahadi S, Leopold SR, Zhang MJ, Rao V, Avina M, Mishra T, et al. Longitudinal multi-omics of host–microbe dynamics in prediabetes. Nature. 2019;569(7758):663.
5. Thaiss CA. Microbiome dynamics in obesity. Science. 2018;362(6417): 903–04.
6. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569(7758):655.
7. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, Huang B, Arodz TJ, Edupuganti L, Glascock AL, et al. The vaginal microbiome and preterm birth. Nat Med. 2019;25(6):1012.
8. Rose S. M. S.-F., Contrepois K, Moneghetti KJ, Zhou W, Mishra T, Mataraso S, Dagan-Rosenfeld O, Ganz AB, Dunn J, Hornburg D, et al. A longitudinal big data approach for precision health. Nat Med. 2019;25(5):792.
9. Koh H, Blaser MJ, Li H. A powerful microbiome-based association test and a microbial taxa discovery framework for comprehensive association mapping. Microbiome. 2017;5(1):45.
10. Koh H, Livanos AE, Blaser MJ, Li H. A highly adaptive microbiome-based association test for survival traits. BMC Genomics. 2018;19(1):210.
11. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med. 2018;24(4): 392.
12. Hu J, Koh H, He L, Liu M, Blaser MJ, Li H. A two-stage microbial association mapping framework with advanced fdr control. Microbiome. 2018;6(1):131.
13. Schmidt TS, Raes J, Bork P. The human gut microbiome: from association to modulation. Cell. 2018;172(6):1198–215.
14. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, Albert PS, Caporaso JG. q2-longitudinal: Longitudinal and paired-sample analyses of microbiome data. mSystems. 2018;3(6):00219–18.
15. Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. Bioinformatics. 2016;32(17):2611–17.
16. Cheng L, Ramchandran S, Vatanen T, Lietzén N, Lahesmaa R, Vehtari A, Lähdesmäki H. An additive gaussian process regression model for interpretable non-parametric analysis of longitudinal data. Nat Commun. 2019;10(1):1–11.
17. Bucci V, Tzen B, Li N, Simmons M, Tanoue T, Bogart E, Deng L, Yeliseyev V, Delaney ML, Liu Q, et al. Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses. Genome Biol. 2016;17(1):121.
18. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol. 2013;9(12):1003388.
19. Gibbons SM, Kearney SM, Smillie CS, Alm EJ. Two dynamic regimes in the human gut microbiome. PLoS Comput Biol. 2017;13(2):1005364.
20. Ridenhour BJ, Brooker SL, Williams JE, Van Leuven JT, Miller AW, Dearing MD, Remien CH. Modeling time-series data from microbial communities. ISME J. 2017;11(11):2526.
21. Shenhav L, Furman O, Briscoe L, Thompson M, Silverman JD, Mizrahi I, Halperin E. Modeling the temporal dynamics of the gut microbial community in adults and infants. PLoS Comput Biol. 2019;15(6):1006960.
22. Chen I, Kelkar YD, Gu Y, Zhou J, Qiu X, Wu H. High-dimensional linear state space models for dynamic microbial interaction networks. PloS ONE. 2017;12(11):0187822.
23. Lugo-Martinez J, Ruiz-Perez D, Narasimhan G, Bar-Joseph Z. Dynamic interaction network inference from longitudinal microbiome data. Microbiome. 2019;7(1):54.
24. Metwally AA, Yang J, Ascoli C, Dai Y, Finn PW, Perkins DL. Metalonda: a flexible r package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. Microbiome. 2018;6(1):32.
25. Shields-Cutler RR, Al-Ghalith GA, Yassour M, Knights D. Splinectomer enables group comparisons in longitudinal microbiome studies. Front Microbiol. 2018;9:785.
26. Zhang Y, Davis R. Principal trend analysis for time-course data with applications in genomic medicine. Ann Appl Stat. 2013;7(4):2205–28.
27. Chen J, Bushman FD, Lewis JD, Wu GD, Li H. Structure-constrained sparse canonical correlation analysis with an application to microbiome data analysis. Biostatistics. 2012;14(2):244–58.
28. Xiao J, Cao H, Chen J. False discovery rate control incorporating phylogenetic tree increases detection power in microbiome-wide multiple testing. Bioinformatics. 2017;33(18):2873–81.
29. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. Elife. 2017;6:21887.
30. Randolph TW, Zhao S, Copeland W, Hullar M, Shojaie A. Kernel-penalized regression for analysis of microbiome data. Ann Appl Stat. 2018;12(1):540–66.
31. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc. 1979;74(368):829–36.
32. Cleveland WS, Devlin SJ. Locally weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc. 1988;83(403):596–610.
33. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology. 1982;143(1):29–36.
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Methodol. 1995;57(1):289–300.
35. Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, McRitchie S, Li H, Chung J, Sohn J, Kim S, et al. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. Nat Microbiol. 2016;1(11):16140.
36. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, et al. Qiime

allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335.

37. Pinnell LJ, Turner JW. Shotgun metagenomics reveals the benthic microbial community response to plastic and bioplastic in a coastal marine environment. Front Microbiol. 2019;10:1252.

38. Tsilimigras MC, Fodor AA. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann Epidemiol. 2016;26(5):330–35.

39. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.

40. Rivera-Pinto J, Egozcue JJ, Pawlowsky-Glahn V, Paredes R, Noguera-Julian M, Calle ML. Balances: a new perspective for microbiome analysis. MSystems. 2018;3(4):e00053–18.

41. Aitchison J. The statistical analysis of compositional data. J R Stat Soc Ser B Methodol. 1982;44(2):139–60.

42. Aitchison J, Egozcue JJ. Compositional data analysis: where are we and where should we be heading?. Math Geol. 2005;37(7):829–50.

43. Zou H, Hastie T, Tibshirani R. Sparse principal component analysis. J Comput Graph Stat. 2006;15(2):265–86.

44. Maxwell O, Chukwudike CN, Chinedu OV, Valentine CO, Paul OC. Comparison of different parametric methods in handling critical multicollinearity: Monte carlo simulation study. Asian J Probab Stat. 2019;3(2):1–16.

45. Lahr DJ, Katz LA. Reducing the impact of pcr-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity dna polymerase. Biotechniques. 2009;47(4):857–66.

46. Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, Bushman FD, Knight R, Kelley ST. Bayesian community-wide culture-independent microbial source tracking. Nat Methods. 2011;8(9): 761–63.

47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. Dada2: high-resolution sample inference from illumina amplicon data. Nat Methods. 2016;13(7):581–83.

48. Cao Q, Sun X, Rajesh K, Chalasani N, Gelow K, Katz B, Shah VH, Sanyal AJ, Smirnova E. Effects of rare microbiome taxa filtering on statistical analysis. Front Microbiol. 2021;11:3203.

49. Cutler A, Breiman L. Archetypal analysis. Technometrics. 1994;36(4): 338–47.

50. Mørup M, Hansen LK. Archetypal analysis for machine learning and data mining. Neurocomputing. 2012;80:54–63.

51. Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J. Comparison of different methods for univariate time series imputation in r. arXiv preprint arXiv:1510.03924. 2015.

52. Gerber GK. The dynamic microbiome. FEBS Lett. 2014;588(22):4131–39.

53. Martiny JB, Jones SE, Lennon JT, Martiny AC. Microbiomes in light of traits: a phylogenetic perspective. Science. 2015;350(6261):9323.

54. Rosipal R, Krämer N. Overview and Recent Advances in Partial Least Squares. In: Saunders C, Grobelnik M, Gunn S, Shawe-Taylor J, editors. International Statistical and Optimization Perspectives Workshop "Subspace, Latent Structure and Feature Selection". Berlin: Springer Berlin Heidelberg; 2005. p. 34–51.

55. Meng C, Zeleznik OA, Thallinger GG, Kuster B, Gholami AM, Culhane AC. Dimension reduction techniques for the integrative analysis of multi-omics data. Brief Bioinform. 2016;17(4):628–41.

56. Good P. Permutation Tests: a Practical Guide to Resampling Methods for Testing Hypotheses. New York: Springer-Verlag; 2013.

57. Haynes W. Wilcoxon rank sum test. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. Encyclopedia of Systems Biology. New York: Springer New York; 2013. p. 2354–55. https://doi.org/10.1007/978-1-4419-9863-7_1185.

58. Dmitrienko A, Tamhane AC. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry. 2007;6(3):171–80.

59. Dmitrienko A, Tamhane AC. Gatekeeping procedures in clinical trials. In: Multiple Testing Problems in Pharmaceutical Statistics. London: Chapman and Hall/CRC; 2009. p. 183–210.

60. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 1992;46(3):175–85.

61. Wang C, Hu J, Blaser MJ, Li H. Estimating and testing the microbial causal mediation effect with high-dimensional and compositional microbiome data. Bioinformatics. 2020;36(2):347–55.

62. Tvedebrink T. dirmult: Estimation in dirichlet-multinomial distribution. R Package Version 0.1; 2009, p. 3.

63. Reynolds AP, Richards G, de la Iglesia B, Rayward-Smith VJ. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J Math Model Algoritm. 2006;5(4):475–504.

Publisher's Note