Phylogeny-Guided Microbiome OTU-Specific Association Test (POST)

Caizhi Huang

North Carolina State University <https://orcid.org/0000-0002-0986-920X>

Benjamin John Callahan

North Carolina State University

Michael C Wu

Fred Hutchinson Cancer Research Center

Shannon T. Holloway

North Carolina State University

Hayden Brochu

North Carolina State University

Wenbin Lu

North Carolina State University

Xinxia Peng

North Carolina State University

Jung-Ying Tzeng (✉ jytzeng@ncsu.edu)

North Carolina State University <https://orcid.org/0000-0002-5505-1775>

Methodology

Keywords: Association test, Phylogenetic tree, Kernel machine regression

Posted Date: November 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1017592/v1>

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

[Read Full License](#)

Phylogeny-Guided Microbiome OTU-Specific Association Test (POST)

Caizhi Huang¹, Benjamin J. Callahan^{1,2}, Michael C. Wu³, Shannon T. Holloway⁴, Hayden Brochu^{1,5}, Wenbin Lu⁴, Xinxia Peng^{1,5} and Jung-Ying Tzeng^{1,4*}

*Correspondence:

jytzeng@ncsu.edu

¹Bioinformatics Research Center,
North Carolina State University,
Raleigh, 27606, USA
Full list of author information is
available at the end of the article

Abstract

Background: The relationship between host conditions and microbiome profiles, typically characterized by operational taxonomic units (OTUs), contains important information about the microbial role in human health. Traditional association testing frameworks are challenged by the high-dimensionality and sparsity of typical microbiome profiles. Incorporating phylogenetic information is often used to address these challenges with the assumption that evolutionarily similar taxa tend to behave similarly. However, this assumption may not always be valid due to the complex effect of microbes, and phylogenetic information should be incorporated in a *data-supervised* fashion.

Results: In this work, we propose a local collapsing test called **Phylogeny-guided microbiome OTU-Specific association Test (POST)**. In POST, whether or not to borrow information and how much information to borrow from the neighboring OTUs in the phylogenetic tree are supervised by phylogenetic distance and the outcome-OTU association. POST is constructed under the kernel machine framework to accommodate complex OTU effects and extends kernel machine microbiome tests from community-level to OTU-level. Using simulation studies, we showed that when the phylogenetic tree is informative, POST has better performance than existing OTU-level association tests. When the phylogenetic tree is not informative, POST achieves similar performance as existing methods. Finally, we show that POST can identify more outcome-associated OTUs that are of biological relevance in real data applications on bacterial vaginosis and on preterm birth.

Conclusions: Using POST, we show that the power of detecting associated microbiome features can be enhanced by adaptively leveraging the phylogenetic information when testing for a target OTU. We developed an user friendly R package *POSTm* which is now available at CRAN (<https://CRAN.R-project.org/package=POSTm>) for public access.

Keywords: Association test; Phylogenetic tree; Kernel machine regression

Background

The human microbiome is the collection of micro-organisms residing in the human body and has been definitely shown to impact human disease and health [1]. Multiple resources have become available, e.g., the Human Microbiome Project [2] and Integrative Human Microbiome Project [3], to characterize the microbial communities at human body and to comprehend how the human microbiota affect host clinical outcomes. One common strategy to determine microbial compositions is marker-gene sequencing, which amplifies and sequences a fingerprint gene (e.g., 16S

rRNA gene) that carries species-specific identities. The generated sequencing reads are then clustered into operational taxonomic units (OTUs) based on the sequence similarity, with a specified amount of dissimilarities allowed within each OTU [4], although the concept of amplicon sequence variant (ASV) has recently been introduced to avoid imposing the arbitrary dissimilarity thresholds and provide a finer taxonomic resolution [5]. The resulting OTUs (or ASVs) are phylogenetically related.

Comparison of OTU profiles with respect to differential conditions or clinical outcomes lends important knowledge towards understanding the microbial roles in affecting health [1, 6], and distance-based approaches are widely adopted to evaluate such association at community-level [7, 8, 9, 10, 11]. It is also of interest to identify specific OTUs or organisms as biomarkers that drive the global association, which can provide crucial insights into the microbial functions and mechanisms on diseases. For example, the correlation between *Lactobacillus crispatus* abundance and host vaginal cytokine IP-10/CXCL10 levels may help to explain the induction of preterm birth [12].

OTU-level analysis can be underpowered due to the weak signal strength of single OTUs and a large number of OTUs. Multiple studies have shown that incorporating phylogenetic structure among OTUs can increase detecting power or predictive accuracy of microbiome analysis, because phylogenetically related species or OTUs are expected to impact or respond to environmental disturbances in a similar way [13]. For example, Xiao et al.(2017)[14] introduced a false discover rate (FDR) control procedure, TreeFDR, which uses the phylogenetic tree to specify the prior correlation among the p-values of individual OTUs. TreeFDR is more powerful than the classic FDR controlling methods (e.g., the Benjamini-Hochberg (BH) FDR method) and those incorporating grouping structure [15, 16]. Xiao et al.(2018)[17] introduced a predictive method, glmmTree, based on a generalized linear mixed model that encourages the nearby OTUs in the phylogenetic tree to share similar effects, and shows that glmmTree improves prediction performance especially when associated OTUs are clustered on the tree. Kim et al.(2020)[18] introduced a genus-level collapsing association test, TMat, which first assesses association for each OTU within a genus and then uses the minimum OTU p-value to determine the genus significance.

However, it has been noted that inappropriate incorporation of phylogenetic information can degrade testing performance when phylogeny is not informative of association patterns among OTUs [19]. Furthermore, evolutionarily close OTUs can act differently and even in opposite ways. For example, two close species in *Lactobacillus* genus, i.e., *Lactobacillus iners* and *Lactobacillus crispatus*, exhibit opposite associations with preterm birth—*Lactobacillus crispatus* dominance was negatively associated with preterm birth while *Lactobacillus iners* dominance was positively associated with preterm birth [20].

To address these challenges in the incorporation of phylogenetic information, we propose the **Phylogeny-guided microbiome OTU-Specific association Test (POST)**, which boosts the detecting power of the target OTU by adaptively borrowing information from its phylogenetically close OTUs. Whether or not to borrow information and how much information to borrow from the neighboring OTUs are supervised

by phylogenetic distance and the outcome-OTU association. POST is built on a kernel machine regression framework, which can flexibly accommodate complex microbiome effects, easily adjust for covariates, and be applicable to continuous and binary outcomes. Compared to existing phylogeny-informed single-OTU methods, POST can better handle sparse OTU data and allow information collapsing from OTUs of opposite effects, because kernel methods collapse OTU information at similarity/kernel level instead of at the level of p-values or abundance counts. POST extends the current community-level kernel tests [10, 21, 11] to OTU-level test. Finally, POST is computationally efficient for OTU-wide analysis as its p-value can be analytically obtained. Through extensive numerical studies, we show that POST has favorable performance in identifying associated OTUs over existing methods in simulations and real data applications, including detecting the association of vaginal microbiome with bacterial vaginosis and with preterm birth.

Methods

Suppose that for subject i , $i = 1, \dots, n$, we observe the outcome value y_i , which can be continuous or binary; the vector of p covariates $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^\top$; and the abundance vector of M OTUs $\mathbf{z}_i = (z_{i1}, \dots, z_{iM})^\top$, after sequence processing. In matrix presentation, we have the $n \times M$ matrix of OTU abundance $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)^\top$, the $n \times p$ covariate matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$, and the outcome vector $\mathbf{y} = (y_1, y_2, \dots, y_n)^\top$.

POST uses a kernel machine regression to model the relationship between OTU m and the outcome variable:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\gamma} + h^m(\mathbf{Z}) \quad (1)$$

where $\boldsymbol{\mu} = E(\mathbf{y}|\mathbf{X}, \mathbf{Z})$, the conditional expectation of $\mathbf{y}$ given the microbiome and covariate information; $g(\cdot)$ is the link function, which can be set to be the identity function for continuous outcomes and the logistic function for binary outcomes; $\boldsymbol{\gamma}$ is the $p \times 1$ vector of covariate regression coefficients; $h^m(\cdot)$ is a smooth function in a reproducing kernel Hilbert space generated by a kernel function $k^m(\mathbf{z}_i, \mathbf{z}_j)$. Function $h^m(\cdot)$ characterizes the effect of OTU m and can be specified by $k^m(\mathbf{z}_i, \mathbf{z}_j)$ using the dual representation of a function in kernel methods, i.e., $h^m(\mathbf{z}_i) = \sum_{j=1}^n \alpha_j^m k^m(\mathbf{z}_i, \mathbf{z}_j)$ with $\alpha_1^m, \dots, \alpha_n^m$ the unknown parameters. The association between OTU m and the outcome can be evaluated by testing $H_0 : h^m(\mathbf{Z}) = \mathbf{0}$.

In POST, instead of using a global kernel as in the community-level test, we propose a OTU-specific kernel function $k^m(\cdot, \cdot)$ for OTU m to permit the local OTU test to borrow information from its neighboring OTUs in the phylogenetic tree. Figure 1 describes the key steps of POST: (1) to quantify the pairwise OTU correlation based on their phylogenetic relationship; (2) to compute the subject kernel matrix $\mathbf{K}^m$ for OTU m using OTU abundance and OTU correlation data, and (3) to evaluate the association of OTU m with the outcome variable using the local kernel test that seeks for the optimal amount of information borrowing. The detail of each step is described below.

Quantifying OTU phylogenetic correlation

The first step of POST is to compute the OTU correlation matrix based on the phylogenetic relationship. To do so, we start with constructing the phylogenetic tree with M tips based on the sequences of the M OTUs using *FastTree2* [22]. We then quantify the pairwise distance between OTUs ℓ and m (denoted by $d_{\ell m}$) by taking the sum of branch lengths from OTU ℓ to OTU m through the first common node in the tree, which can be done using the R function 'cophenetic'. Finally we convert the distance to correlation (denoted by $r_{\ell m}$) using the Gaussian function, i.e., $r_{\ell m} = \exp\left\{-\frac{d_{\ell m}^2}{c \times s}\right\}$, where s is the standard deviation (SD) of $d_{\ell m}$'s, $\ell, m \in \{1, \dots, M\}$; c is a positive data-adaptive parameter that controls how fast the OTU correlation decreases when the between-OTU distance increases. Parameter c is measured on the unit of distance SD so as to be scale-free and robust to trees generated from different tools (e.g., *FastTree2*, *hclust*). As detailed in Step 2, neighboring OTUs with non-zero $r_{\ell m}$, $\ell \neq m$, will contribute in the association test of OTU m , and $r_{\ell m}$ controls the amount of information contributed from OTU ℓ into the test via c . We illustrate in Figure 2 a typical relationship between c and $r_{\ell m}$ using the dataset of [23] considered in the simulation study. Without loss of generality, we focus on OTU 3438 and its 16 neighboring OTUs in the phylogenetic tree (i.e. the left panel of Figure 2), and show the values of $r_{\ell, 3438}$ for the neighboring OTU ℓ under different c 's. We see that smaller c (e.g., $c = 0.02$) yields a more rapid correlation decay with increasing distance, and consequently OTU m can borrow a non-negligible amount of information only from a few OTUs. In contrast, larger c (e.g., $c = 0.08$) yields a slower correlation decay with increasing distance and defines a larger "neighborhood" of OTU m from which OTU m may borrow information. When $c = 0$, $r_{mm} = 1$ and $r_{\ell m} = 0$ for all $\ell \neq m$, and the local association test becomes a strict single OTU test because only information of target OTU m is used (e.g., Figure 2, Panel of $c = 0$). When $c = \infty$, $r_{\ell m} = 1$ for all ℓ , and the local association test becomes a global, community-level test because all OTUs contribute equally into the test (e.g., Figure 2, Panel of $c = 100$).

When conducting POST association test in Step 3, we use the data to determine the optimal c value — specifically, we compute the p-values for a grid of c between 0 and $c_{\max}$, and choose the c value that gives the minimum p-value. As detailed in the simulation studies, we set the $c_{\max} = 0.05$ to ensure information sharing is limited within a small, concentrated neighborhood of the target OTU. The data-adaptive determination of c also ensures that neighboring information is shared with the target OTU only if sufficient support from the data and that no information sharing among those OTUs that do not have related effects on the outcome variable.

Computing subject kernel similarity

In kernel machine regression (1), one crucial component is to specify the local kernel function $k^m(\mathbf{z}_i, \mathbf{z}_j)$, which consequently determines the OTU effects, i.e., $h^m(\mathbf{Z}_i) = \sum_{j=1}^n \alpha_j^m k^m(\mathbf{z}_i, \mathbf{z}_j)$. Several kernel choices are available for microbiome compositions in community-level tests [10, 24]. The key is to start with a distance metric to quantify the pairwise abundance dissimilarities, store them in the $n \times n$ abundance dissimilarity matrix $\mathbf{A}$, and then transform $\mathbf{A}$ to the kernel similarity matrix $\mathbf{K}$ using Gower's matrix, i.e., $\mathbf{K} = -\frac{1}{2}(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n})\mathbf{A}^2(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n})$, where

$\mathbf{A}^2$ is the element-wise square of $\mathbf{A}$, $\mathbf{I}$ is the $n \times n$ identity matrix, and $\mathbf{1}$ is a vector of ones with length n [24]. For POST, we propose to construct a local, OTU-specific kernel $\mathbf{K}^m$ so that only local information around the target OTU is used. First, obtain the local dissimilarity matrix $\mathbf{A}^m$ by incorporating $r_{\ell m}$ into the global dissimilarity matrix $\mathbf{A}$. Next, perform Gower's transformation on $\mathbf{A}^m$ to obtain $\mathbf{K}^m \equiv \{k^m(\mathbf{z}_i, \mathbf{z}_j)\} = -\frac{1}{2}(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n})(\mathbf{A}^m)^2(\mathbf{I} - \frac{\mathbf{1}\mathbf{1}^\top}{n})$. Although such local kernel construction can be applied on arbitrary distance metrics, in POST, we choose to use the Atchison distance, a non-phylogenetic distance measure accounting for the compositional nature of abundance data, to avoid the potential overuse of phylogenetic information. Atchison distance is the Euclidean distance of centered log-ratio (CLR) transformed abundance data, and the "local" Atchison distance for OTU m between subjects i and j can be obtained by

$$\mathbf{A}_{ij}^m = \sqrt{\sum_{\ell=1}^M r_{\ell m} \times (z_{i\ell}^* - z_{j\ell}^*)^2} \quad (2)$$

where $z_{i\ell}^* = \log\left[\frac{z_{i\ell}}{G(\mathbf{z}_i)}\right]$, i.e., the CLR transformation of the original abundance count $z_{i\ell}$ for OTU ℓ , and $G(\mathbf{z}_i) = \sqrt[M]{z_{i1} \cdots z_{iM}}$ is the geometric mean of OTU abundance for subject i . A pseudo-count 0.5 is added to OTU counts table before doing CLR transformation. If proportion data is used, a pseudo-proportion 1e-6 can be added to the proportion table [25].

Performing OTU-specific association test

The association of OTU m can be detected by testing for $H_0 : h^m(\mathbf{Z}) = \mathbf{0}$ in Model (1). Such test can be constructed through the connection of kernel machine regression and generalized linear mixed models (GLMM) [26, 27]. That is, $h^m(\mathbf{Z})$ can be viewed as subject-specific random effects with $h^m(\mathbf{Z}) \sim N(\mathbf{0}, \tau \mathbf{K}^m)$, and then testing for $H_0 : h^m(\mathbf{Z}) = \mathbf{0}$ is equivalent to the variance component test of $H_0 : \tau = 0$. Under the GLMM framework, a variance component score test can be obtained as $T^{m,c} = \frac{1}{2\phi}(\mathbf{y} - \hat{\boldsymbol{\mu}}_0)^\top \mathbf{K}^m(\mathbf{y} - \hat{\boldsymbol{\mu}}_0)$, where $\hat{\boldsymbol{\mu}}_0 = g^{-1}(\mathbf{X}\hat{\boldsymbol{\gamma}})$ with $\hat{\boldsymbol{\gamma}}$ the estimated covariate coefficient under the null model: $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\gamma}$, and $\hat{\phi}$ is the estimator of the dispersion parameter under H_0 . For continuous outcome, $\hat{\boldsymbol{\mu}}_0 = \mathbf{X}\hat{\boldsymbol{\gamma}}$ and $\hat{\phi}$ equals to the estimated residual variance under H_0 . For binary outcome, $\phi = 1$ and $\hat{\boldsymbol{\mu}}_0 = \exp\{\mathbf{X}\hat{\boldsymbol{\gamma}}\}/(1 + \exp\{\mathbf{X}\hat{\boldsymbol{\gamma}}\})$. With a fixed c , $T^{m,c}$ asymptotically follows a weighted mixtures of $\chi^2_{(1)}$ distribution under H_0 , based on which the p-value, denoted by $p_{m,c}$, can be computed. Because the sample size tends to be moderate in microbiome studies, we use the small-sample distribution derived in Chen et al.(2016)[28] to compute $p_{m,c}$. Finally, as the optimal c is unknown in reality, in POST we consider a grid of $c \in \{c_1, \dots, c_J\}$, use the Cauchy combination (CC) method [29, 30] to aggregate the p-values of different c 's by computing $T^m = \sum_{j=1}^J \tan\{(0.5 - p_{m,c_j})\pi\}$, and obtain the p-value of T^m (denoted by p_m) by $p_m = \frac{1}{2} - \{\arctan(T^m/J)\}/\pi$. The CC method behaves like the minimum-p-value method as T^m is dominated by the smallest p-values; yet the CC method becomes a computationally fast alternative to the minimum p-value method, because the p-value of T^m can be analytically obtained from a Cauchy distribution, even with correlated p-values [29, 30].

Simulation study

We conduct simulation studies to evaluate the performance of POST in identifying outcome-associated OTUs. We generate OTU data based on the real dataset from the upper-respiratory-tract microbiome study [23], obtained from the R package *GuniFrac* [7]. The dataset consists of 856 OTU data from 60 subjects (from R object 'throat.otu.tab') and the OTU phylogenetic tree (from R object 'throat.tree' constructed using *FastTree* [7]). We consider 2 different simulation settings (Simulations A and B); for each simulation, we simulate OTU counts of $n = 100$ individuals following Chen and Li (2013)[24] and focus on the $M = 400$ most abundant OTUs in the original data. We model the observed OTU counts using the Dirichlet-Multinomial distribution with parameters $(\pi_1, \dots, \pi_M, \theta)$, where π_ℓ 's are the mean of OTU-proportions and θ is the over-dispersion parameter [31, 32], and obtain their maximum likelihood estimates $(\hat{\pi}_1, \dots, \hat{\pi}_M, \hat{\theta})$. Next, we generate the OTU abundance proportions $(p_1, \dots, p_M)$ from Dirichlet $(\hat{\pi}_1, \dots, \hat{\pi}_M, \hat{\theta})$; we also generate the total counts of individual i , N_i , from a negative binomial distribution with mean 10000 and size 25. Finally we generate the OTU counts of individual i $(z_{i1}, \dots, z_{iM})$ from Multinomial $(N_i, p_1, \dots, p_M)$; the resulting simulated counts are stored in the $n \times M$ OTU count matrix $\mathbf{Z}$.

Simulation A

In Simulation A, we adopt the simulation design of Xiao et al. (2017) and consider a case-control study with 50 cases (i.e., $y_i = 1$) and 50 controls (i.e., $y_i = 0$). Specifically, we use the R package *pam* to partition the 400 OTUs into 20 clusters based on the phylogenetic tree, and select causal OTUs from the 8 most abundant clusters. Given $\mathbf{Z}$ with $M = 400$ and $n = 100$, we then multiply the counts of cases with a fold change vector $[\exp(\beta_1), \dots, \exp(\beta_{400})]$, where β_m is the effect size of OTU m , and is generated from $\text{Normal}(\nu, 1)$ if OTU m is causal and 0 otherwise. We consider $|\nu|$ to be 1 (small effect size) or 2 (large effect size).

We consider five scenarios for causal OTUs as illustrated in Figure 3. Scenarios 1-3 consider the same set of causal OTUs, which form multiple "causal OTU hubs" in the phylogenetic tree, with 1 causal hub from each of the top 8 abundant clusters and 7-10 causal OTUs per causal hub. In Scenario 1, β_m 's are generated from $N(\nu, 1)$ for all causal OTUs. In Scenario 2, $\beta_m \sim N(\nu, 1)$ for half of the causal hubs and $\beta_m \sim N(-\nu, 1)$ for the other half of the causal hubs. In Scenario 3, a random half of the causal OTUs have their β_m 's generated from $N(\nu, 1)$ and the remaining half have their β_m 's from $N(-\nu, 1)$; consequently a causal hub may contain a fair number of positive-effect and negative-effect causal OTUs. Scenario 4 considers the case of smaller causal hubs with 2-3 causal OTUs per hub (i.e., less-informative trees); a random half of the causal hubs with $\beta_m \sim N(\nu, 1)$ and the other half causal hubs with $\beta_m \sim N(-\nu, 1)$. Scenario 5 considers the case where causal OTUs are randomly chosen from the entire phylogenetic tree (i.e., non-informative trees), where a random half of the causal OTUs have $\beta_m \sim N(\nu, 1)$ and the other half OTUs have $\beta_m \sim N(-\nu, 1)$.

Simulation B

In Simulation B, we consider both continuous and binary outcomes and adopt the simulation design of Zhao et al.(2015)[10] and Koh et al.(2017)[11]. Given the simulated OTU counts $\mathbf{Z} = \{z_{im}\}$ and the same set of causal OTUs as Simulation

A, we generate the outcome value of individual i by $y_i \sim N(\eta_i, 1)$ for continuous outcomes and $y_i \sim \text{Bernoulli}(\Pi_i)$ with $\Pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$ for binary outcomes. Value $\eta_i = 0.5\omega \times (\text{scale}(x_{1i}) + \text{scale}(x_{2i})) + \sum_{m=1}^M \beta_m \times \text{scale}(z_{im})$, where x_{1i} and x_{2i} are covariates; z_{im} is the count of the causal OTU m for subject i ; $\omega = 1$ or 0 and is a parameter controlling if there are covariate effects. The 'scale' function is to standardize the variable to mean 0 and standard deviation 1. Variable x_{1i} is simulated from $\text{Bernoulli}(0.5)$ and is independent of the OTU counts, while variable x_{2i} is correlated with causal OTUs by letting $x_{2i} \sim N(\delta_i, 1)$ with $\delta_i = \text{scale}(\sum_{m \in \text{causal OTUs}} z_{im})$. For non-causal OTUs, $\beta_m = 0$; for causal OTUs, $\beta_m \sim \text{Normal}(\nu, \nu/5)$ or $\text{Normal}(-\nu, \nu/5)$ dependent on the scenario. We consider $|\nu|$ to be 0.2 (small effect size) and 0.5 (large effect size) for continuous outcomes; 0.3 (small effect size) and 1 (large effect size) for binary outcomes. In each setting, we consider $\omega = 0$ (no covariate effects) and $\omega = 1$ (with covariate effects).

Determining appropriate $c_{\max}$

To determine the appropriate $c_{\max}$ for POST, we conduct 100 replications under each of the 5 causal scenarios in Simulations A and B, and use the decision rule that a OTU is selected as important if p-value < 0.05 and not important otherwise. Using this decision rule, we report the true positive rate (TPR), false positive rate (FPR) and a composite measure by taking the harmonic mean of TPR and (1-FPR), which is referred to as pseudo-F score. TPR is obtained by first computing the fraction of detected causal OTUs among all causal OTUs in each replication, and then averaging across the 100 replications; FPR is obtained by first computing the fraction of detected non-causal OTUs among all non-causal OTUs in each replication, and then averaging across the 100 replications; pseudo-F is obtained by taking the harmonic mean of the TPR and 1-FPR in each replication and then averaging across the 100 replications.

Performance assessment

To examine the validity of POST, we set $\beta_m = 0$ for all OTUs, and report the type I error rates and the quantile-quantile (Q-Q) plots of the null p-values based on 4,000 replications under Simulations A and B. To examine the performance of selecting causal OTUs, we report the area under the receiver operating characteristic (ROC) curves (AUC in short) based on the 100 replications under each simulation setting from the 5 causal scenarios. AUC can assess the performance over various selection thresholds and account for both TPR and FPR. The AUC of a method is obtained as follows. (1) In a replication, we compute the FDR-adjusted p-values of a method, based on which we then compute the FPR and TPR at varying significant thresholds. (2) We take the average FPR and average TPR across the 100 replications at each of the significant thresholds. (3) We construct the ROC curve using the average FPRs and TPRs, and compute the corresponding AUC. We use the two-stage BH (TSBH) FDR procedure [33] to compute the FDR-adjusted p-values for all methods except for TF, which directly gives FDR-adjust p-values. TSBH has been shown to yield good performance for correlated p-values [34] and is implemented in the R function 'mt.rawp2adjp' of the R package *multtest*.

We compare POST with 5 baseline approaches: (1) TreeFDR (TF) of Xiao et al. (2017)[14], using the R function 'MicrobiomeSeqTreeFDR' in the R package

StructFDR; (2) single-OTU test (SO), by setting $c = 0$ in POST; (3) DESeq2 (DE) of Love et al. (2017)[35], using the R function 'DESeq' with the default Wald test in the R package *DeSeq2*; (4) Wilcoxon rank-sum test on the proportional data (WR-P) and on the CLR transformed data (WR-R), using the R function 'wilcox.test'. Wilcoxon test is only conducted for binary outcomes with no covariate effects. (5) Spearman correlation test on proportional data (SC-P) and on the CLR transformed data (SC-R), using the R function 'cor.test'. Spearman test is only conducted for continuous outcomes with no covariate effects. Among these method, POST and TF incorporate phylogenetic tree information. POST, SO, WR-R and SC-R use the CLR transformation to address the compositional nature of microbiome data.

Results

Simulation results

Appropriate $c_{\max}$

To identify the appropriate value of $c_{\max}$, the upper bound of c , we examine the trajectory of pseudo-F (Addition file 1: Figure S1) over different values of $c_{\max}$, ranging from 0 (which corresponds to zero OTU neighbors) and 0.1 (which corresponds to a relatively large number of OTU neighbors) as shown in Figure 2. An ideal $c_{\max}$ should permit information borrowing only from OTUs within a localized neighborhood in the phylogenetic tree. In addition, because incorporating too much information from irrelevant phylogenetic neighboring OTUs would lead to undesirable TPRs and FPRs, the ideal $c_{\max}$ is expected to occur at some turning point with which the increment of pseudo-F trajectory stops or slows down. Factoring in these two criteria, we find that $c_{\max} = 0.05$ appears to offer a robust choice across different scenarios in Simulations A and B. We hence implement POST with $c_{\max} = 0.05$ hereafter.

Validity of POST

We check the behavior of POST under the global null hypothesis of no causal OTUs. Table 1 shows that the type I error rates is appropriately controlled at nominal levels of 0.05, 0.01, and 0.001. Additional file 2: Figure S2 shows the corresponding Q-Q plots of POST p-values, where the observed POST p-values agree with the expected p-values from Uniform (0,1), confirming the validity of POST.

Selection Performance

Next we examine the relative performance of POST against the baseline methods using AUC (Table 2 for Simulation A and B without covariate effects; Additional file 3: Table S1 for Simulation B with covariate effects). In simulation A, we observe that POST and TF have larger AUC than those methods that do not use tree information when causal OTUs are clustered in large hubs (Scenarios 1-3) or small hubs (Scenario 4). Furthermore, POST has larger AUC than TF, suggesting that POST can further boost power through adaptive information collapsing from evolutionarily close OTUs at the level of OTU abundances, compared to TF which collapses information at the level of OTU p-values. The superior performance of POST is consistently observed regardless of whether the effects of nearby causal OTUs are in the same direction (Scenarios 1 and 2) or opposite directions (Scenarios 3 and

4), and whether the effect sizes are large or small. In scenario 5 where causal OTUs are randomly distributed in the phylogenetic tree, POST has similar performance with other methods. Among the non-tree based methods, SO and DE have similar performance, and are better than the Wilcoxon rank-sum test. The Wilcoxon rank-sum test using proportional data have slightly higher AUC than using CLR transformed data.

In simulation B, POST has the largest AUC or similar AUC to the best method in all scenarios, regardless of the effect sizes, without covariates (Table 2) or with covariates (Additional File 3: Table S1). Besides POST, DE often has the best performance even in some of Scenarios 1-4 although it does not incorporate tree information. We also observe that TF has AUC around 0.5 in most scenarios. A closer examination suggests that this may be because the signal strengths in Simulation B are much smaller than those in Simulation A, based on the way that the data were simulated. Consequently, TF yields many large, tied FDR adjusted p-values in Simulation B, while other methods yield less tied adjusted p-values with relatively smaller in values than TF's.

In summary, across different simulation scenarios, POST has better or comparable performance compared to the baseline methods. POST tends to have the largest AUC among all methods (e.g., in Scenarios 1-4 of informative trees); if not (e.g., in Scenario 5 of non-informative trees), POST has its AUC comparable to the top methods, which can either be TF, DE or SO, depending on simulation settings.

Real data analysis

Association study of vaginal microbiome and bacterial vaginosis

Bacterial vaginosis (BV) is a type of vaginal inflammation and is characterized by low level of lactobacilli and overgrowth of various anaerobic bacteria [36]. BV are most often diagnosed using Amsel criteria or microscopy based Nugent scoring [37] but with limited accuracy [38]. High-throughput sequencing technologies such as 16S rRNA amplicon sequencing have been used to study the species diversity of vaginal microbiota and shown promise as an alternative assessment of BV for practical and accurate diagnosis [39, 40, 41].

We conduct OTU-specific analysis using the 16S rRNA gene sequencing dataset of vaginal microbiome from [41] to evaluate the association between BV and the vaginal microbiome. The dataset consists 39 individuals, and the sequencing data and metadata are publicly available at NCBI SRA database (PRJNA600021). Initial processing leads to the abundance data of 2711 OTUs formed at 97% similarity. We employ the *FastTree2* [22] to infer the phylogenetic tree using the OTU sequences. Finally, we filter out OTUs with abundance <0.005% and prevalence <10%, and analyze the resulting 186 OTUs of 39 individuals for BV association using POST, TF, SO, DE, WR-P and WR-R. In the analysis of POST, TF, SO and DE, we include race as a covariate. For POST, SO, DE, WR-P and WR-R, the TSBH procedure [33] is used to compute the FDR-adjusted p-values. We select important OTUs using the decision rule of FDR-adjusted p-values < 0.05.

The Upset plot (Figure 5A) shows that POST, TF, SO, DE, WR-P and WR-R identified 7, 2, 1, 5, 3 and 1 significant OTUs, respectively. Table 3 lists the OTUs detected by each method and their mapped genus/species with 100% sequence identity among vaginal microbiome. Among the 7 POST-detected OTUs,

OTU3 (mapped to *Lactobacillus crispatus*) is also identified by SO, DE and WR-C. Depletion of *Lactobacillus crispatus* has been shown to be highly associated with BV [42]. POST also identifies 6 additional OTUs as important (among which OTU66 is also identified by DE); these OTUs are all mapped to *Lactobacillus* species, including *Lactobacillus jensennii*, *Lactobacillus gasseri* and *Lactobacillus iners*. These vaginal *Lactobacillus* species have been shown to have different abundance in women with vs. without BV [42] and play important roles for vaginal ecosystem [43]. On the other hand, TF identifies two OTUs, i.e., OTU112 (mapped to *Peptoniphilus* sp. and also identified by WR-P) and OTU85 (mapped to *Gemella* sp.). Although the 2 OTUs are not detected by POST, SO, DE or WR-R, these genera have also been detected in BV cases [44, 45]. Besides the OTUs discussed above, DE and WR-P also identify 4 other OTUs (i.e., OTU11, OTU12, OTU16 and OTU91), which are all mapped to Genus *Prevotella*, and OTU16 (*Prevotella timonensis*) has been found associated with BV [46]. It seems that POST, SO and WR-R (i.e., methods based on CLR-transformed data) miss some of the OTUs identified by TF, DE and WR-P (i.e., methods based on proportional data), although there are also OTUs detected by both types of methods. Finally, Figure 5B shows the relative positions of the identified OTUs in the phylogenetic tree, where POST identified OTUs tend to be more clustered together compared to other methods.

We further illustrate the data-driven information borrowing of POST using OTU2 (average proportional abundance 0.22, mapped to *Lactobacillus iners*, and only identified by POST) and OTU3 (average proportional abundance 0.26, mapped to *Lactobacillus crispatus*, and identified by POST, SO, DE and WR-C). The abundance boxplots (Additional file 4: Figure S3) show that OTU2 has higher abundance in BV patients compared to healthy women, while OTU3 has an opposite pattern. Our findings agree with the literature that BV patients have loss of many *Lactobacillus* species except *Lactobacillus iners* [42, 47]. When testing for OTU2, the best c determined by the data is $c = 0.03$, suggesting some information borrowing from neighboring OTUs (Additional file: Figure S4A), e.g., OTU7 has $r_{7,2} = 0.74$ and OTU3 has $r_{3,2} = 0.52$ in Equation (2). This example illustrates that (i) POST can incorporate information with opposite directions, and (ii) the information borrowing is OTU-specific and data-adaptive — when testing for OTU3, the best c is $c = 0$, which means no information borrowing from any neighboring OTUs (Additional file: Figure S4B).

Association study of vaginal microbiome and preterm birth

Preterm birth is a major cause of neonatal morbidity and mortality. Previous studies have suggested that the risk of preterm birth is associated with vaginal microbiota composition, especially certain species/genera including *Lactobacillus crispatus*, *Lactobacillus iners*, *BVAB1*, *Sneathia amnii*, and some *Prevotella* species [48, 12, 20]. We perform OTU-specific analysis on the Stanford cohort data of Callahan et al. (2017) [48] to identify OTUs associated with preterm birth. We apply the same data processing steps as in the BV study and obtain 746 OTUs and 39 individuals. After excluding OTUs with the abundance $< 0.005\%$ and the present rate $< 10\%$, we base our analysis on the 95 remaining OTUs. Association analysis are performed using POST, TF, SO, and DE (including race as a covariate), as

well as WR-P and WR-C (without adjusting for covariates). For POST, SO, DE, WR-P and WR-C, the TSBH procedure [33] is used to compute the FDR-adjusted p-values.

Figure 6A shows that POST, TF, SO DE, WR-P and WR-R identified 3, 2, 1, 11, 10 and 2 significant OTUs, respectively. Additional file 6: Table S2 shows the significant OTUs, their detection methods and the corresponding genus/species. Figure 6B shows the relative positions of the identified OTUs in the phylogenetic tree. The 3 POST-identified OTUs have also been detected by other methods, i.e., OTU131 (mapped to *Prevotella* sp.), which is also found significant by SO, DE, WR-P and WR-R; OTU40 (mapped to *Prevotella melaninogenica*), which is also identified by DE; and OTU153 (mapped to *Neisseria* sp.), which is also identified by TF, DE, WR-P and WR-R. *Prevotella* genus has been reported to be associated with preterm birth [12]; OTU153 has 97.45% (229/235) sequence identity with *Neisseria gonorrhoeae*, which causes one type of sexually transmitted infections and has been reported associated with preterm birth [49, 50]. Besides OTU153, TF also identifies OTU72 (mapped to *Haemophilus parainfluenzae* and also identified by WR-P). *Haemophilus parainfluenzae* has been reported to play a potential role in preterm birth [51]. For DE, besides the 3 OTUs identified by both POST and DE, DE also identify 8 additional OTUs as listed in Additional file 7: Table S2. Among these OTUs, OTU2 (mapped to *Lactobacillus crispatus*) and OTU8 (mapped to *Prevotella* sp.) have been found associated with preterm birth in the literature [20, 12]. OTU4 (mapped to Veillonellaceae bacterium), OTU71 (mapped to *Alloscardovia omnicolens*) and OTU19 (mapped to *Staphylococcus aureus*) have also be discussed in the literature although they are reported as not associated with preterm birth [52, 53]. We could not find much literature discussing the association between preterm birth and OTU114 (mapped to *Lacticaseibacillus rhamnosus*), OTU126 (mapped to *Cloacibacterium* sp., also identified by WR-P), or OTU15 (mapped to *Mycoplasma hominis*). Finally, WR-P also identifies 6 additional unique OTUs (see Additional file 6: Table S2 for detail). Except for OTU44 (mapped to *Fusobacterium nucleatum*), which has been reported associated with preterm birth [54], other OTUs have not been discussed much in the literature and the associations diminish if switching to WR-R.

A final note is that the two data applications show different relative performance of POST compared to the baseline methods. In the BV study, POST detects several significant unique OTUs that are not identified by other methods. We discuss the biological relevance of these identified OTUs, and examine the phylogenetic relationship among the identified OTUs (Figure 5B). The results suggest that these OTUs tend to be evolutionarily close and their signals might be easier to be detected by borrowing appropriate amount information from appropriate neighboring OTUs. In contrast, in preterm birth study, POST only identified OTUs that are also identified by other non-tree based methods. In particular, POST has similar performance to SO and WR-R, suggesting the compatibility of POST to the non-tree based methods on CLR transformed data and the robustness of POST to non-informative phylogeny.

Discussion and conclusions

In this work, we proposed a phylogeny-guided OTU-specific test, POST. POST extends the kernel-based approaches from community-level analysis to OTU-level analysis, and evaluates the association of a focal OTU based on the abundance of itself as well as its phylogenetically close neighbors. POST allows closer neighboring OTUs to contribute more than distant neighbors, and includes the standard single OTU tests as a special case (i.e., zero contribution from all neighboring OTUs). The actual contribution from neighboring OTUs is adaptively determined by data, using a tuning parameter c to identify the appropriate OTUs and the optimal amount to borrow information from. The best c is the one yielding the smallest p-value of the kernel test, whose p-value is then calculated using the Cauchy combination method [30]. We evaluate the performance of POST in different simulated scenarios and in real data applications; we see that POST has comparable or superior performance compared to commonly used tree-free methods (e.g., SO, DESeq, Wilcoxon rank-sum test and Spearman correlation test) and TreeFDR, a recently developed tree-guided procedure. Compared to these baseline methods, POST tends to have higher AUC when the phylogenetic tree is informative; and have similar AUC when the phylogenetic tree is not informative.

Here we describe POST based on OTUs formed at $\leq 3\%$ sequence dissimilarity and the corresponding phylogenetic tree constructed using *FastTree2* [22]. Although our specific testing results are restricted to this framework, the POST method is relevant to a broader class of metagenomic sequencing methods including marker-gene and shotgun sequencing methods where reads are assigned taxonomy against a reference database with a pre-calculated phylogeny. However, there are a couple points to keep in mind when extending the described framework to these more general scenarios. Recall that for a target OTU, its neighborhood and the amount of information to borrow are controlled via c . Because c is measured on the scale of the SD of pairwise OTU distances in the phylogenetic tree, c does not depend on the actual values of the OTU distance in a tree and hence POST can be applied on phylogenetic trees from different tools. For different taxonomic levels, some modifications may be needed because the upper bound of c , $c_{\max}$, which controls the neighborhood boundary of an OTU, may or may not always be generalizable to OTUs defined at different level. For example, it would be dangerous to directly apply the current $c_{\max}$ to OTUs formed at $>3\%$ dissimilarity or higher taxonomic ranks (e.g., genus) because the current $c_{\max}$ may lead to too large a neighborhood in the tree for the focal OTU. On the other hand, the current $c_{\max}$ can be used for OTUs formed at ASV levels (i.e., 0% dissimilarity), with the price of possibly having a smaller OTU neighborhood than the actual optimal neighborhood to borrow information from. When needed, the proper $c_{\max}$ can be explored and determined using a similar simulation procedure as conducted in the paper so to assure that only OTUs from a meaningful, localized neighborhood would contribute to the association assessment of the focal OTU.

In POST we build the kernel matrix based on the non-phylogenetic Aitchison distance to account for the compositional nature of microbiome data [55]. It is possible to construct local kernel using other distance metrics, such as the UniFrac distance family. For example, we can compute the local weighted UniFrac (WU) distance

for OTU m between subjects i and j by $WU_{ij}^m = \frac{\sum_{t=1}^T b_t |[r_m]_t^\top (q_t^i - q_t^j)|}{\sum_{t=1}^T b_t |[r_m]_t^\top (q_t^i + q_t^j)|}$ where b_t is the length of branch t and T is the total number of branches. Assuming that there are n_t OTUs connected to branch t in the phylogenetic tree, then q_t^i is a length- n_t vector recording the abundance proportions of the n_t OTUs connected to branch t for subject i ; $[r_m]_t$ is a length- n_t vector recording the correlation between OTU m and those OTUs connected to branch t as obtained in Section 2.1 and dependent on c . The above equation quantifies the "local" weighted UniFrac distance for OTU m by weighting the proportion difference of the branch- t -related OTUs according to their correlations with OTU m . If r_m equals $\mathbf{1}$, Above equation is the global weighted UniFrac distance. However, in these phylogenetic-distances, the phylogenetic information seems to be used twice, one embedded in the original distance definition and one in the use of r_m . Further examinations are needed to understand the pros and cons of local tests on phylogenetic-based distance.

Acknowledgements

We thank Dr. Michael McLaren for his helpful discussions on this work.

Funding

This work has been partially supported by National Institute of Health Grants T32ES007329 (to CH), R35GM133745 (to CH and BC), P01CA142538 (to STH, WL and JYT), RF1AG074328 (to STH, WL and JYT), R01GM129512 (to MW), R01HL155417 (to MW), R21AI120713 (to HB and XP), U19AI144181 (subcontract to HB and XP), and National Institute of Health Contract HHSN272201800008C (subcontract to XP and HB).

Abbreviations

POST: Phylogeny-guided microbiome OTU-specific association test; OTU: Operational taxonomic units; ASV: Amplicon sequence variant; FPR: False positive rate; TPR: True positive rate; AUC: Area under the receiver operating characteristic curves; TSBH: Two-stage Benjamini–Hochberg procedure; TF: TreeFDR; DE: DESeq2; WR-P: Wilcoxon rank-sum test using proportional data; WR-R: Wilcoxon rank-sum test using CLR transformed data; SO: Single-OTU test; BV: Bacterial vaginosis; WU: weighted UniFrac.

Availability of data and materials

The dataset of upper-respiratory-tract microbiome study [23] used in our simulation study is available in the R package *GUniFrac*. The 16S rRNA sequencing dataset and metadata of the vaginal microbiome and bacterial vaginosis study [41] used in the real data analysis are publicly available at NCBI SRA database with project number PRJNA600021. The processed sequencing data and metadata of the vaginal microbiome and preterm birth [48] used in the real data analysis are publicly available at the the Stanford Digital Repository (<https://purl.stanford.edu/yb681vm1809>). Our POST method is implemented at the R package *POSTm* and it is now available at CRAN (<https://CRAN.R-project.org/package=POSTm>) for public access. R codes used for simulation, real data analysis and data visualization are available at the GitHub by request.

Ethics approval and consent to participate

No ethics approval or consent to participate was required.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

No consent for publication was required.

Authors' contributions

CH, MCW, WL and JYT conceived the presented ideas and study design. CH implemented the methods and performed the numerical studies under the supervision of BC, MCW, XP and JYT. CH, BC, HB, XP and JYT contributed to data processing, analysis and result interpretation of the real data analysis. CH and STH created the R package with input from BC, MCW, HB, WL, XP and JYT. CH, BC and JYT draft the manuscript with input from MCW, STH, HB, WL and XP. All authors helped shape the research, discussed the results and contributed to the final manuscript.

Author details

¹Bioinformatics Research Center, North Carolina State University, Raleigh, 27606, USA. ²Department of Population Health and Pathobiology, North Carolina State University, Raleigh, 27607, USA. ³Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, 98109, USA. ⁴Department of Statistics, North Carolina State University, Raleigh, 27606, USA. ⁵Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, 27607, USA.

References

1. Cho, I., Blaser, M.J.: The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* **13**(4), 260–270 (2012)
2. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., Gordon, J.I.: The human microbiome project. *Nature* **449**(7164), 804–810 (2007)
3. iHMP Research Network Consortium, I.H., et al.: The integrative human microbiome project. *Nature* **569**, 641–648 (2019)
4. Schloss, P.D., Handelsman, J.: Introducing dotur, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**(3), 1501–1506 (2005)
5. Callahan, B.J., McMurdie, P.J., Holmes, S.P.: Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal* **11**(12), 2639 (2017)
6. Alekseyenko, A.V., Perez-Perez, G.I., De Souza, A., Strober, B., Gao, Z., Bihan, M., Li, K., Methé, B.A., Blaser, M.J.: Community differentiation of the cutaneous microbiota in psoriasis. *Microbiome* **1**(1), 31 (2013)
7. Chen, J., Bittinger, K., Charlson, E.S., Hoffmann, C., Lewis, J., Wu, G.D., Collman, R.G., Bushman, F.D., Li, H.: Associating microbiome composition with environmental covariates using generalized unifracs distances. *Bioinformatics* **28**(16), 2106–2113 (2012)
8. Fukuyama, J., McMurdie, P.J., Dethlefsen, L., Relman, D.A., Holmes, S.: Comparisons of distance methods for combining covariates and abundances in microbiome studies. In: *Biocomputing 2012*, pp. 213–224. World Scientific, ??? (2012)
9. Tang, Z.-Z., Chen, G., Alekseyenko, A.V.: Permanova-s: association test for microbial community composition that accommodates confounders and multiple distances. *Bioinformatics* **32**(17), 2618–2625 (2016)
10. Zhao, N., Chen, J., Carroll, I.M., Ringel-Kulka, T., Epstein, M.P., Zhou, H., Zhou, J.J., Ringel, Y., Li, H., Wu, M.C.: Testing in microbiome-profiling studies with mirkat, the microbiome regression-based kernel association test. *The American Journal of Human Genetics* **96**(5), 797–807 (2015)
11. Koh, H., Blaser, M.J., Li, H.: A powerful microbiome-based association test and a microbial taxa discovery framework for comprehensive association mapping. *Microbiome* **5**(1), 45 (2017)
12. Fettweis, J.M., Serrano, M.G., Brooks, J.P., Edwards, D.J., Girerd, P.H., Parikh, H.I., Huang, B., Arodz, T.J., Edupuganti, L., Glascock, A.L., et al.: The vaginal microbiome and preterm birth. *Nature medicine* **25**(6), 1012–1021 (2019)
13. Martiny, J.B., Jones, S.E., Lennon, J.T., Martiny, A.C.: Microbiomes in light of traits: a phylogenetic perspective. *Science* **350**(6261), 9323 (2015)
14. Xiao, J., Cao, H., Chen, J.: False discovery rate control incorporating phylogenetic tree increases detection power in microbiome-wide multiple testing. *Bioinformatics* **33**(18), 2873–2881 (2017)
15. Hu, J.X., Zhao, H., Zhou, H.H.: False discovery rate control with groups. *Journal of the American Statistical Association* **105**(491), 1215–1227 (2010)
16. Yekutieli, D.: Hierarchical false discovery rate-controlling methodology. *Journal of the American Statistical Association* **103**(481), 309–316 (2008)
17. Xiao, J., Chen, L., Johnson, S., Yu, Y., Zhang, X., Chen, J.: Predictive modeling of microbiome data using a phylogeny-regularized generalized linear mixed model. *Frontiers in microbiology* **9**, 1391 (2018)
18. Kim, K.J., Park, J., Park, S.-C., Won, S.: Phylogenetic tree-based microbiome association test. *Bioinformatics* **36**(4), 1000–1006 (2020)
19. Bichat, A., Plassais, J., Ambroise, C., Mariadassou, M.: Incorporating phylogenetic information in microbiome differential abundance studies has no effect on detection power and fdr control. *Frontiers in Microbiology* **11**, 649 (2020)
20. Kindinger, L.M., Bennett, P.R., Lee, Y.S., Marchesi, J.R., Smith, A., Cacciatore, S., Holmes, E., Nicholson, J.K., Teoh, T., MacIntyre, D.A.: The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. *Microbiome* **5**(1), 6 (2017)
21. Wu, C., Chen, J., Kim, J., Pan, W.: An adaptive association test for microbiome data. *Genome medicine* **8**(1), 56 (2016)
22. Price, M.N., Dehal, P.S., Arkin, A.P.: Fasttree 2—approximately maximum-likelihood trees for large alignments. *PloS one* **5**(3) (2010)
23. Charlson, E.S., Chen, J., Custers-Allen, R., Bittinger, K., Li, H., Sinha, R., Hwang, J., Bushman, F.D., Collman, R.G.: Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PloS one* **5**(12) (2010)
24. Chen, J., Li, H.: Kernel methods for regression analysis of microbiome compositional data. In: *Topics in Applied Statistics*, pp. 191–201. Springer, ??? (2013)
25. Plantinga, A.M., Chen, J., Jenq, R.R., Wu, M.C.: pldist: ecological dissimilarities for paired and longitudinal microbiome association analysis. *Bioinformatics* **35**(19), 3567–3575 (2019)
26. Liu, D., Lin, X., Ghosh, D.: Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* **63**(4), 1079–1088 (2007)
27. Liu, D., Ghosh, D., Lin, X.: Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. *BMC bioinformatics* **9**(1), 292 (2008)
28. Chen, J., Chen, W., Zhao, N., Wu, M.C., Schaid, D.J.: Small sample kernel association tests for human genetic and microbiome association studies. *Genetic epidemiology* **40**(1), 5–19 (2016)
29. Liu, Y., Xie, J.: Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association* **115**(529), 393–402 (2020)
30. Liu, Y., Chen, S., Li, Z., Morrison, A.C., Boerwinkle, E., Lin, X.: Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *The American Journal of Human Genetics* **104**(3), 410–421 (2019)
31. Mosimann, J.E.: On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika* **49**(1/2), 65–82 (1962)

32. Tvedebrink, T.: Overdispersion in allelic counts and θ -correction in forensic genetics. *Theoretical population biology* **78**(3), 200–210 (2010)
33. Benjamini, Y., Krieger, A.M., Yekutieli, D.: Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**(3), 491–507 (2006)
34. Stevens, J.R., Al Masud, A., Suyundikov, A.: A comparison of multiple testing adjustment methods with block-correlation positively-dependent tests. *Plos one* **12**(4) (2017)
35. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for rna-seq data with *deseq2*. *Genome biology* **15**(12), 1–21 (2014)
36. Sobel, J.D.: Bacterial vaginosis. *Annual review of medicine* **51**(1), 349–356 (2000)
37. Gutman, R.E., Peipert, J.F., Weitzen, S., Blume, J.: Evaluation of clinical methods for diagnosing bacterial vaginosis. *Obstetrics & Gynecology* **105**(3), 551–556 (2005)
38. Kahwati, L.C., Clark, R., Berkman, N., Urrutia, R., Patel, S.V., Zeng, J., Viswanathan, M.: Screening for bacterial vaginosis in pregnant adolescents and women to prevent preterm delivery: updated evidence report and systematic review for the us preventive services task force. *Jama* **323**(13), 1293–1309 (2020)
39. Dols, J.A., Molenaar, D., van der Helm, J.J., Caspers, M.P., de Kat Angelino-Bart, A., Schuren, F.H., Speksnijder, A.G., Westerhoff, H.V., Richardus, J.H., Boon, M.E., et al.: Molecular assessment of bacterial vaginosis by lactobacillus abundance and species diversity. *BMC infectious diseases* **16**(1), 180 (2016)
40. Vitali, B., Cruciani, F., Picone, G., Parolin, C., Donders, G., Laghi, L.: Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis. *European Journal of Clinical Microbiology & Infectious Diseases* **34**(12), 2367–2376 (2015)
41. Subramaniam, A., Kumar, R., Cliver, S.P., Zhi, D., Szychowski, J.M., Abramovici, A., Biggio, J.R., Lefkowitz, E.J., Morrow, C., Edwards, R.K.: Vaginal microbiota in pregnancy: evaluation based on vaginal flora, birth outcome, and race. *American journal of perinatology* **33**(04), 401–408 (2016)
42. Srinivasan, S., Fredricks, D.N.: The human vaginal bacterial biota and bacterial vaginosis. *Interdisciplinary perspectives on infectious diseases* **2008** (2008)
43. Petrova, M.I., Lievens, E., Malik, S., Imholz, N., Lebeer, S.: Lactobacillus species as biomarkers and agents that can promote various aspects of vaginal health. *Frontiers in physiology* **6**, 81 (2015)
44. Diop, K., Diop, A., Michelle, C., Richez, M., Rathored, J., Bretelle, F., Fournier, P.-E., Fenollar, F.: Description of three new peptoniphilus species cultured in the vaginal fluid of a woman diagnosed with bacterial vaginosis: *Peptoniphilus pacaensis* sp. nov., *peptoniphilus raoultii* sp. nov., and *peptoniphilus vaginalis* sp. nov. *MicrobiologyOpen* **8**(3), 00661 (2019)
45. Coleman, J.S., Gaydos, C.A.: Molecular diagnosis of bacterial vaginosis: an update. *Journal of Clinical Microbiology* **56**(9), 00342–18 (2018)
46. van Teijlingen, N.H., Helgers, L.C., Zijlstra-Willems, E.M., van Hamme, J.L., Ribeiro, C.M., Strijbis, K., Geijtenbeek, T.B.: Vaginal dysbiosis associated-bacteria *megaspheera elsdonii* and *prevotella timonensis* induce immune activation via dendritic cells. *Journal of reproductive immunology* **138**, 103085 (2020)
47. Fredricks, D.N., Fiedler, T.L., Marrazzo, J.M.: Molecular identification of bacteria associated with bacterial vaginosis. *New England Journal of Medicine* **353**(18), 1899–1911 (2005)
48. Callahan, B.J., DiGiulio, D.B., Goltsman, D.S.A., Sun, C.L., Costello, E.K., Jeganathan, P., Biggio, J.R., Wong, R.J., Druzyn, M.L., Shaw, G.M., et al.: Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of us women. *Proceedings of the National Academy of Sciences* **114**(37), 9966–9971 (2017)
49. Pararas, M., Skevaki, C., Kafetzis, D.: Preterm birth due to maternal infection: causative pathogens and modes of prevention. *European Journal of Clinical Microbiology and Infectious Diseases* **25**(9), 562–569 (2006)
50. Choi, S.J., Park, S.D., Jang, I.H., Uh, Y., Lee, A.: The prevalence of vaginal microorganisms in pregnant women with preterm labor and preterm birth. *Annals of laboratory medicine* **32**(3), 194–200 (2012)
51. Mendz, G.L., Petersen, R., Quinlivan, J.A., Kaakoush, N.O.: Potential involvement of *campylobacter curvus* and *haemophilus parainfluenzae* in preterm birth. *Case Reports* **2014**, 2014205282 (2014)
52. Tabatabaei, N., Eren, A., Barreiro, L., Yotova, V., Dumaine, A., Allard, C., Fraser, W.: Vaginal microbiome in early pregnancy and subsequent risk of spontaneous preterm birth: a case-control study. *BJOG: An International Journal of Obstetrics & Gynaecology* **126**(3), 349–358 (2019)
53. Son, K.-A., Kim, M., Kim, Y.M., Kim, S.H., Choi, S.-J., Oh, S.-y., Roh, C.-R., Kim, J.-H.: Prevalence of vaginal microorganisms among pregnant women according to trimester and association with preterm birth. *Obstetrics & gynecology science* **61**(1), 38–47 (2018)
54. Han, Y.W., Redline, R.W., Li, M., Yin, L., Hill, G.B., McCormick, T.S.: *Fusobacterium nucleatum* induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. *Infection and immunity* **72**(4), 2272–2279 (2004)
55. Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J.: Microbiome datasets are compositional: and this is not optional. *Frontiers in microbiology* **8**, 2224 (2017)

Figures

Figure 1 Overview of Phylogeny-guided microbiome OTU-specific association test (POST)

Tables

Additional Files

Additional file 1 — Figure S1.

Pseudo-F plots from Simulations A and B with large effect size (left) and small effect size (right) of causal OTUs for all five scenarios. Simulation A considers binary outcomes (1st row), which were simulated assuming no covariate

Figure 2 OTU phylogenetic correlation $r_{\ell m}$ under different values of the tuning parameter c , illustrated by setting $m=OTU3438$ as the target OTU. The x-axis is the OTUs shown in the same order as the tree tips. The y-axis is the correlation between OTU3438 and the remaining OTUs.

Figure 3 Illustration of the five causal-OTU scenarios considered in the simulation. Scenarios 1 to 3 consider larger "causal hubs", each containing about 7-10 causal OTUs; Scenario 4 considers smaller causal hubs of 2-3 causal OTUs; Scenario 5 considers causal OTUs with random positions in the phylogenetic tree. Red (blue) circles indicate that causal effect size tends to be positive (negative).

Figure 4 ROC curves of Simulation A with large OTU effect size (top row) and small OTU effect size (bottom row) for POST, Single-OTU test (SO), TreeFDR (TF), DESeq2 (DE) and Wilcoxon rank-sum test (WR) under the 5 causal scenarios. Scenarios 1 to 3 consider larger "causal hubs", each containing about 7-10 causal OTUs; Scenario 4 considers smaller causal hubs of 2-3 causal OTUs; Scenario 5 considers causal OTUs with random positions in the phylogenetic tree.

Figure 5 Upset plot of detected OTUs at FDR level of 0.05 and phylogenetic trees of the analyzed OTUs with detected OTUs for bacterial vaginosis study. SO: Single-OTU test implemented by POST with $c = 0$; TF: TreeFDR; DE: DESeq2; WR-P: Wilcoxon rank-sum test using proportional data; WR-R: Wilcoxon rank-sum test using CLR transformed data.

Figure 6 Upset plot of detected OTUs at FDR level of 0.05 and phylogenetic trees of the analyzed OTUs with detected OTUs for preterm birth study. SO: Single-OTU test implemented by POST with $c = 0$; TF: TreeFDR; DE: DESeq2; WR-P: Wilcoxon rank-sum test using proportional data; WR-R: Wilcoxon rank-sum test using CLR transformed data.

Table 1 Type I error rates at the significance levels of 0.05, 0.01, and 0.001 for POST.

Simulation	Outcome	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$	$\alpha=0.0001$
A	Binary	0.047	0.007	0.0006	0.00005
B	Continuous	0.051	0.010	0.0010	0.00008
	Binary	0.047	0.008	0.0006	0.00007

Table 2 AUC of different methods in Simulations A and B. Methods considered include POST, TreeFDR (TF), Single-OTU test (SO) implemented by POST with $c = 0$, DESeq2 (DE), Wilcoxon rank-sum test with proportional data (WR-P) or CLR transformed data (WR-R) for binary outcomes and Spearman correlation test with proportional data (SC-P) or CLR transformed data (SC-R) for continuous outcomes. The outcome values are simulated assuming no covariate effects and 5 different causal-OTU scenarios (shown in Addition file 1: Figure S1). Scenarios 1 to 3 consider larger "causal hubs", each containing about 7-10 causal OTUs; Scenario 4 considers smaller causal hubs of 2-3 causal OTUs; Scenario 5 considers causal OTUs with random positions in the phylogenetic tree. Methods with the highest AUC are shown in bold.

Simulation	Simulation A						Simulation B											
Outcome	Binary Outcome						Continuous Outcome						Binary Outcome					
Method	POST	TF	SO	DE	WR-P	WR-R	POST	TF	SO	DE	SC-P	SC-R	POST	TF	SO	DE	WR-P	WR-R
Scenario1	0.80	0.78	0.62	0.69	0.51	0.51	0.64	0.52	0.58	0.61	0.57	0.61	0.58	0.53	0.55	0.52	0.54	0.55
Scenario2	0.88	0.77	0.71	0.71	0.62	0.58	0.66	0.50	0.62	0.65	0.57	0.51	0.60	0.48	0.56	0.57	0.52	0.51
Scenario3	0.88	0.78	0.72	0.70	0.60	0.60	0.70	0.49	0.62	0.63	0.58	0.51	0.63	0.50	0.59	0.55	0.57	0.51
Scenario4	0.78	0.75	0.70	0.69	0.62	0.58	0.63	0.48	0.61	0.63	0.58	0.52	0.56	0.49	0.56	0.54	0.55	0.51
Scenario5	0.66	0.72	0.67	0.64	0.57	0.55	0.59	0.48	0.60	0.61	0.54	0.51	0.54	0.50	0.54	0.52	0.53	0.50
Scenario1	0.72	0.65	0.59	0.62	0.53	0.51	0.62	0.48	0.56	0.58	0.55	0.56	0.56	0.49	0.53	0.52	0.52	0.52
Scenario2	0.76	0.65	0.61	0.62	0.58	0.55	0.62	0.49	0.59	0.61	0.54	0.51	0.54	0.49	0.53	0.55	0.51	0.50
Scenario3	0.76	0.65	0.63	0.62	0.57	0.55	0.66	0.49	0.60	0.60	0.56	0.50	0.59	0.50	0.56	0.55	0.54	0.50
Scenario4	0.69	0.65	0.62	0.62	0.58	0.55	0.60	0.48	0.58	0.60	0.56	0.51	0.54	0.50	0.53	0.53	0.53	0.51
Scenario5	0.59	0.58	0.59	0.58	0.54	0.52	0.56	0.50	0.56	0.57	0.53	0.50	0.53	0.50	0.52	0.52	0.52	0.50

* The top of the results is for large OTU effect size and the bottom is for small OTU effect size. For Simulation A, small OTU effect size is from $N(\pm 1, 1)$ and large OTU effect size is from $N(\pm 2, 1)$. For Simulations B, small OTU effect size is simulated from $N(\pm 0.2, 0.04)$ and $N(\pm 0.3, 0.06)$ for continuous and binary outcomes, respectively; large OTU effect size is simulated from $N(\pm 0.5, 0.1)$ and $N(\pm 1, 0.2)$ for continuous and binary outcomes, respectively.

Table 3 OTUs significantly associated with bacterial vaginosis (BV) at FDR level of 0.05. TF: TreeFDR; SO: Single-OTU test implemented by POST with $c = 0$; DE: DESeq2; WR-P: Wilcoxon rank-sum test using proportional data; WR-R: Wilcoxon rank-sum test using CLR transformed data.

OTU	FDR-adjusted p-value (raw p-value)						Detected method	Genus/Species	Direction***
	POST	TF*	SO	DE	WR-P**	WR-R**			
OTU3	0.0386 (0.0004)	0.0895 (-)	0.0444 (0.0002)	0.0021 (0.0000)	0.1777 (0.0262)	0.0245 (0.0001)	POST/SO/DE/WR-R	<i>Lactobacillus crispatus</i>	-
OTU90	0.0386 (0.0006)	0.2448 (-)	0.9527 (0.8209)	0.9503 (0.2597)	0.3752 (0.2341)	0.9889 (0.6267)	POST	<i>Lactobacillus sp.</i>	+
OTU7	0.0386 (0.0008)	0.6535 (-)	0.8805 (0.0909)	0.1471 (0.006)	0.7604 (0.687)	0.9319 (0.0305)	POST	<i>Lactobacillus jensenii</i>	-
OTU82	0.0386 (0.0009)	0.5175 (-)	0.9898 (0.9191)	0.9701 (0.9968)	0.2584 (0.1158)	0.9889 (0.687)	POST	<i>Lactobacillus gasseri</i>	-
OTU66	0.0386 (0.0013)	0.5441 (-)	0.7006 (0.0138)	0.0267 (0.0007)	0.3782 (0.2397)	0.9319 (0.0117)	POST/DE	<i>Lactobacillus sp.</i>	-
OTU2	0.0386 (0.0015)	0.1804 (-)	0.917 (0.3501)	0.8758 (0.1746)	0.2348 (0.0436)	0.9889 (0.3088)	POST	<i>Lactobacillus iners</i>	+
OTU58	0.0386 (0.0015)	0.8392 (-)	0.917 (0.3455)	0.8758 (0.2085)	0.7912 (0.7311)	0.9889 (0.3088)	POST	<i>Lactobacillus sp.</i>	+
OTU112	0.4112 (0.0351)	0.0455 (-)	0.7006 (0.0233)	0.4217 (0.0419)	0.0402 (0.0006)	0.9319 (0.0192)	TF/WR-P	<i>Peptoniphilus sp.</i>	+
OTU85	0.4698 (0.0579)	0.0455 (-)	0.9065 (0.2613)	0.3725 (0.0291)	0.0968 (0.0081)	0.9889 (0.3363)	TF	<i>Gemella sp.</i>	+
OTU11	0.9184 (0.4987)	0.2863 (-)	0.9427 (0.4837)	0.0000 (0.0000)	0.2205 (0.0394)	0.9889 (0.7076)	DE	<i>Prevotella sp.</i>	+
OTU12	0.3905 (0.0284)	0.0895 (-)	0.7006 (0.0142)	0.0128 (0.0003)	0.0402 (0.0006)	0.9417 (0.0536)	DE/WR-P	<i>Prevotella sp.</i>	+
OTU16	0.6796 (0.1802)	0.1548 (-)	0.8805 (0.2094)	0.0006 (0.0000)	0.0968 (0.0059)	0.9889 (0.5687)	DE	<i>Prevotella timonensis</i>	+
OTU91	0.6796 (0.1543)	0.0916 (-)	0.8805 (0.106)	0.505 (0.0586)	0.0402 (0.0007)	0.9889 (0.835)	WR-P	<i>Prevotella sp.</i>	+

* : TF only reports FDR-adjusted p-values ** : WR-P and WR-R did not adjust for race. ***: + (and -) indicates that the OTU is positively (and negatively) associated with BV risk from a logistic regression.

effects besides OTU effects; Simulations B consider both binary outcomes (2nd rows) and continuous outcomes (3rd rows), which were simulated assuming OTU and covariate effects.

Additional file 2 — Figure S2.

QQ plot of POST p-values under the null hypothesis of no causal OTUs. The plots from left to right are for Simulations A, B-continuous and B-binary respectively. The outcomes in Simulation A were generated assuming no OTU effects and no covariate effects; the outcomes in Simulations B were generated assuming no OTU effects but with covariate effects.

Additional file 3 — Table S1.

AUC of different methods in Simulations B and C. The methods considered include POST, TreeFDR (TF), Single-OTU test (SO) and DESeq2 (DE). The outcome values are simulated assuming covariate effects and 5 different causal-OTU scenarios. Scenarios 1 to 3 consider larger "causal hubs", each containing about 7-10 causal OTUs; Scenario 4 considers smaller causal hubs of 2-3 causal OTUs; Scenario 5 considers causal OTUs with random positions in the phylogenetic tree. Wilcoxon rank-sum test (WR; for binary outcomes) and Spearman correlation test (SC; for continuous outcomes) are not included because they cannot account for covariates. Methods with the highest AUC are shown in bold.

Additional file 4 — Figure S3.

Boxplots of OTU2 (*Lactobacillus iners*) and OTU3 (*Lactobacillus crispatus*) for BV and healthy women in the BV study.

Additional file 5 — Figure S4.

OTU phylogenetic correlation for OTU2 and OTU3 with best c values 0.03 and 0, respectively. Only show neighboring OTUs inside genus *Lactobacillus*.

Additional file 6 — Table S2.

OTUs significantly associated with preterm birth at FDR level of 0.05. TF: TreeFDR; SO: Single-OTU test implemented by POST with $c=0$; DE: DESeq2; WR-P: Wilcoxon rank-sum test using proportional data; WR-R: Wilcoxon rank-sum test using CLR transformed data.

Figure 1

Figure 2

Figure 3

scenario 1

scenario 2

scenario 3

scenario 4

scenario 5

Figure 4

Figure 5

Figure 6

Supplementary Files

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

- [AdditionalFile1FigureS1.jpeg](#)
- [AdditionalFile2FigureS2.jpeg](#)
- [AdditionalFile3TableS1.docx](#)
- [AdditionalFile4FigureS3.png](#)
- [AdditionalFile5FigureS4.png](#)
- [AdditionalFile6TableS2.docx](#)