

METHODOLOGY

Open Access



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*}

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. Phylogenetic information is often incorporated 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 effects 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 show 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, in real data applications on bacterial vaginosis and on preterm birth, we find that POST can identify similar or more outcome-associated OTUs that are of biological relevance compared to existing methods.

Conclusions: Using POST, we show that adaptively leveraging the phylogenetic information can enhance the selection performance of associated microbiome features by improving the overall true-positive and false-positive detection. We developed a user friendly R package *POSTm* which is freely available on CRAN (<https://CRAN.R-project.org/package=POSTm>).

Keywords: Association test, Phylogenetic tree, Kernel machine regression

Background

The human microbiome is the collection of microorganisms residing in the human body and has been definitively shown to impact human disease and health [1]. Multiple resources are now available that enable the characterization of microbial communities in the human body and further our understanding of how the human

microbiota affect clinical outcomes of the host, e.g., the Human Microbiome Project [2] and Integrative Human Microbiome Project [3]. 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 identifiers. The generated sequencing reads can be either clustered into operational taxonomic units (OTUs) based on sequence similarity [4] or denoised into amplicon sequence variants (ASVs) with exact sequences [5]. For simplicity, hereafter,

*Correspondence: jtzeng@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



© The Author(s). 2022 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

we use “OTU” as a general reference to a taxonomic unit when discussing the overall concepts. We will explicitly distinguish OTU and ASV when the two terms need to be discerned, such as in the real data analysis.

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 the community level [7–11]. 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 in 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 challenging due to sparse OTU counts, low OTU abundance, and high OTU dimensionality. Multiple studies have shown that incorporating phylogenetic structure among OTUs can increase the detecting power or predictive accuracy of a 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. [14] introduced a false discovery rate (FDR) control procedure, TreeFDR, that 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) as well as those that incorporate grouping structure [15, 16]. Xiao et al. [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 showed that glmmTree improves prediction performance, especially when associated OTUs are clustered on the tree. Kim et al. [18] introduced a genus-level collapsing association test, TMat, that first assesses the 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, 11, 21] 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 $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$, 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 [10, 22] 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 an 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 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.

Step 1: 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* [23]. 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, $\ell, m \in \{1, \dots, M\}$; and 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 Fig. 2 a typical relationship between c and $r_{\ell m}$ using the dataset of Charlson et al. [24] considered in the simulation study. Without loss of generality, we focus on OTU3438 and its 16 neighboring OTUs in the phylogenetic tree (Fig. 2A) and show the values of $r_{\ell,3438}$ for the neighboring OTU ℓ under different c 's in Fig. 2B. 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., Fig. 2B, 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., Fig. 2B, panel of $c = 100$).

When conducting POST association test in step 3, we use data to determine the optimal c value. Specifically, we compute the p -values for a grid of c between 0 and c_{\max} and use the Cauchy combination (CC) method [25, 26] to aggregate the p -values of different c 's. As detailed in the simulation studies, we set the $c_{\max} = 0.05$ to ensure information sharing is limited within a small, concentrated

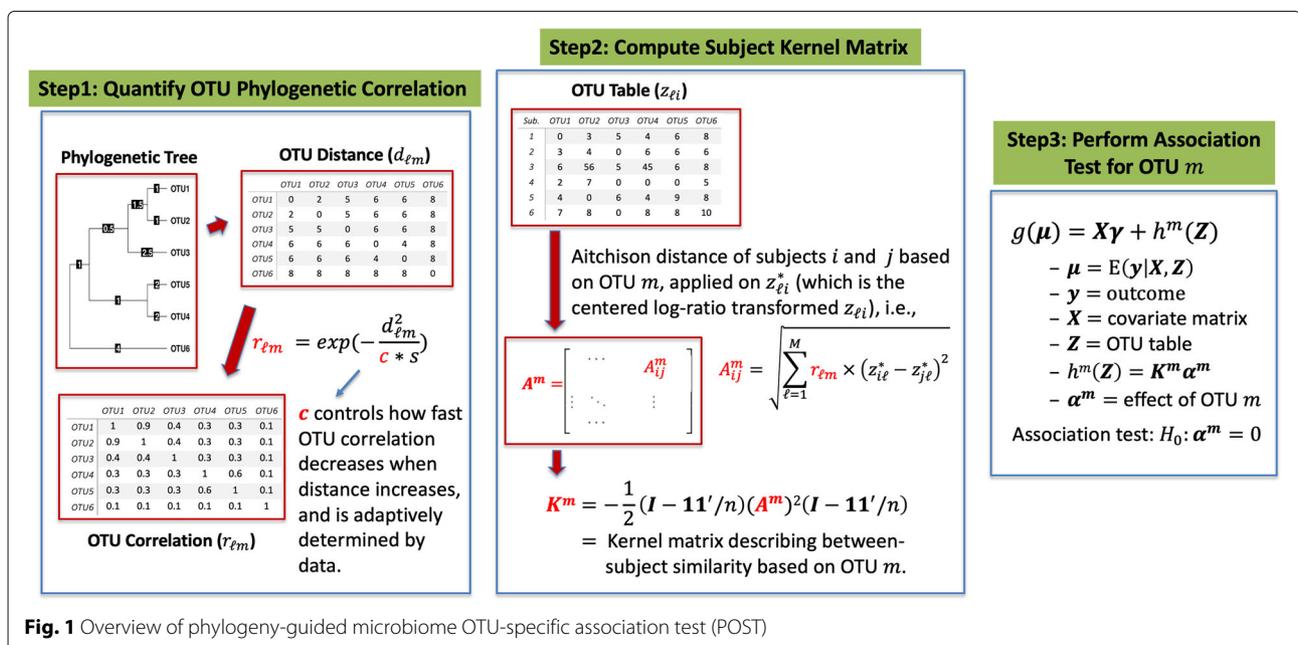
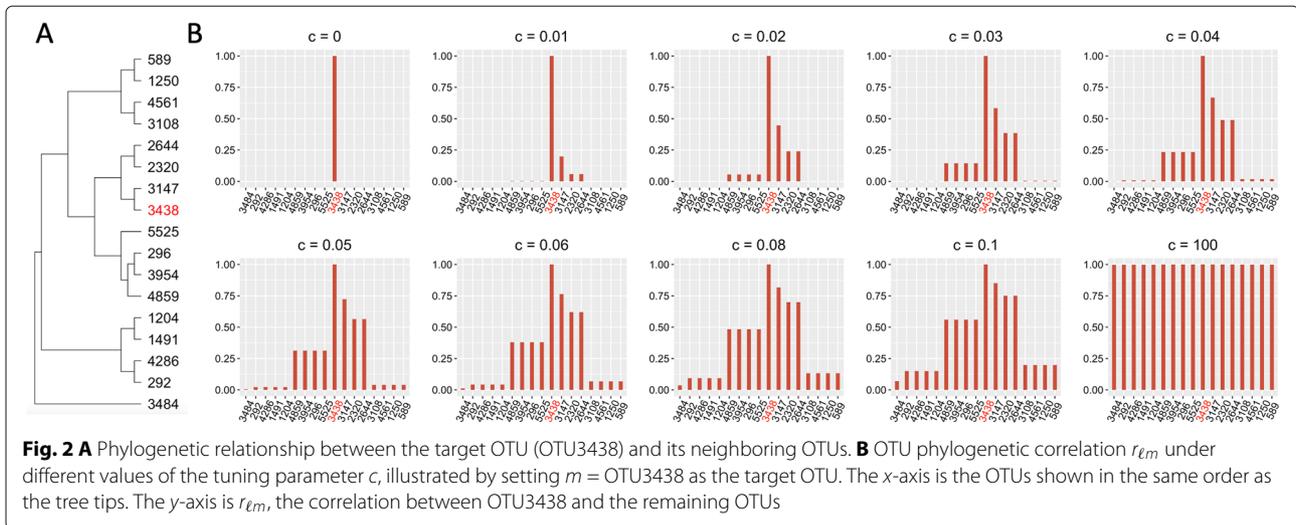


Fig. 1 Overview of phylogeny-guided microbiome OTU-specific association test (POST)



neighborhood of the target OTU. The adaptively determined c also ensures that only informative neighbors will make contributions to the target OTU.

Step 2: 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, 22]. 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 A , and then transform A to the kernel similarity matrix K using Gower’s matrix, i.e., $K = -\frac{1}{2} \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n} \right) A^2 \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n} \right)$, where A^2 is the element-wise square of A , I is the $n \times n$ identity matrix, and $\mathbf{1}$ is a vector of ones with length n [22]. For POST, we propose to construct a local, OTU-specific kernel K^m so that only local information around the target OTU is used. First, obtain the local dissimilarity matrix A^m by incorporating $r_{\ell m}$ into the original dissimilarity matrix A . Next, perform Gower’s transformation on A^m to obtain $K^m \equiv \{k^m(\mathbf{z}_i, \mathbf{z}_j)\} = -\frac{1}{2} \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n} \right) (A^m)^2 \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n} \right)$. Although such local kernel construction can be applied on arbitrary distance metrics, in POST, we choose to use the Aitchison distance, a non-phylogenetic distance measure accounting for the compositional nature of abundance data, to avoid the potential overuse of phylogenetic information. Aitchison distance is the Euclidean distance of centered log-ratio (CLR) transformed abundance data, and the “local” Aitchison distance for OTU m between subjects i and j can be obtained by:

$$A_{ij}^m = \sqrt{\sum_{\ell=1}^M r_{\ell m} \times (z_{i\ell}^* - z_{j\ell}^*)^2}, \tag{2}$$

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 the OTU counts table before doing CLR transformation. If proportion data is used, a pseudo-proportion $1e-6$ can be added to the proportion table [27].

Step 3: 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 a test can be constructed through the connection of kernel machine regression and generalized linear mixed models (GLMM) [28, 29]. That is, $h^m(\mathbf{Z})$ can be viewed as subject-specific random effects with $h^m(\mathbf{Z}) \sim N(\mathbf{0}, \tau K^m)$, 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 $T^{m,c} = \frac{1}{2\hat{\phi}} (\mathbf{y} - \hat{\boldsymbol{\mu}}_0)^T K^m (\mathbf{y} - \hat{\boldsymbol{\mu}}_0)$, where $\hat{\boldsymbol{\mu}}_0 = g^{-1}(X\hat{\boldsymbol{\gamma}})$ with $\hat{\boldsymbol{\gamma}}$ the estimated covariate coefficient under the null model: $g(\boldsymbol{\mu}) = X\boldsymbol{\gamma}$, and $\hat{\phi}$ is the estimator of the dispersion parameter under H_0 . For continuous outcome, $\hat{\boldsymbol{\mu}}_0 = X\hat{\boldsymbol{\gamma}}$ and $\hat{\phi}$ equals to the estimated residual variance under H_0 . For binary outcome, $\phi = 1$ and $\hat{\boldsymbol{\mu}}_0 = \exp\{X\hat{\boldsymbol{\gamma}}\} / (1 + \exp\{X\hat{\boldsymbol{\gamma}}\})$. With a fixed c , $T^{m,c}$ asymptotically follows a weighted mixtures of $\chi_{(1)}^2$ 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. [30] 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 CC method [25, 26] to combine the transformed 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 [25, 26].

Simulation study

We conducted simulation studies to evaluate the performance of POST in identifying outcome-associated OTUs. We generated OTU data based on the real dataset from the upper respiratory tract microbiome study [24], 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 considered 2 different simulation settings (simulations A and B); for each simulation, we simulated OTU counts for $n = 100$ individuals following Chen and Li [22] and focused on the $M = 400$ most abundant OTUs in the original data. We modeled the observed OTU counts using the Dirichlet-multinomial distribution with parameters $(\pi_1, \dots, \pi_M, \theta)$, where π_ℓ 's are the means of the OTU proportions and θ is the over-dispersion parameter [31, 32], and obtained their maximum likelihood estimates $(\hat{\pi}_1, \dots, \hat{\pi}_M, \hat{\theta})$. Next, we generated the OTU abundance proportions (p_1, \dots, p_M) from Dirichlet $(\hat{\pi}_1, \dots, \hat{\pi}_M, \hat{\theta})$; we also generated the total counts of individual i , N_i , from a negative binomial distribution with mean 10,000 and size 25. Finally, we generated the OTU counts of individual i (z_{i1}, \dots, z_{iM}) from multinomial (N_i, p_1, \dots, p_M) ; the OTU count matrix \mathbf{Z} comprises the resulting simulated counts.

Simulation A

In simulation A, we adopted the simulation design of Xiao et al. and considered a case-control study with 50 cases (i.e., $y_i = 1$) and 50 controls (i.e., $y_i = 0$). Specifically, we used the R package *pam* to partition the 400 OTUs into 20 clusters based on the phylogenetic tree and selected causal OTUs from the 8 most abundant clusters. Given \mathbf{Z} with $M = 400$ and $n = 100$, we then multiplied 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 generated from normal($v, 1$) if OTU m is causal and 0 otherwise. We considered $|v|$ to be 1 (small effect size) or 2 (large effect size).

We consider five scenarios for causal OTUs as illustrated in Fig. 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 $N(v, 1)$ for all causal OTUs. In scenario 2, $\beta_m \sim N(v, 1)$ for half of

the causal hubs and $\beta_m \sim N(-v, 1)$ 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 $N(v, 1)$, and the remaining half have their β_m 's from $N(-v, 1)$; 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 $\beta_m \sim N(v, 1)$ and the other half causal hubs with $\beta_m \sim N(-v, 1)$. 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 $\beta_m \sim N(v, 1)$ and the other half OTUs have $\beta_m \sim N(-v, 1)$.

Simulation B

In simulation B, we considered both continuous and binary outcomes and adopted the simulation design of Zhao et al. [10] and Koh et al. [11]. Given the simulated OTU counts $\mathbf{Z} = \{z_{im}\}$ and the same set of causal OTUs as simulation A, we generated 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 causal OTU m for subject i ; and $\omega = 1$ or 0 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 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, $\beta_m \sim \text{normal}(v, v/5)$ or $\text{normal}(-v, v/5)$ dependent on the scenario. We considered $|v|$ to be 0.2 (small effect size) and 0.5 (large effect size) for continuous outcomes and to be 0.3 (small effect size) and 1 (large effect size) for binary outcomes. In each setting, we considered $\omega = 0$ (no covariate effects) and $\omega = 1$ (with covariate effects).

Determining appropriate c_{\max}

To determine the appropriate c_{\max} for POST, we conducted 100 replications under each of the 5 causal scenarios in simulations A and B and used the decision rule that an OTU is selected as important if p -value < 0.05 and as 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 the pseudo- F score. The TPR was obtained by first computing the fraction of detected causal OTUs among all causal OTUs in each replication and then averaging across the 100 replications; the FPR was obtained by first computing the fraction of detected non-causal OTUs among all non-

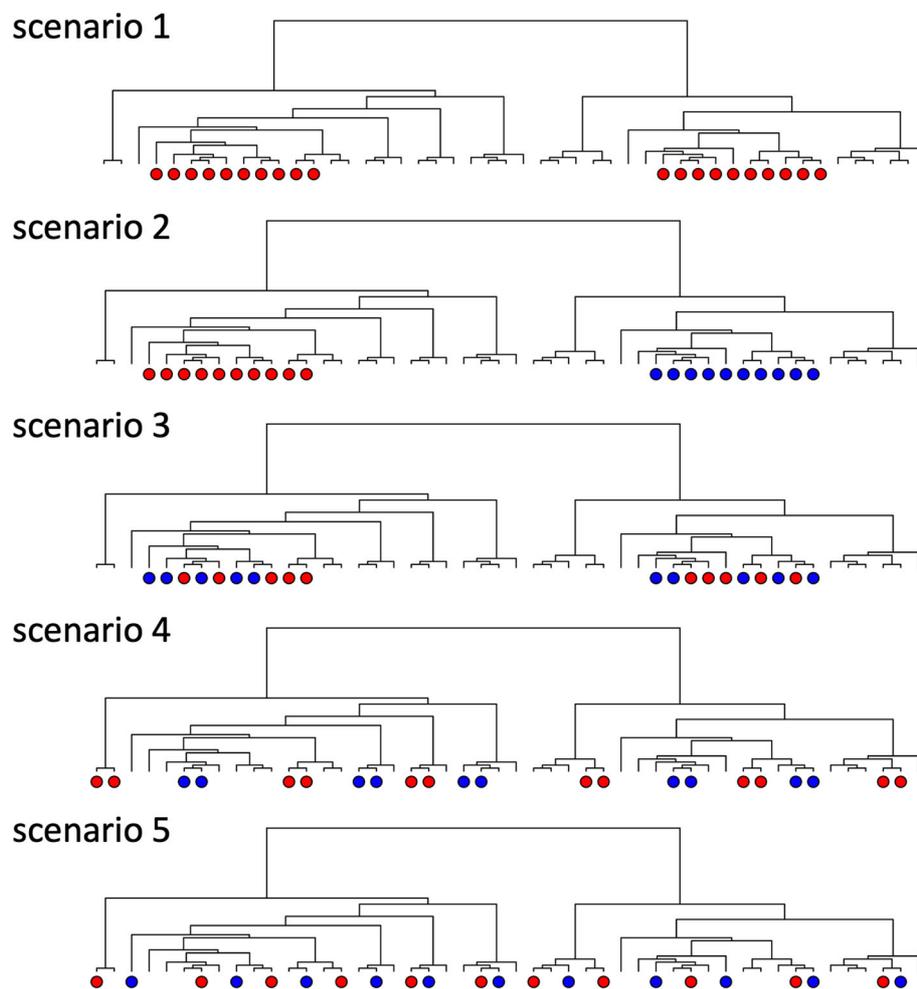


Fig. 3 Illustration of the five causal OTU scenarios considered in the simulation. Scenarios 1 to 3 consider larger “causal hubs,” each containing about 7–10 causal OTUs; scenario 4 considers smaller causal hubs of 2–3 causal OTUs; scenario 5 considers causal OTUs with random positions in the phylogenetic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

causal OTUs in each replication and then averaging across the 100 replications; and the pseudo- F was obtained by taking the harmonic mean of the TPR and $1 - \text{FPR}$ 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 4000 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 was obtained as follows. (1) In each replication, we computed the FDR-adjusted p -values of a method, based on which

we then computed the FPR and TPR at varying significance thresholds. (2) We took the average FPR and average TPR across the 100 replications at each of the significant thresholds. (3) We constructed the ROC curve using the average FPRs and TPRs and computed the corresponding AUC. We used 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-adjusted p -values. TSBH has been shown to yield good performance for correlated p -values [34] and is implemented in R function “mt.rawp2adjp” of R package *multtest*.

We compare POST with 7 baseline approaches: (1) TreeFDR (TF) of Xiao et al. [14] as implemented by the function “MicrobiomeSeqTreeFDR” in R package *StructFDR*; (2) single-OTU test (SO), by setting $c = 0$ in POST; (3) DESeq2 (DE) of Love et al. [35] as implemented by the function “DESeq” with the default Wald test in the R package *DeSeq2*; (4) ANCOM-BC (AB) of

Lin et al. [36] as implemented by the function “*ancombc*” in the R package *ANCOMBC*; (5) LinDA (LD) of Zhou et al. [37] as implemented in the function “*linda*” in R package *LinDA*; and (6) Wilcoxon rank-sum test on the proportional data (WR-P) and on the CLR transformed data (WR-R) as implemented in the function “*wilcox.test*.” The Wilcoxon test is only appropriate for binary outcomes with no covariate effects. We also consider (7) Spearman correlation test on proportional data (SC-P) and on the CLR transformed data (SC-R) as implemented by the function “*cor.test*” of the *stats* R package. The Spearman test is only appropriate for continuous outcomes with no covariate effects. Among these methods, POST and TF incorporate phylogenetic tree information. POST, SO, LD, 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 examined the trajectory of pseudo- F (Additional file 1: Fig. 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 Fig. 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 found that $c_{\max} = 0.05$ appears to offer a robust choice across the different scenarios in simulations A and B and implemented POST with $c_{\max} = 0.05$ hereafter.

Validity of POST

We checked the behavior of POST under the global null hypothesis of no causal OTUs. Table 1 shows that the type I error rate is appropriately controlled at nominal levels of 0.05, 0.01, and 0.001. Additional file 2: Fig. 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.

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

Selection performance

Next, we examined the relative performance of POST against the baseline methods using AUC (Table 2 for simulations 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) (Table 2; Fig. 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 performed comparably to the best-performing methods. Among the non-tree based methods, SO, DE, AB, and LD had similar AUCs 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-performing method in all scenarios, regardless of the effect sizes, without covariates (Table 2) or with covariates (Additional File 3: Table S1). DE, AB, and LD had similar AUCs and can perform the best even in some of scenarios 1–4 although they did not incorporate tree information. We also observe that TF has an 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 fewer tied adjusted p -values with relatively smaller values than those of TF.

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, SO, DE, AB or LD, depending on simulation settings.

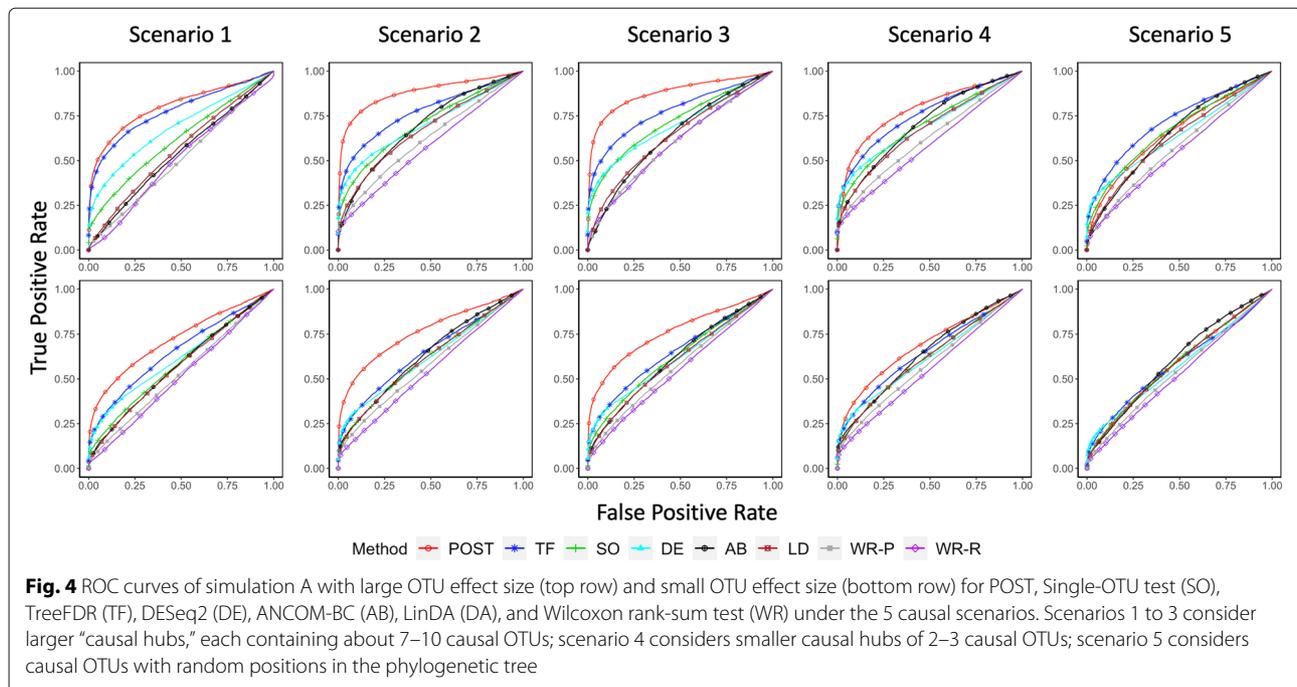
Real data analysis

We applied POST to two microbiome datasets, both at the level of OTUs formed at $\leq 3\%$ dissimilarity and at the level of ASVs. We present the OTU-level results below and the ASV-level results in Additional file 7: Section S1.

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), ANCOM-BC (AB) and LinDA (LD), 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. 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 Outcome	Simulation A										Simulation B													
	Binary outcome					Continuous outcome					Binary outcome					Continuous outcome								
	POST	TF	SO	DE	AB	LD	WR-P	WR-R	POST	TF	SO	DE	AB	LD	SC-P	SC-R	POST	TF	SO	DE	AB	LD	WR-P	WR-R
Large effect size*																								
Scenario1	0.80	0.78	0.62	0.69	0.54	0.56	0.51	0.51	0.64	0.52	0.58	0.61	0.57	0.61	0.57	0.61	0.58	0.53	0.55	0.52	0.57	0.59	0.54	0.55
Scenario2	0.88	0.77	0.71	0.71	0.70	0.67	0.62	0.58	0.66	0.50	0.62	0.65	0.62	0.62	0.57	0.51	0.60	0.48	0.56	0.57	0.61	0.56	0.52	0.51
Scenario3	0.88	0.78	0.72	0.70	0.64	0.64	0.60	0.60	0.70	0.49	0.62	0.63	0.61	0.62	0.58	0.51	0.63	0.50	0.59	0.55	0.58	0.58	0.57	0.51
Scenario4	0.78	0.75	0.70	0.69	0.70	0.67	0.62	0.58	0.63	0.48	0.61	0.63	0.58	0.61	0.58	0.52	0.56	0.49	0.56	0.54	0.55	0.55	0.55	0.51
Scenario5	0.66	0.72	0.67	0.64	0.65	0.63	0.57	0.55	0.59	0.48	0.60	0.61	0.59	0.60	0.54	0.51	0.54	0.50	0.54	0.52	0.57	0.55	0.53	0.50
Small effect size*																								
Scenario1	0.72	0.65	0.59	0.62	0.57	0.57	0.53	0.51	0.62	0.48	0.56	0.58	0.56	0.59	0.55	0.56	0.56	0.49	0.53	0.52	0.54	0.57	0.52	0.52
Scenario2	0.76	0.65	0.61	0.62	0.63	0.62	0.58	0.55	0.62	0.49	0.59	0.61	0.59	0.59	0.54	0.51	0.54	0.49	0.53	0.55	0.56	0.53	0.51	0.50
Scenario3	0.76	0.65	0.63	0.62	0.62	0.60	0.57	0.55	0.66	0.49	0.60	0.60	0.58	0.60	0.56	0.50	0.59	0.50	0.56	0.55	0.55	0.56	0.54	0.50
Scenario4	0.69	0.65	0.62	0.62	0.64	0.62	0.58	0.55	0.60	0.48	0.58	0.60	0.57	0.58	0.56	0.51	0.54	0.50	0.53	0.53	0.52	0.54	0.53	0.51
Scenario5	0.59	0.58	0.59	0.58	0.60	0.58	0.54	0.52	0.56	0.50	0.56	0.57	0.56	0.57	0.53	0.50	0.53	0.50	0.52	0.52	0.53	0.52	0.52	0.50

*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



In all analyses, we considered TF, SO, AB, LD, WR-P, and WR-R as baselines. We included race as a covariate except for WR-based methods. We used the TSBH procedure [33] to compute the FDR-adjusted p -values except for TF, which outputted its own FDR-adjusted p -values. We selected important OTUs/ASVs using the decision rule of FDR-adjusted p -values < 0.05.

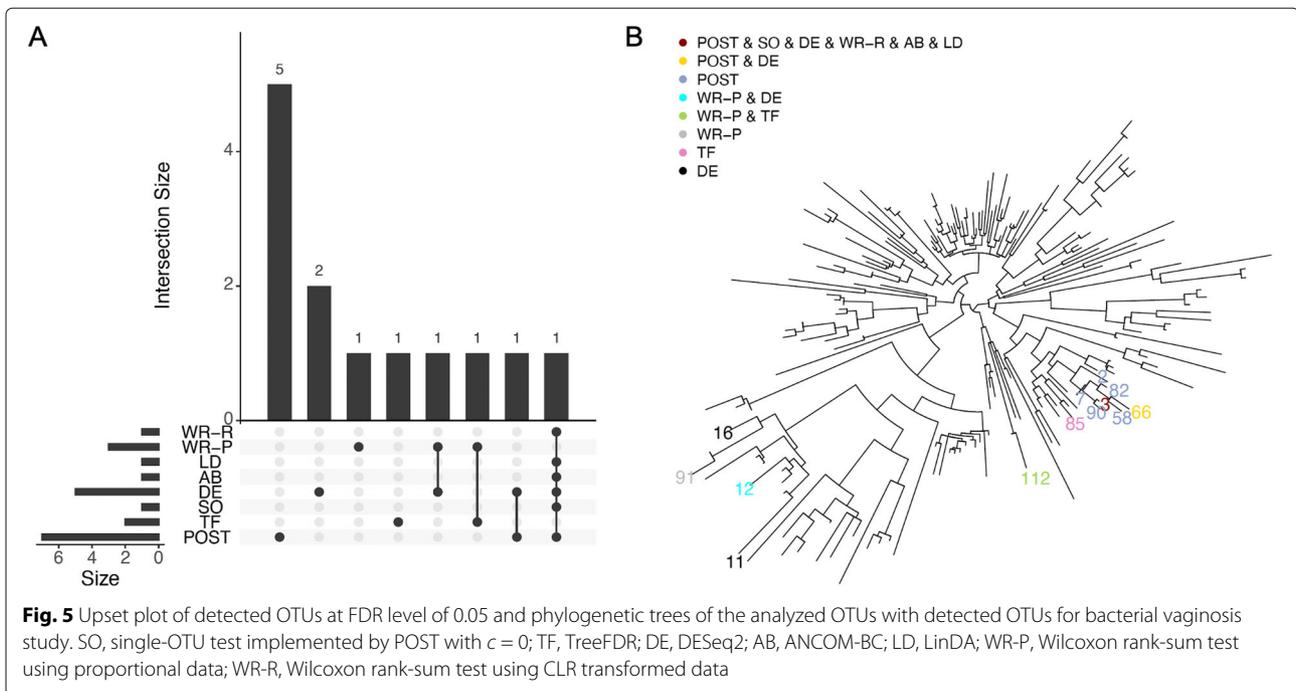
Association study of vaginal microbiome and bacterial vaginosis

Bacterial vaginosis (BV) is a type of vaginal inflammation and is characterized by low levels of lactobacilli and overgrowth of various anaerobic bacteria [38]. BV is most often diagnosed using the Amsel criteria or microscopy-based Nugent scoring [39] but with limited accuracy [40]. High-throughput sequencing technologies such as 16S rRNA amplicon sequencing have been used to study the species diversity of vaginal microbiota and have shown promise as an alternative assessment of BV for practical and accurate diagnosis [41–43].

We conducted an OTU-specific analysis using the 16S rRNA gene sequencing dataset of vaginal microbiome from [43] to evaluate the association between BV and the normal 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 employed *FastTree2* [23] to infer the phylogenetic tree using the OTU sequences. Finally, we filtered out OTUs with abundance < 0.005%

and prevalence < 10% and analyzed the resulting 186 OTUs of 39 individuals for BV association studies.

The Upset plot (Fig. 5A) shows that POST, TF, SO, DE, AB, LD, WR-P, and WR-R identified 7, 2, 1, 5, 1, 1, 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. We organized our findings as follows. (i) Among the 7 POST-detected OTUs, OTU3 (mapped to *Lactobacillus crispatus*) is also identified by SO, DE, AB, LD, and WR-R. Depletion of *Lactobacillus crispatus* has been shown to be highly associated with BV [44]. 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 jensenii*, *Lactobacillus gasseri*, and *Lactobacillus iners*. These vaginal *Lactobacillus* species have been shown to have different abundances in women with vs. without BV [44] and play important roles in the vaginal ecosystem [45]. (ii) TF identified 2 OTUs, i.e., OTU112 (mapped to *Peptoniphilus* sp. and also identified by WR-P) and OTU85 (mapped to *Gemella* sp.); these genera have also been detected in BV cases [46, 47]. (iii) Besides the OTUs discussed above, DE and/or 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 to be associated with BV [48]. (iv) It seems that POST, SO, LD, 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. (v) Fig. 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 illustrated the data-adaptiveness 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, AB, LD, and WR-R). The abundance box-plots (Additional file 4: Fig. S3) show that OTU2 has higher abundance in BV patients compared to healthy women, while OTU3 has an opposite pattern. Our find-

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; AB, ANCOM-BC; LD, LinDA; WR-P, Wilcoxon rank-sum test using the CLR transformed data

OTU	FDR-adjusted p -value								Detected method	Genus/species	Direction**
	POST	TF	SO	DE	AB	LD	WR-P*	WR-R*			
OTU3	0.039	0.089	0.045	0.002	0.006	0.041	0.178	0.025	POST/SO/DE/AB/LD/WR-R	<i>Lactobacillus crispatus</i>	-
OTU90	0.039	0.245	0.953	0.950	0.983	0.961	0.375	0.989	POST	<i>Lactobacillus</i> sp.	+
OTU7	0.039	0.653	0.881	0.147	0.766	0.729	0.760	0.932	POST	<i>Lactobacillus jensenii</i>	-
OTU82	0.039	0.517	0.985	0.970	0.963	0.976	0.258	0.989	POST	<i>Lactobacillus gasseri</i>	-
OTU66	0.039	0.544	0.701	0.027	0.379	0.615	0.378	0.932	POST/DE	<i>Lactobacillus</i> sp.	-
OTU2	0.039	0.180	0.917	0.876	0.766	0.923	0.235	0.989	POST	<i>Lactobacillus iners</i>	+
OTU58	0.039	0.839	0.917	0.876	0.871	0.923	0.791	0.989	POST	<i>Lactobacillus</i> sp.	+
OTU112	0.411	0.046	0.701	0.422	0.379	0.678	0.040	0.932	TF/WR-P	<i>Peptoniphilus</i> sp.	+
OTU85	0.470	0.046	0.906	0.372	0.766	0.870	0.097	0.989	TF	<i>Gemella</i> sp.	+
OTU11	0.918	0.286	0.943	0.000	0.869	0.923	0.220	0.989	DE	<i>Prevotella</i> sp.	+
OTU12	0.391	0.089	0.701	0.013	0.338	0.615	0.040	0.942	DE/WR-P	<i>Prevotella</i> sp.	+
OTU16	0.680	0.155	0.881	0.001	0.766	0.918	0.097	0.989	DE	<i>Prevotella timonensis</i>	+
OTU91	0.680	0.092	0.881	0.505	0.766	0.852	0.040	0.989	WR-P	<i>Prevotella</i> sp.	+

*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

ings agree with the literature that BV patients have loss of many *Lactobacillus* species except *Lactobacillus iners* [44, 49]. When testing for OTU2, the best c determined by the data is $c = 0.03$, suggesting some information borrowing from neighboring OTUs (Additional file 5: Fig. S4A), e.g., OTU7 has $r_{7,2} = 0.74$ and OTU3 has $r_{3,2} = 0.52$ in Eq. (2). This example illustrates that (i) POST can incorporate information with opposite directions, and (ii) the information borrowing is OTU-specific and data-driven—when testing for OTU3, the best c is $c = 0$, i.e., no information borrowing from any neighboring OTUs (Additional file 5: Fig. S4B).

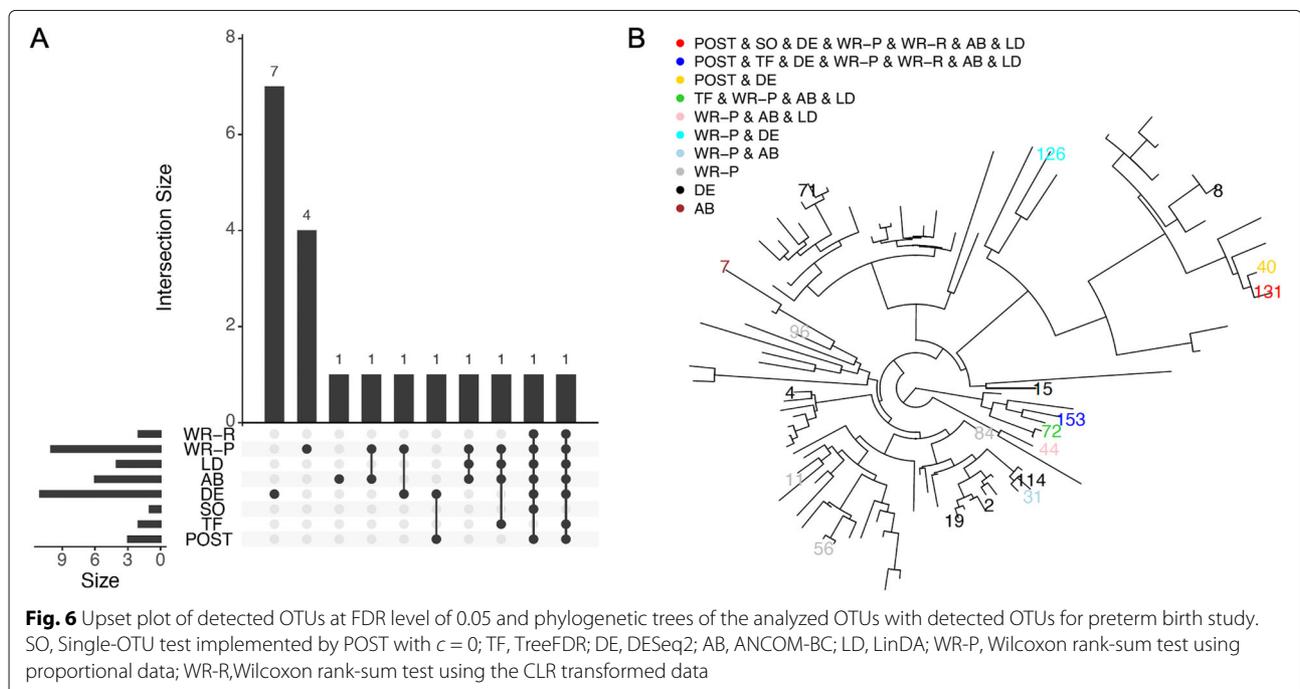
Association study of vaginal microbiome and preterm birth

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 [12, 20, 50]. We performed an OTU-specific analysis on the Stanford cohort data of Callahan et al. [50] to identify OTUs associated with preterm birth. We applied the same data processing steps as in the BV study and obtained 746 OTUs and 39 individuals. After excluding OTUs with abundance < 0.005% and present rate < 10%, we based our analysis on the 95 remaining OTUs.

Figure 6A shows that POST, TF, SO, DE, AB, LD, WR-P, and WR-R identified 3, 2, 1, 11, 6, 4, 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 results can be summarized as follows. (i) All of the 3 POST-detected OTUs have also been detected by at least 1 baseline method, i.e., OTU131 (mapped to *Prevotella* sp.), which is also found significant by all baseline methods except TF; OTU40 (mapped to *Prevotella melaninogenica*), which is also identified by DE; and OTU153 (mapped to *Neisseria* sp.), which is also identified by all baseline methods except SO. The *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 infection and has been reported to be associated with preterm birth [51, 52]. (ii) There were 4 OTUs that were identified by ≥ 1 baseline method but missed by POST: OTU72, OTU126, OTU31, and OTU44. OTU72 (mapped to *Haemophilus parainfluenzae*) was identified by TF, AB, LD, and WR-P, and *Haemophilus parainfluenzae* has been reported to play a potential role in preterm birth [53]. OTU44 (mapped to *Fusobacterium nucleatum* and identified by AB, LD, and WR-P), has been reported associated with preterm birth [54]. We found no supporting evidence in the literature regarding the association with preterm birth for OTU126 (mapped to *Cloacibacterium* sp. and identified by DE and WR-P) or for OTU31 (mapped to *Staphylococcus anginosus* and identified by AB and WR-P).

(iii) Besides the OTUs discussed above, DE uniquely identified 7 OTUs. Among these OTUs, OTU2 (mapped to *Lactobacillus crispatus*) and OTU8 (mapped to *Pre-*



votella sp.) have been found associated with preterm birth in the literature [12, 20]. OTU4 (mapped to Veillonellaceae bacterium), OTU71 (mapped to *Alloscardovia omnicoles*), and OTU19 (mapped to *Staphylococcus aureus*) have also been discussed in the literature, although they are reported as not associated with preterm birth [55, 56]. We found no supporting evidence in the literature regarding the association with preterm birth for OTU114 (mapped to *Lacticaseibacillus rhamnosus*) or for OTU15 (mapped to *Mycoplasma hominis*). (iv) AB uniquely identified OTU7 (mapped to *Clostridiales genomsp. BVAB1*), which has been reported to be associated with preterm birth [12]. (v) WR-P uniquely identified 4 OTUs, for which we did not find supporting evidence in the literature, and the associations diminish if switching to WR-R. (vi) We also observed that the CLR-based methods (e.g., POST, SO, LD, and WR-R) tend to identify fewer OTUs compared to the proportional-based methods (e.g., DE and WR-P).

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 detected several significant unique OTUs that were not identified by other methods. We discussed the biological relevance of these identified OTUs and examined the phylogenetic relationship among the identified OTUs (Fig. 5B). The results suggest that these OTUs tend to be evolutionarily close and their signals might be easier to be detected by borrowing an appropriate amount information from appropriate neighboring OTUs. In contrast, in the preterm birth study, POST only identified OTUs that are also identified by other non-tree-based methods. In particular, POST had similar performance to SO, LD, 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 from which to borrow information. We evaluated the performance of POST in different simulated scenarios and in real data applications. We showed that POST can enhance the selection performance by yielding more desirable true

positive and false-positive detection when compared to commonly used tree-free and tree-guided 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 described POST based on OTUs formed at $\leq 3\%$ sequence dissimilarity and the corresponding phylogenetic tree constructed using *FastTree2* [23]. 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 of 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 thus, 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.

It is possible to extend the POST framework to incorporate OTU correlation based on abundance/co-occurrence instead of phylogeny. Method 1 is to replace the OTU phylogenetic correlation ($r_{\ell m}$) in step 1 with the OTU abundance correlations (such as computing $r_{\ell m}$ using SparCC [57]). A hyperparameter c will need to be integrated into $r_{\ell m}$ to permit that the information borrowing is conducted in an outcome-supervised fashion. Method 2 is to replace the input phylogenetic tree in step 1 with a tree constructed based on abundance correlation or some network structures [58], e.g., by first converting the abundance correlation to a dissimilarity matrix and then performing hierarchical clustering to form a "tree" as outlined in Bichat et al. [19]. Method 2 can

be directly implemented by the existing POST R package *POSTm*, although additional numerical studies would be needed to determine an appropriate c_{\max} value. However, method 2 appears to incorporate some redundancy because in the existing POST workflow, the “tree” will then be converted into correlation (i.e., r_{lm}). Finally, given that which correlation type (abundance vs. phylogeny) would be informative of the underlying OTU association patterns is unknown in real practice, an important next step will be to construct an omnibus framework that can adaptively determine “which types” of correlation information to use besides “how much” information to borrow from correlated OTUs.

In the real data analyses, we added a pseudo-count of 0.5 to the OTU count tables to handle the zero-count issues. It has been shown in the literature that the choice of pseudo-count can affect the association results [59, 60]. For POST, we conducted sensitivity analyses to assess the impact of pseudo-counts, by comparing the results of pseudo-count 0.5 and pseudo-count 1 under simulation A (Additional file 8: Table S3) and in real data analyses (Additional file 9: Fig. S5). We observed that (1) the AUCs based on different pseudo-counts (Additional file 8: Table S3) are very similar across different scenarios and effect sizes. (2) The p -values obtained using different pseudo-counts (Additional file 9: Fig. S5) are highly correlated and fall along the 45-degree lines in the scatter plots. While some deviations may occur (e.g., in the BV analysis), the deviations were from non-small p -values. These results suggest that POST is fairly robust to the choices of pseudo-counts.

In POST, we build the kernel matrix based on the non-phylogenetic Aitchison distance to account for the compositional nature of microbiome data [61]. 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 $WU_{ij}^m = \frac{\sum_{t=1}^T b_t |r_m|_t^T (q_i^t - q_j^t)|}{\sum_{t=1}^T b_t |r_m|_t^T (q_i^t + q_j^t)}$ 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 the “Step 1: Quantifying OTU phylogenetic correlation” section 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 $\mathbf{1}$, the 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.

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; AB: ANCOM-BC; LD: LinDA; 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

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-022-01266-3>.

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 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 Simulation B. The methods considered include POST, TreeFDR (TF), Single-OTU test (SO), DESeq2 (DE), ANCOM-BC (AB) and LinDA (LD). 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 neighboring OTUs inside genus *Lactobacillus* are shown.

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.

Additional file 7: Section S1. In this section, we conducted association analysis at the ASV level (i.e., 0% dissimilarity) using the same analysis strategies as described in the main texts for the bacterial vaginosis (BV) and the preterm birth (PTB) association studies.

Additional file 8: Table S3. The AUC using different pseudo-counts in Simulation A. The AUCs based on pseudo-count 0.5 are very close to the AUCs based on pseudo-count 1 across different scenarios and effect sizes.

Additional file 9: Figure S5. Scatter plots of $-\log_{10} p$ -values obtained using pseudo-count 0.5 (X-axis) and those obtained using pseudo-count 1 (Y-axis) in the BV study (left) and the PTB study (right). The p -values of different pseudo-counts are highly correlated ($R=0.995$ in BV and 0.998 in PTB) and fall along the 45-degree lines.

Acknowledgements

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

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 the 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 drafted 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. All authors read and approved the final manuscript.

Funding

This work has been partially supported by the 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).

Availability of data and materials

The dataset of upper respiratory tract microbiome study [24] used in our simulation study is available in R package *GUniFrac*. The 16S rRNA sequencing dataset and metadata of the vaginal microbiome and bacterial vaginosis study [43] 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 [50] used in the real data analysis are publicly available at the Stanford Digital Repository (<https://purl.stanford.edu/yb681vm1809>). Our POST method is implemented by R package *POSTm*, and it is now available at CRAN (<https://CRAN.R-project.org/package=POSTm>). The R codes used for simulation, real data analysis, and data visualization are available through GitHub by request.

Declarations

Ethics approval and consent to participate

No ethics approval or consent to participate was required.

Consent for publication

No consent for publication was required.

Competing interests

The authors declare that they have no competing interests.

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.

Received: 25 October 2021 Accepted: 1 April 2022

Published online: 07 June 2022

References

- Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet*. 2012;13(4):260–70.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature*. 2007;449(7164):804–10.
- et al. The integrative human microbiome project. *Nature*. 2019;569:641–8.
- Schloss PD, Handelsman J. Introducing dotur, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol*. 2005;71(3):1501–6.
- Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J*. 2017;11(12):2639.
- Alekseyenko AV, Perez-Perez GI, De Souza A, Strober B, Gao Z, Bihan M, Li K, Methé BA, Blaser MJ. Community differentiation of the cutaneous microbiota in psoriasis. *Microbiome*. 2013;1(1):31.
- Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, Collman RG, Bushman FD, Li H. Associating microbiome composition with environmental covariates using generalized unifracs distances. *Bioinformatics*. 2012;28(16):2106–13.
- Fukuyama J, McMurdie PJ, Dethlefsen L, Relman DA, Holmes S. Comparisons of distance methods for combining covariates and abundances in microbiome studies. In: *Biocomputing 2012*. Singapore: World Scientific; 2012. p. 213–224.
- Tang Z-Z, Chen G, Alekseyenko AV. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. *Bioinformatics*. 2016;32(17):2618–25.
- Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, Zhou H, Zhou JJ, Ringel Y, Li H, Wu MC. Testing in microbiome-profiling studies with mirkat, the microbiome regression-based kernel association test. *Am J Hum Genet*. 2015;96(5):797–807.
- 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.
- 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):1012–21.
- Martiny JB, Jones SE, Lennon JT, Martiny AC. Microbiomes in light of traits: a phylogenetic perspective. *Science*. 2015;350(6261):9323.
- 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.
- Hu JX, Zhao H, Zhou HH. False discovery rate control with groups. *J Am Stat Assoc*. 2010;105(491):1215–27.
- Yekutieli D. Hierarchical false discovery rate—controlling methodology. *J Am Stat Assoc*. 2008;103(481):309–16.
- 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. *Front Microbiol*. 2018;9:1391.
- Kim KJ, Park J, Park S-C, Won S. Phylogenetic tree-based microbiome association test. *Bioinformatics*. 2020;36(4):1000–6.
- 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. *Front Microbiol*. 2020;11:649.
- Kindinger LM, Bennett PR, Lee YS, Marchesi JR, Smith A, Cacciatore S, Holmes E, Nicholson JK, Teoh T, MacIntyre DA. The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. *Microbiome*. 2017;5(1):6.
- Wu C, Chen J, Kim J, Pan W. An adaptive association test for microbiome data. *Genome Med*. 2016;8(1):56.
- Chen J, Li H. Kernel methods for regression analysis of microbiome compositional data. In: Hu M, Liu Y, Lin J, editors. *Topics in Applied Statistics*. New York: Springer; 2013. p. 191–201.
- Price MN, Dehal PS, Arkin AP. Fasttree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5(3):e9490.
- Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, Hwang J, Bushman FD, Collman RG. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS ONE*. 2010;5(12):e15216.
- Liu Y, Xie J. Cauchy combination test: a powerful test with analytic *p*-value calculation under arbitrary dependency structures. *J Am Stat Assoc*. 2020;115(529):393–402.
- Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. Acat: a fast and powerful *p* value combination method for rare-variant analysis in sequencing studies. *Am J Hum Genet*. 2019;104(3):410–21.
- Plantinga AM, Chen J, Jenq RR, Wu MC. pldist: ecological dissimilarities for paired and longitudinal microbiome association analysis. *Bioinformatics*. 2019;35(19):3567–75.
- Liu D, Lin X, Ghosh D. Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics*. 2007;63(4):1079–88.
- 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*. 2008;9(1):292.
- Chen J, Chen W, Zhao N, Wu MC, Schaid DJ. Small sample kernel association tests for human genetic and microbiome association studies. *Genet Epidemiol*. 2016;40(1):5–19.

31. Mosimann JE. On the compound multinomial distribution, the multivariate β -distribution, and correlations among proportions. *Biometrika*. 1962;49(1/2):65–82.
32. Tvedebrink T. Overdispersion in allelic counts and θ -correction in forensic genetics. *Theor Popul Biol*. 2010;78(3):200–10.
33. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*. 2006;93(3):491–507.
34. Stevens JR, Al Masud A, Suyundikov A. A comparison of multiple testing adjustment methods with block-correlation positively-dependent tests. *Plos ONE*. 2017;12(4):e0176124.
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with *deseq2*. *Genome Biol*. 2014;15(12):1–21.
36. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun*. 2020;11(1):1–11.
37. Zhou H, He K, Chen J, Zhang X. Linda: Linear models for differential abundance analysis of microbiome compositional data. *Genome Biol*. 2022;23(1):1–23.
38. Sobel JD. Bacterial vaginosis. *Annu Rev Med*. 2000;51(1):349–56.
39. Gutman RE, Peipert JF, Weitzen S, Blume J. Evaluation of clinical methods for diagnosing bacterial vaginosis. *Obstet Gynecol*. 2005;105(3):551–6.
40. Kahwati LC, Clark R, Berkman N, Urrutia R, Patel SV, 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*. 2020;323(13):1293–309.
41. Dols JA, Molenaar D, van der Helm JJ, Caspers MP, de Kat Angelino-Bart A, Schuren FH, Speksnijder AG, Westerhoff HV, Richardus JH, Boon ME, et al. Molecular assessment of bacterial vaginosis by lactobacillus abundance and species diversity. *BMC Infect Dis*. 2016;16(1):180.
42. Vitali B, Cruciani F, Picone G, Parolin C, Donders G, Laghi L. Vaginal microbiome and metabolome highlight specific signatures of bacterial vaginosis. *Eur J Clin Microbiol Infect Dis*. 2015;34(12):2367–76.
43. Subramaniam A, Kumar R, Cliver SP, Zhi D, Szychowski JM, Abramovici A, Biggio JR, Lefkowitz EJ, Morrow C, Edwards RK. Vaginal microbiota in pregnancy: evaluation based on vaginal flora, birth outcome, and race. *Am J Perinatol*. 2016;33(04):401–8.
44. Srinivasan S, Fredricks DN. The human vaginal bacterial biota and bacterial vaginosis. *Interdisc Perspect Infect Dis*. 2008;750479.
45. Petrova MI, Lievens E, Malik S, Imholz N, Lebeer S. Lactobacillus species as biomarkers and agents that can promote various aspects of vaginal health. *Front Physiol*. 2015;6:81.
46. 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*. 2019;8(3):00661.
47. Coleman JS, Gaydos CA. Molecular diagnosis of bacterial vaginosis: an update. *J Clin Microbiol*. 2018;56(9):00342–18.
48. van Teijlingen NH, Helgers LC, Zijlstra-Willems EM, van Hamme JL, Ribeiro CM, Strijbis K, Geijtenbeek TB. Vaginal dysbiosis associated bacteria *Megasphaera elsdenii* and *Prevotella timonensis* induce immune activation via dendritic cells. *J Reprod Immunol*. 2020;138:103085.
49. Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med*. 2005;353(18):1899–911.
50. Callahan BJ, DiGiulio DB, Goltsman DSA, Sun CL, Costello EK, Jeganathan P, Biggio JR, Wong RJ, Druzin ML, Shaw GM, et al. Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of us women. *Proc Natl Acad Sci*. 2017;114(37):9966–71.
51. Pararas M, Skevaki C, Kafetzis D. Preterm birth due to maternal infection: causative pathogens and modes of prevention. *Eur J Clin Microbiol Infect Dis*. 2006;25(9):562–9.
52. Choi SJ, Park SD, Jang IH, Uh Y, Lee A. The prevalence of vaginal microorganisms in pregnant women with preterm labor and preterm birth. *Ann Lab Med*. 2012;32(3):194–200.
53. Mendz GL, Petersen R, Quinlivan JA, Kaakoush NO. Potential involvement of *Campylobacter curvus* and *Haemophilus parainfluenzae* in preterm birth. *Case Rep*. 2014;2014:2014205282.
54. Han YW, Redline RW, Li M, Yin L, Hill GB, McCormick TS. *Fusobacterium nucleatum* induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. *Infect Immun*. 2004;72(4):2272–9.
55. 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 Int J Obstet Gynaecol*. 2019;126(3):349–58.
56. Son K-A, Kim M, Kim YM, Kim SH, 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. *Obstet Gynecol Sci*. 2018;61(1):38–47.
57. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. USA: Public Library of Science San Francisco; 2012.
58. Machado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, List M. Network analysis methods for studying microbial communities: a mini review. *Comput Struct Biotechnol J*. 2021;19:2687–98.
59. Costea PI, Zeller G, Sunagawa S, Bork P. A fair comparison. *Nat Methods*. 2014;11(4):359.
60. Paulson JN, Bravo HC, Pop M. Reply to: "a fair comparison". *Nat Methods*. 2014;11(4):359–60.
61. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol*. 2017;8:2224.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

