# Supplementary Materials

## A Computational strategies

The crux of doing inference with BCR is to calculate the marginal likelihoods $M_s(A)$. We use the computational strategies in Ma and Soriano (2018) to compute the integrals. Specifically, for fixed $\nu$, the inner integrals on the regression parameters are evaluated based on Laplace approximation; then the outer integral on $\nu$ is calculated with finite Riemann approximation.

With a bit abuse of the notations, in this section, we use $\boldsymbol{\beta}(A)$ to denote the 'active' regression coefficient in (2.10). That is, under the null, $\boldsymbol{\beta}(A)$ is just the $\boldsymbol{\beta}(A)$ in (2.10); under the alternative, $\boldsymbol{\beta}(A)$ denotes $(\boldsymbol{\beta}(A)^\top, \gamma(A))^\top$. $\boldsymbol{x}_{ij}$ is redefined to be the 'active' covariates in the same sense. With these notations, we have $g(\theta_{\boldsymbol{x}_{ij}}(A)) = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}(A)$ and $\theta_{ij}(A) \mid \boldsymbol{x}_{ij}, \nu(A) \sim \text{Beta}(\theta_{\boldsymbol{x}_{ij}}(A)\nu(A), (1-\theta_{\boldsymbol{x}_{ij}}(A))\nu(A))$ for each local beta-binomial model.

The computational strategies are the same under both hypotheses. Let $\pi_A(\boldsymbol{\beta})$ be the prior density of $\boldsymbol{\beta}(A)$ under either hypothesis. For fixed $\nu$ in the support of $G_A(\nu)$, by the Laplace approximation, the inner integral is

$$
\begin{aligned}
L_\nu(A) &= \int \prod_{i=1}^{2} \prod_{j=1}^{n_i} \mathcal{L}_{BB}(g^{-1}(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}), \nu \mid y_{ij}(A_l), y_{ij}(A_r)) \pi_A(\boldsymbol{\beta}) d\boldsymbol{\beta} \\
&= \int \exp\left\{ \sum_{i,j} \log \mathcal{L}_{BB}(g^{-1}(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}), \nu \mid y_{ij}(A_l), y_{ij}(A_r)) + \log \pi_A(\boldsymbol{\beta}) \right\} d\boldsymbol{\beta} \\
&= \int \exp\{h_\nu(\boldsymbol{\beta})\} d\boldsymbol{\beta} \\
&\approx \exp\left\{ h_\nu(\hat{\boldsymbol{\beta}}_\nu) \right\} \cdot (2\pi)^{d/2} \cdot |-H_\nu(\hat{\boldsymbol{\beta}}_\nu)|^{-1/2} \\
&= \hat{L}_\nu(A)
\end{aligned}
$$

where $h_\nu(\boldsymbol{\beta}) = \sum_{i,j} \log \mathcal{L}_{BB}(g^{-1}(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}), \nu \mid y_{ij}(A_l), y_{ij}(A_r)) + \log \pi_A(\boldsymbol{\beta})$, $\hat{\boldsymbol{\beta}}_\nu$ is the maximizer of $h_\nu(\boldsymbol{\beta})$, $H_\nu(\hat{\boldsymbol{\beta}}_\nu)$ is the Hessian matrix of $h_\nu(\boldsymbol{\beta})$ at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_\nu$. $d$ is the degrees of freedom of $\boldsymbol{\beta}$, which is $(p+2)$ under the alternative and $(p+1)$ under the null. We describe a Newton-

Raphson algorithm to solve for $\hat{\boldsymbol{\beta}}_\nu$ below. The log-likelihood function is strictly log-concave and the Newton-Raphson method generally converges after only a few iterations. Finally, to get $M_s(A)$, we compute the outer integral on $\nu$, $\int L_\nu(A)dG_A(\nu)$, with finite Riemann approximations. Specifically, after calculating $\hat{L}_\nu(A)$ at a grid of $\nu$'s: $\nu_1, \nu_2, \ldots, \nu_M$, we have

$$\int L_\nu(A)dG_A(\nu) \approx \sum_{m=2}^{M} \hat{L}_{\nu_m}(A)(G_A(\nu_m) - G_A(\nu_{m-1})).$$

*Newton-Rhaphson for* $\hat{\boldsymbol{\beta}}_\nu$. In this subsection, we shall fix our attention on a specific node $A$ and suppress the '(A)' in the notations. Moreover, we let $y_{ij}(A_l) = t_{ij}$ for simplicity and express the local Beta-Binomial regression model on $A$ as

$$t_{ij} \sim \text{Binomial}(y_{ij}, \theta_{ij}), \quad \theta_{ij} \sim \text{Beta}(\theta_{\boldsymbol{x}_{ij}}\nu, (1-\theta_{\boldsymbol{x}_{ij}})\nu), \quad \text{and} \quad g(\theta_{\boldsymbol{x}}) = \boldsymbol{x}^\top \boldsymbol{\beta}.$$

The contribution to the log marginal likelihood from the $j$-th observation in group $i$ is

$$\begin{aligned} l_{ij} &= \log \mathcal{L}_{BB}(g^{-1}(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta}), \nu \mid t_{ij}, y_{ij} - t_{ij}) \\ &= \log \Gamma(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) + \log \Gamma((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \log \Gamma(\nu + y_{ij}) \\ &\quad - \log \Gamma(\theta_{\boldsymbol{x}_{ij}}\nu) - \log \Gamma((1-\theta_{\boldsymbol{x}_{ij}})\nu) + \log \Gamma(\nu). \end{aligned}$$

Taking the first derivative w.r.t. $\boldsymbol{\beta}$,

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}} = \frac{\partial l_{ij}}{\partial \theta_{\boldsymbol{x}_{ij}}} \cdot \frac{\partial \theta_{\boldsymbol{x}_{ij}}}{\partial \eta_{ij}} \cdot \frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}$$

where $\eta_{ij} = \boldsymbol{x}_{ij}^\top \boldsymbol{\beta}$. Now with $\phi$ denoting the digamma function,

$$\frac{\partial l_{ij}}{\partial \theta_{\boldsymbol{x}_{ij}}} = \nu[\phi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) - \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \phi(\theta_{\boldsymbol{x}_{ij}}\nu) + \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu)].$$

With the logit link, $\theta_{\boldsymbol{x}_{ij}} = g^{-1}(\eta_{ij}) = 1/(1+e^{-\eta_{ij}})$, and

$$\frac{\partial \theta_{\boldsymbol{x}_{ij}}}{\partial \eta_{ij}} = (g^{-1})'(\eta_{ij}) = \theta_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}}).$$

Thus

$$\frac{\partial l_{ij}}{\partial \boldsymbol{\beta}} = \nu \theta_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}})[\phi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) - \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \phi(\theta_{\boldsymbol{x}_{ij}}\nu) + \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu)]\boldsymbol{x}_{ij}.$$

The second derivative of $l_{ij}$ w.r.t. $\boldsymbol{\beta}$ is

$$\frac{\partial^2 l_{ij}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} = \frac{\partial^2 l_{ij}}{\partial \theta^2_{\boldsymbol{x}_{ij}}} \cdot \left(\frac{\partial \theta_{\boldsymbol{x}_{ij}}}{\partial \eta_{ij}}\right)^2 \cdot \left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}^\top}\right) + \frac{\partial l_{ij}}{\partial \theta_{\boldsymbol{x}_{ij}}} \cdot \frac{\partial^2 \theta_{\boldsymbol{x}_{ij}}}{\partial \eta^2_{ij}} \cdot \left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}^\top}\right) + \frac{\partial l_{ij}}{\partial \theta_{\boldsymbol{x}_{ij}}} \cdot \frac{\partial \theta_{\boldsymbol{x}_{ij}}}{\partial \eta_{ij}} \frac{\partial^2 \eta_{ij}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top}.$$

The third term on the right-hand side is equal to zero. With $\psi$ being the trigamma function,

$$\frac{\partial^2 l_{ij}}{\partial \theta^2_{\boldsymbol{x}_{ij}}} = \nu^2[\psi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) + \psi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \psi(\theta_{\boldsymbol{x}_{ij}}\nu) - \psi((1-\theta_{\boldsymbol{x}_{ij}})\nu)].$$

Thus the first term is

$$\frac{\partial^2 l_{ij}}{\partial \theta^2_{\boldsymbol{x}_{ij}}} \cdot \left(\frac{\partial \theta_{\boldsymbol{x}_{ij}}}{\partial \eta_{ij}}\right)^2 \cdot \left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)^\top$$
$$= \nu^2[\psi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) + \psi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \psi(\theta_{\boldsymbol{x}_{ij}}\nu) - \psi((1-\theta_{\boldsymbol{x}_{ij}})\nu)]\theta^2_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}})^2\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^\top.$$

The second term, which has expectation zero, is

$$\frac{\partial l_{ij}}{\partial \theta_{\boldsymbol{x}_{ij}}} \cdot \frac{\partial^2 \theta_{\boldsymbol{x}_{ij}}}{\partial \eta^2_{ij}} \cdot \left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \eta_{ij}}{\partial \boldsymbol{\beta}}\right)^\top$$
$$= \nu[\phi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) - \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \phi(\theta_{\boldsymbol{x}_{ij}}\nu) + \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu)]\theta_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}})(1-2\theta_{\boldsymbol{x}_{ij}})\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^\top.$$

For each $i = 1, 2,\ j = 1, 2, \ldots n_i$, let

$$a_{ij} = \phi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) - \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \phi(\theta_{\boldsymbol{x}_{ij}}\nu) + \phi((1-\theta_{\boldsymbol{x}_{ij}})\nu)$$
$$b_{ij} = \psi(\theta_{\boldsymbol{x}_{ij}}\nu + t_{ij}) + \psi((1-\theta_{\boldsymbol{x}_{ij}})\nu + y_{ij} - t_{ij}) - \psi(\theta_{\boldsymbol{x}_{ij}}\nu) - \psi((1-\theta_{\boldsymbol{x}_{ij}})\nu).$$

Since the total log likelihood is $l = \sum_{i,j} l_{ij}$,

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \nu \sum_{i,j} a_{ij}\theta_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}})\boldsymbol{x}_{ij} = \nu \boldsymbol{X}^\top W_1 \boldsymbol{z},$$

where the rows of $\boldsymbol{X}$ are $\boldsymbol{x}_{ij}^\top$, $W_1 = \mathrm{diag}(a_{ij})$ and $\boldsymbol{z} = (\theta_{\boldsymbol{x}_{11}}(1-\theta_{\boldsymbol{x}_{11}}), \ldots, \theta_{\boldsymbol{x}_{1n_1}}(1-\theta_{\boldsymbol{x}_{1n_1}}), \ldots, \theta_{\boldsymbol{x}_{2n_2}}(1-\theta_{\boldsymbol{x}_{2n_2}}))^\top$. The rows of $\boldsymbol{X}, W_1$ and the elements of $\boldsymbol{z}$ are ordered first by $j$ and then $i$.

$$\begin{aligned}\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= \nu^2 \sum_{i,j} b_{ij}\theta_{\boldsymbol{x}_{ij}}^2(1-\theta_{\boldsymbol{x}_{ij}})^2\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^\top + \nu\sum_{i,j} a_{ij}\theta_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}})(1-2\theta_{\boldsymbol{x}_{ij}})\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^\top \\ &= -\nu\boldsymbol{X}^\top W_2 \boldsymbol{X},\end{aligned}$$

where $W_2 = -\mathrm{diag}(\nu b_{ij}\theta_{\boldsymbol{x}_{ij}}^2(1-\theta_{\boldsymbol{x}_{ij}})^2 + a_{ij}\theta_{\boldsymbol{x}_{ij}}(1-\theta_{\boldsymbol{x}_{ij}})(1-2\theta_{\boldsymbol{x}_{ij}}))$. The columns of $W_2$ is also ordered first by $j$ and then by $i$.

When applying Laplace approximation to evaluate the marginal likelihood for a fixed $\nu$,

$$L_\nu = \int \exp\{l(\boldsymbol{\beta}) + \log \pi(\boldsymbol{\beta})\} d\boldsymbol{\beta},$$

where $\pi$ is the prior on $\boldsymbol{\beta}$. For example, with $\pi(\boldsymbol{\beta})$ is the independent normal $\mathrm{N}(0, \sigma_k^2)$ on the $k$-th element of $\boldsymbol{\beta}$, let $h_\nu(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \log\pi(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \boldsymbol{\beta}^\top\Sigma^{-1}\boldsymbol{\beta}/2$, where $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_d^2)$, we have

$$\begin{aligned}\frac{\partial h_\nu(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{\partial l}{\partial \boldsymbol{\beta}} - \Sigma^{-1}\boldsymbol{\beta} \\ \frac{\partial^2 h_\nu(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} &= \frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^\top} - \Sigma^{-1}.\end{aligned}$$

Hence the Newton-Raphson step for solving the MLE of $\boldsymbol{\beta}$ given $\nu$ is given by

$$\hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + \left(\boldsymbol{X}^\top W_2^{(t)} \boldsymbol{X} + \Sigma^{-1}/\nu\right)^{-1} \left(\boldsymbol{X}^\top W_1^{(t)} \boldsymbol{z}^{(t)} - \Sigma^{-1}\hat{\boldsymbol{\beta}}^{(t)}/\nu\right).$$

Under the alternative, suppose that $\pi(\boldsymbol{\beta}) = \pi(\boldsymbol{\beta}_1)\pi(\gamma)$, where $\boldsymbol{\beta}_1$ are the coefficients for the covariates and $\gamma$ for the group indicator. Instead of using independent normal prior on $\gamma$, the LIM $g$-prior (Li and Clyde, 2015) could be adopted. Using the independent normal prior for $\boldsymbol{\beta}_1$, we have

$$\begin{aligned} h_\nu(\boldsymbol{\beta}) &= l(\boldsymbol{\beta}) + \log \pi(\boldsymbol{\beta}) \\ &= l(\boldsymbol{\beta}) - \boldsymbol{\beta}_1^\top \Sigma^{-1} \boldsymbol{\beta}_1/2 - g^{-1} \mathcal{J}_\nu(\hat{\gamma})\gamma^2/2 \\ &= l(\boldsymbol{\beta}) - \boldsymbol{\beta}_1^\top \Sigma^{-1} \boldsymbol{\beta}_1/2 - g^{-1} \nu (\boldsymbol{X}^\top \hat{W}_2 \boldsymbol{X})_2 \gamma/2 \end{aligned}$$

where $(\boldsymbol{X}^\top \hat{W}_2 \boldsymbol{X})_2$ denote the block of the Hessian matrix corresponding to $\gamma$. Therefore,

$$\begin{aligned} \frac{\partial h_\nu(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \nu \boldsymbol{X}^\top W_1 \boldsymbol{z} - \begin{pmatrix} \Sigma^{-1}\boldsymbol{\beta}_1 \\ g^{-1}\nu(\boldsymbol{X}^\top \hat{W}_2 \boldsymbol{X})_2 \gamma \end{pmatrix} \\ \frac{\partial^2 h_\nu(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= -\nu \boldsymbol{X}^\top W_2 \boldsymbol{X} - \begin{pmatrix} \Sigma^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & g^{-1}\nu(\boldsymbol{X}^\top \hat{W}_2 \boldsymbol{X})_2 \end{pmatrix}. \end{aligned}$$

The resulting NR update is

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(t+1)} = \hat{\boldsymbol{\beta}}^{(t)} + & \left(\boldsymbol{X}^\top W_2^{(t)} \boldsymbol{X} + \begin{pmatrix} \Sigma^{-1}/\nu & \boldsymbol{0} \\ \boldsymbol{0} & g^{-1}(\boldsymbol{X}^\top W_2^{(t)} \boldsymbol{X})_2 \end{pmatrix}\right)^{-1} \\ & \times \left(\boldsymbol{X}^\top W_1^{(t)} \boldsymbol{z}^{(t)} - \begin{pmatrix} \hat{\Sigma}^{-1}\boldsymbol{\beta}_1^{(t)}/\nu \\ g^{-1}(\boldsymbol{X}^\top W_2^{(t)} \boldsymbol{X})_2 \hat{\gamma}^{(t)} \end{pmatrix}\right). \end{aligned}$$

## B More on decision making

We first consider the original hypothesis that there is no cross-group difference. Taking a decision theoretic perspective, let $d(\boldsymbol{y}) \in \{0, 1\}$ be some decision rule, with $d(\boldsymbol{y}) = 1$ corresponding to the rejection of the global null that there are no cross-group differences in the OTU composition. When the loss function is

$$L(d(\boldsymbol{y}), c) = c \cdot \mathbb{1}_{[H_0 \text{ is true}]} d(\boldsymbol{y}) + (1 - c) \cdot \mathbb{1}_{[H_1 \text{ is true}]} (1 - d(\boldsymbol{y}))$$

for some $0 \leq c \leq 1$, one can show that the Bayes optimal decision rule is $d(\boldsymbol{y}) = \mathbb{1}_{[\text{PJAP}>c]}$. In particular, when $c = 0.5$, this gives the optimal decision under the simple 0-1 loss.

The decision on reporting the significant nodes is essentially a multiple testing problem. One way to address this problem is to use loss functions specified with the false positives and false negatives (Müller et al., 2006). For example, let $d_i(\boldsymbol{y}) \in \{0, 1\}$ be the decision rule on the $i$-th node; again, $d_i(\boldsymbol{y}) = 1$ corresponds to the rejection of the node-specific null. Let FD and FN denote the number of false positives and false negatives. The posterior expectation of FD and FN are

$$\overline{\text{FD}} = \sum (1 - \text{PMAP}_i) \times d_i(\boldsymbol{y}),$$
$$\overline{\text{FN}} = \sum \text{PMAP}_i \times (1 - d_i(\boldsymbol{y})).$$

It can be shown that under the loss $L(d(\boldsymbol{y}), t) = t \times \text{FD} + \text{FN}$, the Bayes optimal decision rule, which minimizes the posterior expected loss $\overline{L}(d(\boldsymbol{y}), t) = t \times \overline{\text{FD}} + \overline{\text{FN}}$ has the form $d_i(\boldsymbol{y}) = \mathbb{1}_{[\text{PMAP}_i > c']}$ with the optimal threshold $c' = t/(t + 1)$, $t \geq 0$ (Müller et al., 2004). In our application, we use $c' = 0.5$ that corresponds to $t = 1$ which is also recommended by Barbieri et al. (2004) from a Bayesian model choice perspective. Note that one can also consider loss functions that directly take into account the dependency among the hypotheses being tested. In our framework, such dependency is incorporated only through the probability model, not in the decision theoretic part.

## C Covariate selection

As we noted in Section 2.4, covariate selection is achievable in BGCR by putting a spike-and-slab prior on the regression coefficients (George and McCulloch, 1997). For example, let $r_l \overset{\text{ind}}{\sim} \text{Bernoulli}(q_l)$ where $q_l \in (0,1)$, $l = 2, \ldots, p+1$. For $A \in \mathcal{I}$, we can modify the prior on $\boldsymbol{\beta}(A)$ to be

$$\beta_l(A) \overset{\text{ind}}{\sim} (1 - r_l)\delta_0 + r_l \text{N}(0, \sigma_l^2(A)), \quad l = 2, \ldots, p+1, \tag{S1}$$

where $\delta_0$ is a point mass at zero, $\sigma_l^2(A)$'s are chosen for $\text{N}(0, \sigma_l^2(A))$ to cover all reasonable values of $\beta_l(A)$ while not supporting unreasonable values of $\beta_l(A)$.

Let $\boldsymbol{r} = (r_2, \ldots, r_{p+1}) \in \{0,1\}^p$. The independent Bernoulli priors on $r_l$ induce the following prior on $\boldsymbol{r}$

$$\pi(\boldsymbol{r}) = \prod_{l=2}^{p+1} q_l^{r_l}(1 - q_l)^{1-r_l}.$$

Conditioning on $\boldsymbol{r}$, the marginal likelihood of the data, $\phi_1(\Omega \mid \boldsymbol{r})$, is available as a byproduct of the BGCR inference algorithm (Section 2.4). When the number of covariates is not too large, this allows us to get the posterior of $\boldsymbol{r}$ by Bayes theorem:

$$\pi(\boldsymbol{r} \mid \boldsymbol{Y}) \propto \pi(\boldsymbol{r})\phi_1(\Omega \mid \boldsymbol{r}).$$

We modify our simulation scenario IV in Section 3.3 to give a simple illustration of the covariate selection procedure. Consider the data simulated under the alternative, in which the counts of OTU '4481131' ($\omega_s$) are increased by 175% in the second group. Instead of using "gender" as a confounder, we generate two covariates for each sample:

$$x_{ij2} \overset{\text{iid}}{\sim} \text{N}(0,1), \quad x_{ij3} \overset{\text{iid}}{\sim} \text{N}(0,1).$$

Suppose that the first covariate is relevant to the counts of a specific OTU while the second covariate has nothing to do with the OTU counts. Specifically, we increase the counts of OTU '4352657' ($\omega_c$) in the $j$-th sample in group $i$ by ($x_{ij2} \times 175\%$) (when this value is less than $-1$, we set the count to zero). We note that due to the large variation in OTU counts, the signal injected on $\omega_c$ is quite weak.

Consider a specific round of simulation. We let $r_2 = r_3 = 0.5$, $\sigma_l^2(A) = 10$ for $l = 2, 3$ and fit BGCR with the prior in (S1). Table S1 summaries the posterior probabilities of the four possible models. In comparison, each model has equal prior probabilities. Therefore, the important variable is correctly identified.

| Covariate in the model | None | 2 | 3 | 2 and 3 |
|---|---|---|---|---|
| Posterior probability | 0.223 | 0.320 | 0.186 | 0.270 |

Table S1: Posterior probabilities of different models (no confounding).

Although a covariate selection procedure can be incorporated in BGCR, one must proceed with caution since this can substantially affect or even invalidate the meaning of the testing result on the two-group difference. To see this intuitively, consider the following simplistic but representative scenario. Suppose there is a (close-to) perfect confounding covariate which explains virtually all the difference across the two groups. Once this covariate is included in the model then there is no remaining cross-group difference and the two-group comparison will not favor the alternative. However, including the covariate into the model may not improve the fit to the observed data in any substantive manner as its effect is largely overlapping with that of the intercept (i.e., the group label). Consequently, statistical model selection strategies, both Bayesian or frequentist, would very likely to exclude this covariate from the model. This would lead to a significant testing result on the two-group differences. As a simple illustration, in the previous example, suppose instead we have

$$x_{1j2} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1), \quad x_{2j2} \overset{\text{iid}}{\sim} \mathrm{N}(2, 1).$$

In this case, the first covariate is a confounding variable. Table S2 summaries the posterior probabilities of the four possible models. Due to the strong confounding effect, the first covariate is excluded from the model, which would lead to false positives in the testing scenario.

| Covariate in the model | None | 2 | 3 | 2 and 3 |
|---|---|---|---|---|
| Posterior probability | 0.599 | 0.001 | 0.400 | $\approx 0$ |

Table S2: Posterior probabilities of different models (with confounding).

## D Additional figures

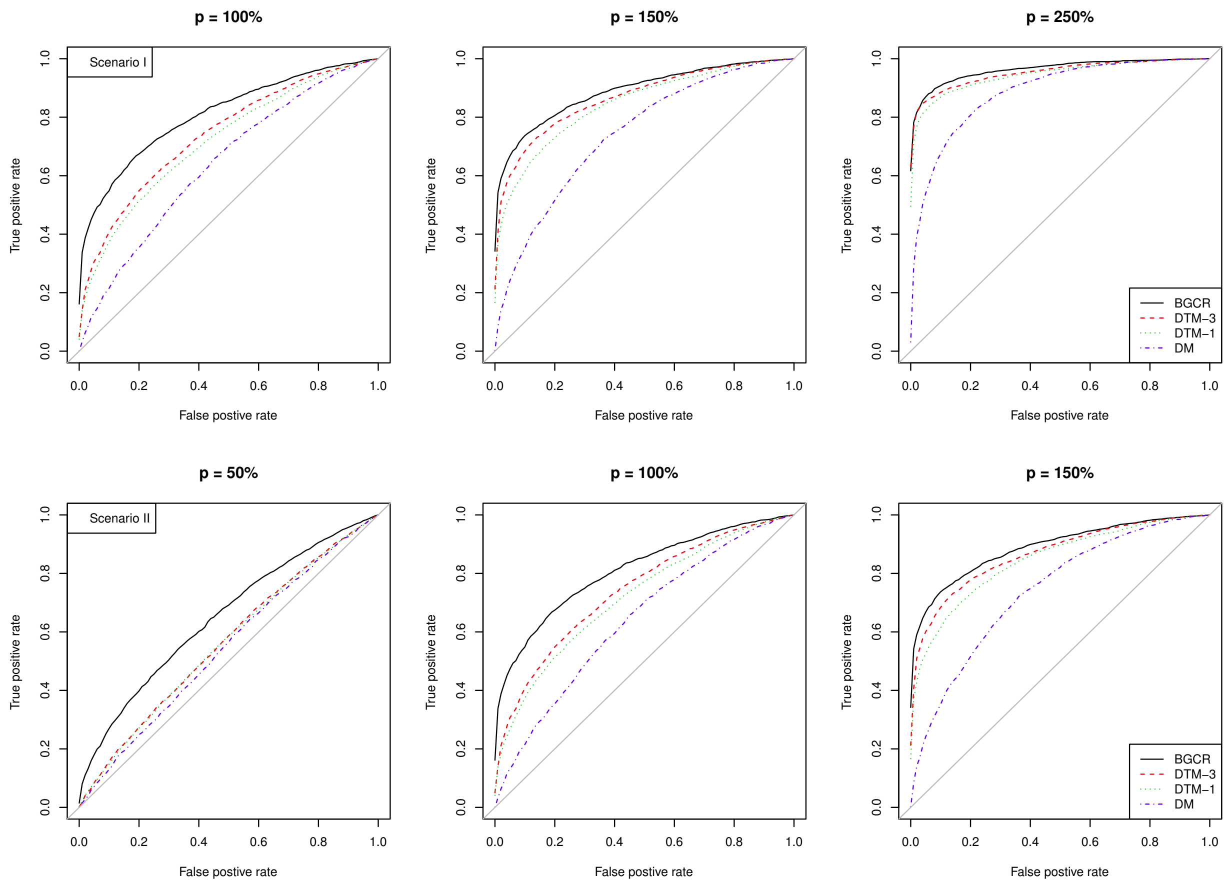


Figure S1: ROC curves for Scenario I and II with $K = 50$. The columns are indicated by the percent of count increased in the second group ($p$).

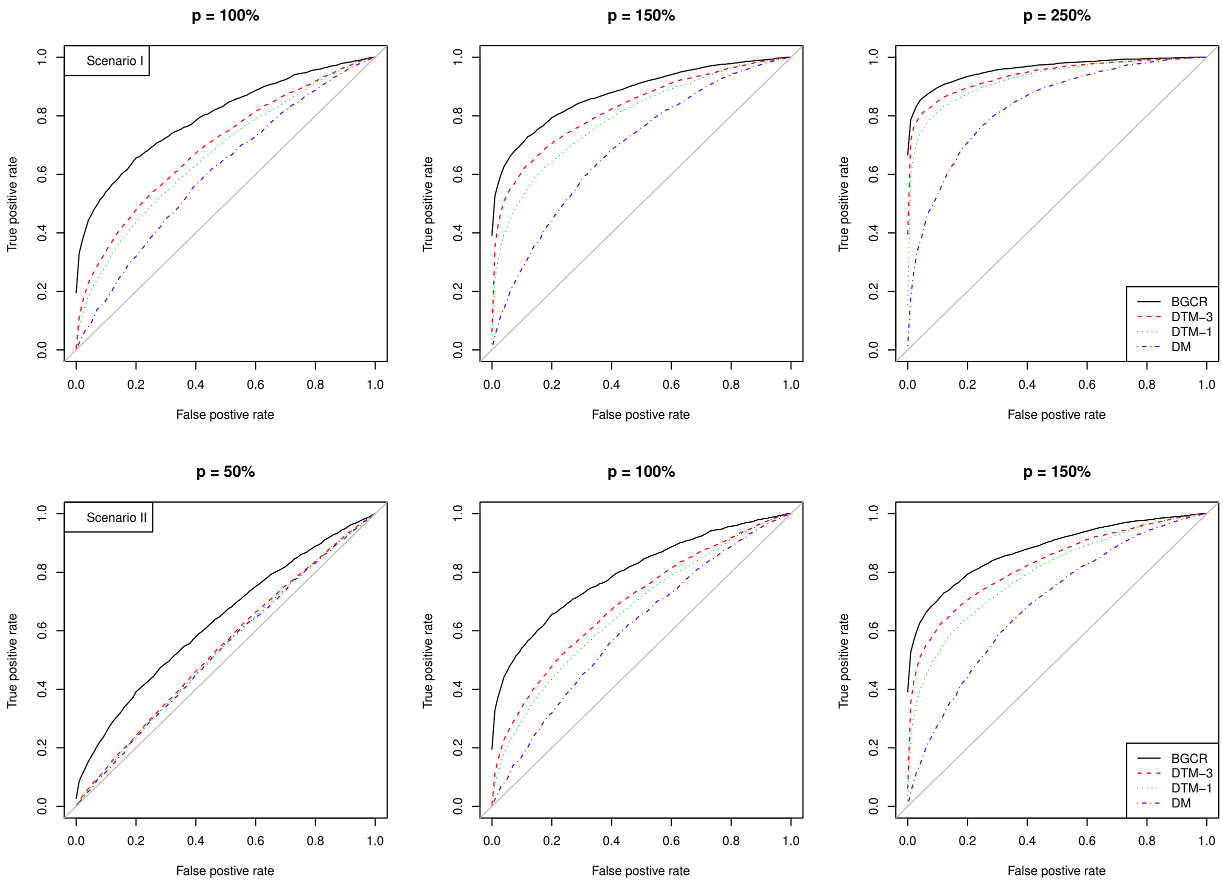


Figure S2: ROC curves for Scenario I and II with $K = 75$. The columns are indicated by the percent of count increased in the second group ($p$).

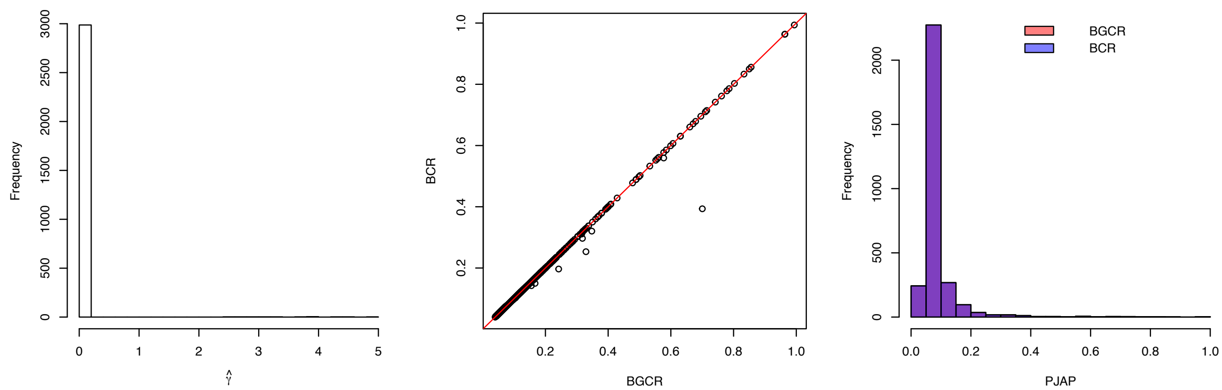


Figure S3: BGCR vs BCR under the null in scenario 3. Left: Histogram of the estimated $\gamma$ in BGCR; Middle: PJAPs of BGCR vs BCR; Right: Histograms of the PJAPs of BGCR and BCR.

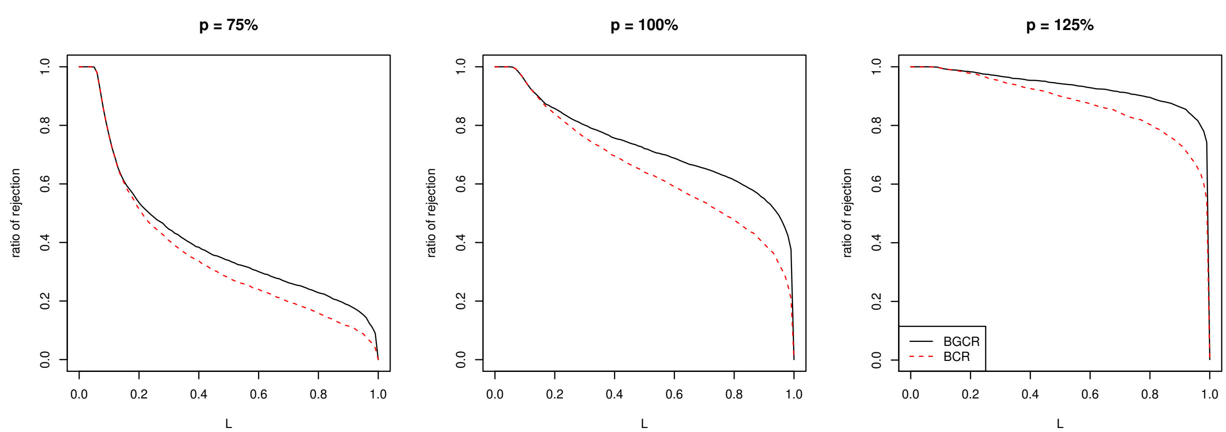


Figure S4: Ratio of rejection under the alternatives in Scenario III. The columns are indicated by the percent of count increased in the second group ($p$).

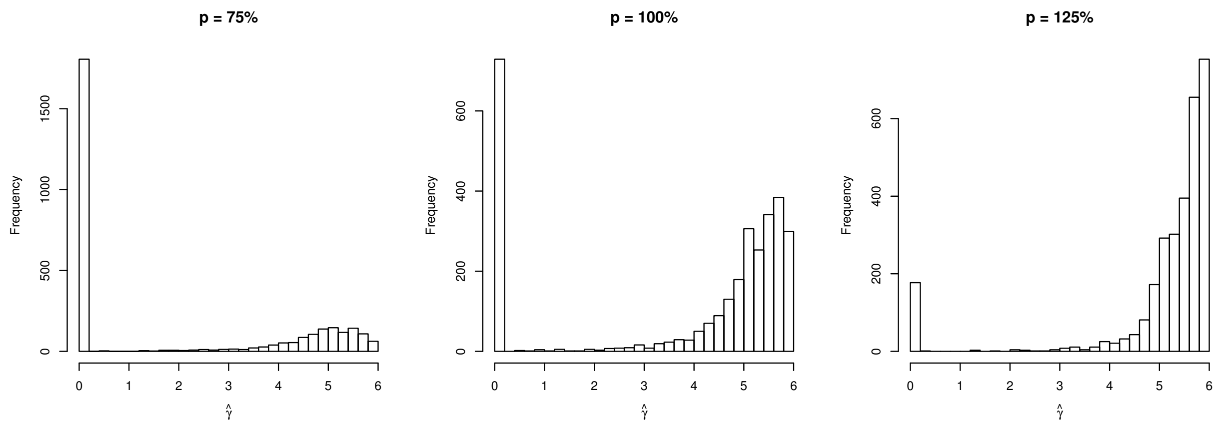


Figure S5: Estimated $\gamma$ under the alternatives in scenario 3.

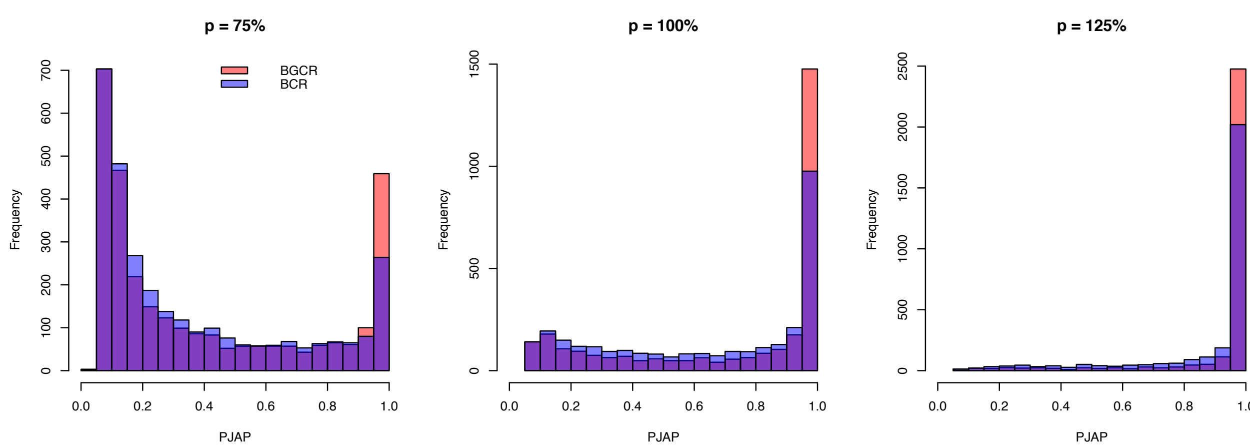


Figure S6: Histograms of the PJAPs under the alternatives in scenario 3.

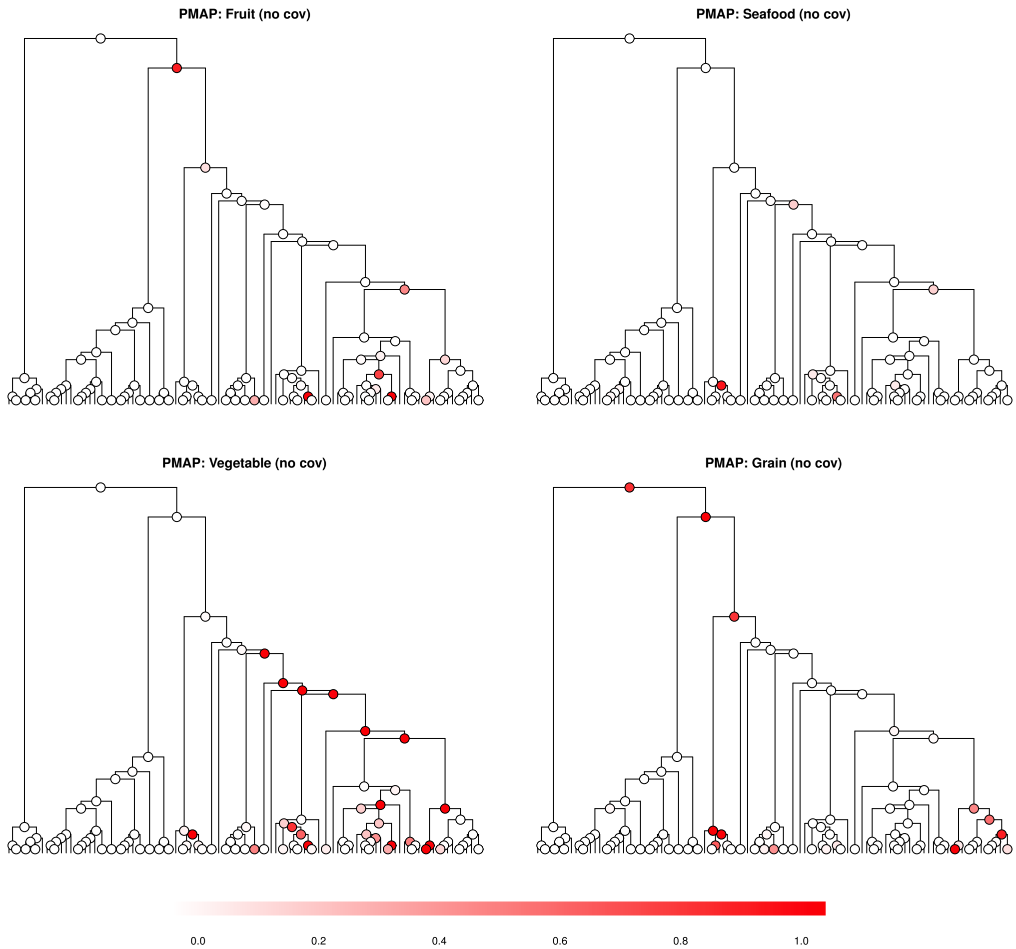


Figure S7: PMAPs for the four comparisons that reject the global null. The nodes are colored by PMAPs reported by BGCR with no covariate adjusted.

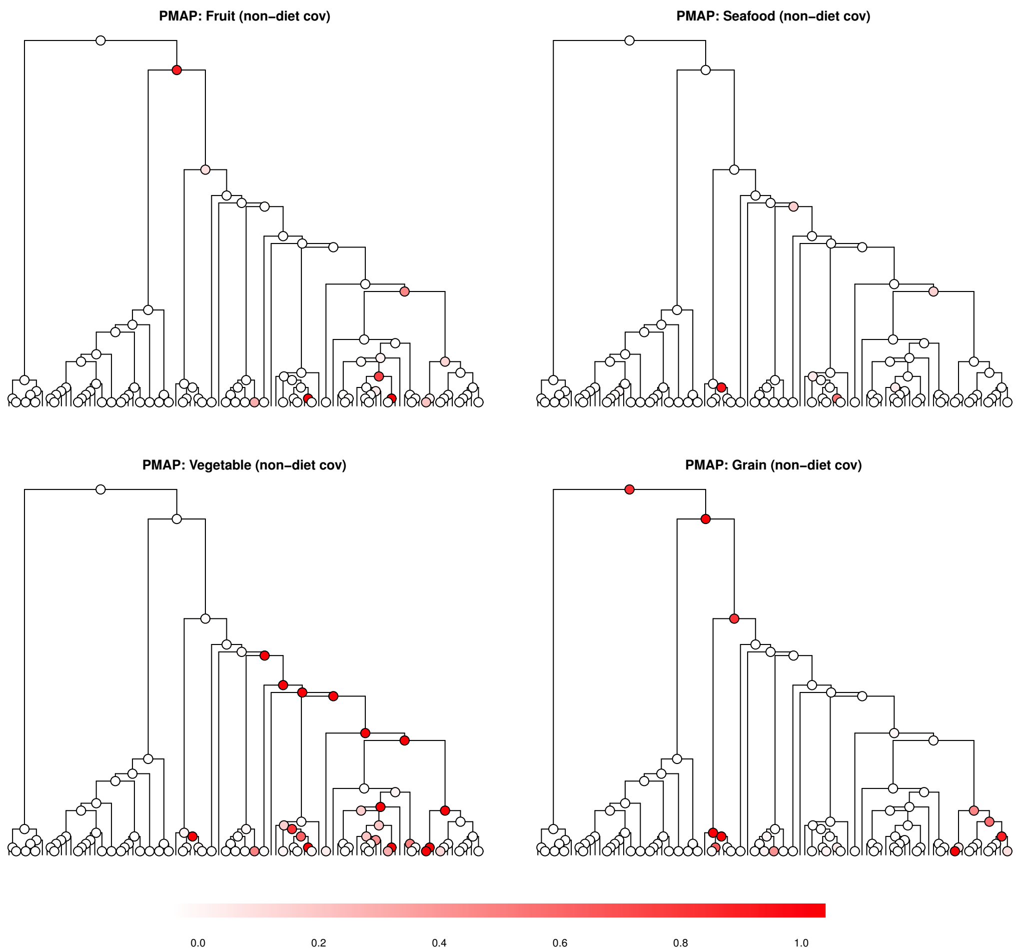


Figure S8: PMAPs for the four comparisons that reject the global null. The nodes are colored by PMAPs reported by BGCR with only non-dietary covariates adjusted.

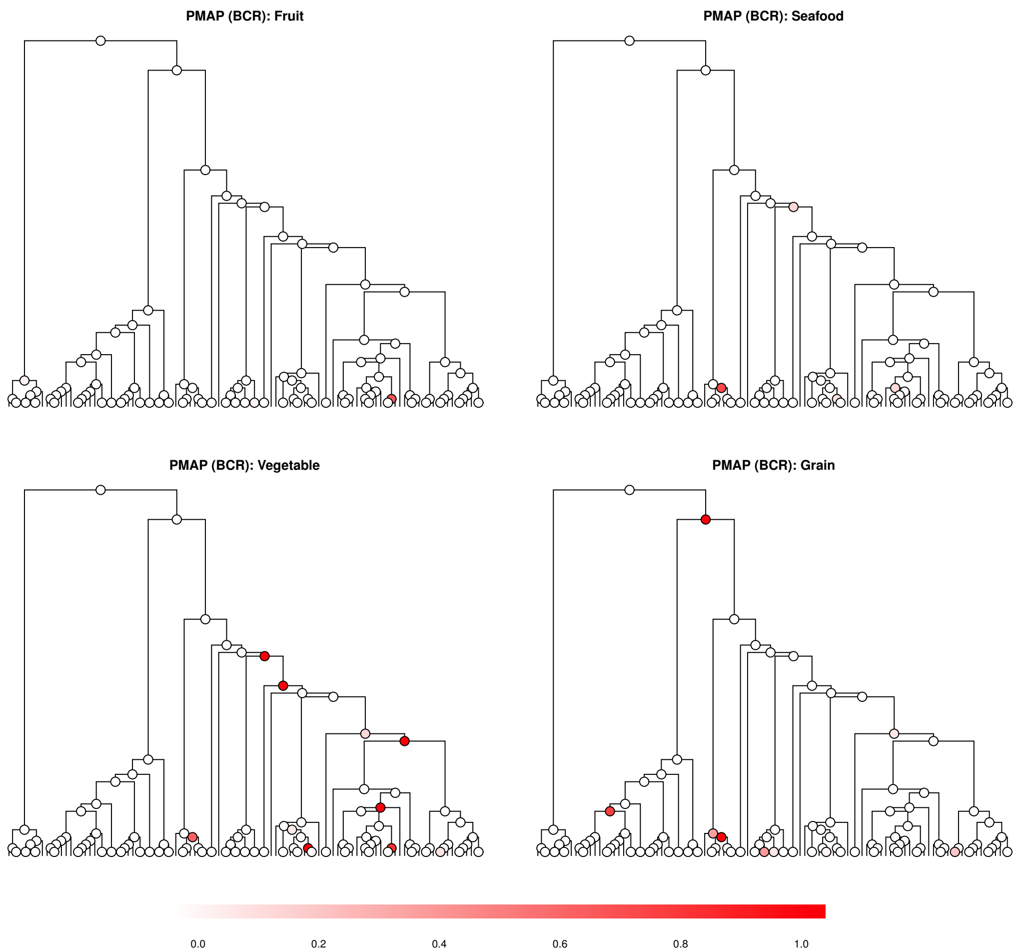


Figure S9: PMAPs for the four comparisons that reject the global null. The nodes are colored by PMAPs reported by BCR with both non-dietary covariates and dietary covariates adjusted.