This supplementary section consists of three subsections that correspond to some of the aspects of the differential abundance simulations presented in this paper and in McMurdie and Holmes (2014). In subsection A we demonstrate that the simulation set-up of McMurdie and Holmes (2014) does not necessarily contain true nulls but all are non-null hypotheses. Therefore, all discoveries made by a procedure are true discoveries (non-nulls), there are no false discoveries. In subsection B we describe our refinement of McMurdie and Holmes (2014) set-up so that the set of hypotheses to be tested includes some true nulls and some true non-nulls. In doing so, we can estimate the false discovery rate of any given method (i.e. among all hypotheses rejected, the expected proportion of null hypotheses rejected). In subsection C, we describe why ANCOM (Mandal *et al*., 2015) has a better control of false discovery rate than any of the other methods considered in this paper when testing hypotheses regarding taxa abundance in two ecosystems (or environments) of unit volume.

1. **(McMurdie and Holmes, 2014 [**[**1**](#_ENREF_1)**]) ‘Original’ Simulation corresponding to Figure S5a**

**Step 1 (Generate samples to form Group 1):** Take the raw data of counts available in one of the environments studied in Caporaso et al. [[2](#_ENREF_2)], e.g. freshwater – creek, replicate 1. Suppose is the sum of observed abundances of taxa in all the samples and let.

* Step 1: Generate multinomial random vector  to represent the taxa counts for the -th sample.
* Steps 1 repeatedtimes to obtain random samples to represent Group 1.

**Step 2 (Generate samples to form Group 2):**

* Step 2(a). As in Step 1, generate multinomial random vector  to represent the taxa counts for the -th sample. By construction,and  are identically and independently distributed.
* According to Figure 2B on page 6 of McMurdie and Holmes (2014) [[1](#_ENREF_1)], the OTU counts for a random subset of taxa (e.g. 30 in their paper) is given by, where represents the effect size (e.g., 10) and the rest are taken to be. In their paper, they take To restate more precisely:



**Note :** By construction, for the-th sample and -th taxa,, 

Unlessor , in general Hence the simulations conducted according to Figure 2B in McMurdie and Holmes (2013) [[1](#_ENREF_1)] there will be no non-differentially abundant taxa by this simulation set up. So it is not surprising that false positive rates are out of control.

1. **‘Balanced Simulation’ corresponding to Figure S5a**

**Step1:** Take the raw data of counts available in freshwater – creek, replicate 1. Suppose is the sum of observed abundances of taxa in all the samples, where. We define

Let 

and 

Note that. The relative abundance of the-th taxa in Group 1 is given by and in Group 2 it is given by

**Step 2(a)** **Generation of Group 1 samples:** Generate a random library size for the -th sample from the collection of libraries in Caporaso *et al.* [[2](#_ENREF_2)]. Then, using the relative abundance vector, generate multinomial random vector  to represent the taxa counts for the -th sample in Group 1. Similarly, by generating a random library size for the -th sample from the collection of libraries in Caporaso *et al.* [[2](#_ENREF_2)] and using the relative abundance vector, we generate multinomial random vector  to represent the taxa counts for the -th sample in Group 2.

Observe that



and



Hence we have



Thus first components are true positives because for  and the last are true negatives because  for.

1. **Inference regarding taxa abundance: Figure S5b**

Suppose anddenote the true abundance of thetaxon,, in the environments A and B respectively. Suppose , but for all. Thus, taxon 1 is true positive and rest of the taxa, , are true negatives in the two environments A and B.

For, , let and. Suppose a researcher draws samples from a multinomial distribution with probability vectors and

then note that, for, , even though.

Among all the methods discussed in the paper, ANCOM (Mandal *et al.,* 2015) is the only method that is designed for testing the null hypotheses and hence is expected to provide a better control of false discovery rate than any of the other methods because none of the other methods described in this paper are designed to test hypotheses regarding the actual parameters of interest, namely, and. Consequently, they are expected to have inflated false discovery rates. Previously, Mandal *et al.* (2015) demonstrated the FDR control of ANCOM using a Poisson-Gamma simulation set-up and we now duplicate those results for multinomial data [3].

**References**

1. McMurdie PJ, Holmes S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS computational biology. 2014;10(4). doi 10.1371/Journal.Pcbi.1003531.

2. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences of the United States of America. 2011;108 Suppl 1:4516-22. doi:10.1073/pnas.1000080107.

3. Mandal S, Van Treuren W, White RA, Eggesbø M, Knight R, Peddada SD (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. Microbial Ecology in Health and Disease, 26, 1 – 7.