Simulation strategies:

There are 3 possible simulation strategies:

* The multinomial method based on McMurdie and Holmes (2013)
  + The balanced update based on [Weiss et al. 2017](https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-017-0237-y#Sec16).
  + Dirichlet distribution based on Chen et al. 2013 which is now propagated to [Xiao et al. 2018](https://www.frontiersin.org/articles/10.3389/fmicb.2018.03112/full) as well as [Ma et al. 2020](https://academic.oup.com/bioinformatics/article/36/13/3959/5822876?rss=1#205233180)
    - Generating the probability vectors for the multinomial distribution using the Dirichlet approximation.
* NORTA method with constrained correlated structure based on the [SpiecEasi approach](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004226#sec002)

Simulation procedure

* Basic data:
  + Generate data by first define a data set to calibrate to (HMP 16S dataset)
    - Remove all taxa with 0 abundance across all stool sites
    - Sample with replacement from the total read count for each sample to get a read count
    - Use read count and the proportion of selected taxa across all samples as the basis for the Dirichlet distribution
    - Use sampling from Dirichlet distribution to generate basic data using the multinomial distribution
* Simulation 1: FDR and power. For each sample choose random taxa to get elevated. Evaluate the potential hypothesis testing capacity of this
  + Generate taxa that are elevated using the procedure defined in McMurdie and Holmes for simulation
* Simulation 2: Differential abundance analysis. Check to see whether or not the method can distinguish between elevated taxa groups across case/control status
  + Generate abundance analysis using balanced simulation adjustment from Weiss et al. 2017
* Simulation 3: Predictive performance. Taxa enrichment scores can be used as predictors to a prediction model.
  + Generate outcome similar to that of Xiao et al. 2018