MICB 475 Meeting #6 Minutes
2023-11-01
Minutes: Emily
Virtual Zoom meeting

Reading week team meetings: by request → we will have to let Avril know if we want to meet during the week of reading week

1. Want to remove NA from DESEQ plot on x axis (mixed in with genus names; attempt to filter out genus from phyloseq didn't solve the problem)

- NAs are caused from ASVs not being annotated to the genus level
- When generating this plot, some sort of dataframe should be used as an input, in the dataset → should be able to filter genus column and remove anything with an NA value
- In the merged results table → filter out !is.na (anything that's not NA), does not = "NA.1" and "NA.2"

2. PiCRUST analysis - what's our input? We found some sources saying biom / some saying qza. → not quite sure where to progress

- Google "PiCRUST qiime example code" to see how other people are using this analysis
- Stack overflow: helpful resource for coding
- Inputs = table.qza, rep-seqs.qza
- .biom file format overlaps significantly with .qza files, so likely both are allowed, however Avril has always used .qza.
- PiCRUST results do not change based on if all the samples are run at the same time or if they are divided up and run separately
- For –p-threads, the number does not change the results, just changes how fast it runs
- For –p-max-nsti, Avril also uses 2

3. Show Avril our data!

- **Beta diversity graphs:** add stat ellipse to make graphs clearer
- **DESEq:** add in what direction the Log2 fold change is in
    - Remove NAs
    - Ensure when presenting results, ensure it's clear what was used to run DESeq vs. how the graph is plotted

General feedback/suggestions:

- worth doing a core microbiome analysis for groups with large differences from alpha diversity
- **For alpha diversity plots:** add data points to the box plots if possible → contextualizes the box plots
- **Controlling for birth mode:** should be controlled for for everything
    - Can't directly control in this for
    - Run the stats separately, then add to the plot, can annotate plots manually or can have Avril give us the code

    - **Permanovas:** are fine, + birth mode to control for birth mode (don't need to convert to linear model)
    - For stuff like alpha diversity, will need to convert data to linear model to control for birthmode before input
    - **Example of linear model:** controlling for birthmode
    - summary(glm(alpha ~ feed + birthmode, data = dataset))$coefficents
    - this is all the output we need, will give a p value for feed, birthmode, etc.

- Should create a plot with 4 box plots for Shannon's and Faith's plots (High BMI formula, High BMI breastmilk, Low BMI formula, Low BMI breastmilk)
- **Adding new column to phyloseq in R:** sample_data(phyloseq)$new_column = […]

```
sample_data(phyloseq)$new_column =
paste(sample_data(phyloseq)$feed,sample_data(phyloseq)$bmi,sep=' - ')
```

Goals for next week: finish DESeq2, PiCRUST, core microbiome analysis, cleaning up graphs