

## **Importing Data into QIIME2**

1. The sample, recip.460.WT.HC3.D14, had the lowest sequencing depth.
2. The median sequence length is 5101.5.
3. The median quality score at position 125 is 38.
4. For me, the plot stayed the same when I reran it.

## **Sequencing Quality Control and Feature Table**

1. There are 287 features.
2. The sample, recip.539.ASO.PD4.D14, had the highest total count of features. Prior to denoising, the sample had 5475 counts.
3. 23 samples had fewer than 4250 total features.
4. QIIME2 labels “04c8be5a3a6ba2d70446812e99318905”, “ea2b0e4a93c24c6c3661cbe347f93b74, and “1ad289cd8f44e109fd95de0382c5b252”, as the features that contains at least 47 samples.
5. The sample, recip.460.WT.HC3.D49, had the fewest of 347.

## **Alpha Rarefaction and Selecting a Rarefaction Depth**

1. Sample name was excluded because it is assigned barcodes. Days post transplant were also excluded because they are continuous variables.
2. Shannon index shows the saturation and stabilization of the diversity.
3. The wildtype has higher diversity. The susceptible genotype has a shallower sampling depth.
4. 8.33% of the samples are lost if we set the rarefaction depth to 2500 sequences per sample.
5. The missing samples came from mice with the wild type genotype.

## **Diversity analysis**

1. The depth of 2000 was chosen from the rarefaction visualization because it let us keep 47 out of 48 high quality samples.

## **Alpha diversity**

1. There is no difference in evenness or phylogenetic diversity between genotypes.
2. Based on group significant tests, there is not a difference in phylogenetic diversity between genotypes, but there is a difference between donors.

## **Beta diversity**

1. Donor reflects the separation of data in the unweighted and weighted emperor plot.
2. I do not see clustering by cage.
3. Yes, there is a significant effect of donor.

4. There is not a significant difference in microbial community between C31 and C35, but there is a significant difference between C31 and C43. Yes, I would expect mice from the different donors to differ based on the boxplots.
5. There is not a significant difference in variance for any of the cages.
6. Yes, you still retain an effect of genotype. It explains 4.1% of variation.

### **Taxonomic classification**

1. Taxonomy classification is k\_\_Bacteria; p\_\_Firmicutes; c\_\_Clostridia; o\_\_Clostridiales; f\_\_Christensenellaceae; g\_\_ with a confidence of 0.9836.
2. There are only two features classified as g\_\_Akkermansia.
3. Yes, I got the same taxonomic identifier when I BLAST it.

### **Taxonomy bar chart**

1. Yes, there is a consistent difference of phylum between donors. This does not surprise me because the effect of donors was significant.

### **Differential abundance with ANCOM-BC**

1. There are more differentially abundant features between the donors than genotype. This was expected based on the beta diversity.
2. No, there were no features differentially abundant in both genotype and donor.
3. In the combined formula, there are more differentially abundant features than genotype, but less than in the donor barplot.

### **Taxonomic classification again**

1. No enriched ASVs had differing taxonomic resolutions.
2. The taxonomies were the same.
3. No, this is not what we expected. It was expected to be more specific.

### **Longitudinal analysis**

1. It seems that the trajectory is making a “V” pattern, as if it escalates and then regulates.
2. We can visualize the change, but it is not as clear without the animation provided. It’s difficult to see a difference based on day.
3. PC1 shows a clear distinction between donors. In PC2, it appears half of the cages increase, whereas the other half decreases. In PC3, there is a consistent pattern among all lines.

### **Distance-based analysis**

1. There is no donor that changes more over time. The same pattern is shown for genotype and cage.
2. Susceptible genotype has lower variation.
3. Temporal change is associated with donor. This was also shown in the volatility plot.
4. There is an interaction between donor and genotype based on the model P value of 0.003.

#### **Machine-learning classifiers for predicting sample characteristics**

1. Wild type + donors have specific features of 7ce470a3f833253f6d667fa6830abe07, 7121132c894a196a69c89b56593e2f42, 1c94105978b1aa2095e3c0096774f240, 4c2189b6b5ced3cb80093414e9449de9, c18afe570abfe82d2f746ecc6e291bab, 8677fc01f5578165f7c059b12414fdd4, 64d0a182b41ddc2ed31ebfc1dd2dab3b, and a6659d7332df023910005e4404ed1537.