Parkinson's Mouse Tutorial

Importing data into QIIME 2

1. Sample recip.460.WT.HC3.D14 has the lowest sequencing depth
2. Median sequence length is 5101.5
3. Median quality score at position 125 is 38
4. I ran the command 4 times and ended up with the same results and graph each time. I am assuming it could be different at some point and be caused by some slight variations in the metadata or randomization.

Sequence quality control and feature table

1. After denoising, 287 features are left
2. Sample recip.539.ASO.PD4.D14 had the highest total count of features. It had 5475 sequences prior to DADA2 denoising.
3. 23 samples had fewer than 4250 total features.
4. Features 04c8be5a3a6ba2d70446812e99318905, ea2b0e4a93c24c6c3661cbe347f93b74, and 1ad289cd8f44e109fd95de0382c5b252 are observed in at least 47 samples.
5. Sample recip.460.WT.HC3.D49 had the fewest features of 347.

Alpha Rarefaction and Selecting a Rarefaction Depth

1. Not all metadata columns are represented in the visualization. Days_post_transplant is excluded because it did not contain categorical data and consisted of only one value.
2. The observed features metric shows saturation and stabilization of the diversity.

3. The wildtype has a higher diversity based on the curve, while the susceptible has a shallower sampling depth.
4. If we set the rarefaction depth to 2500 sequences per sample, 8.33% of samples are lost
5. The missing samples came from wild type mice.

Diversity analysis

1. We picked 2000 as the sampling depth to remove the samples with the lowest amount of features, which in this case was only one. This allows an adequate amount of samples to still be used in the diversity analysis.

Alpha Diversity

1. There does seem to be a difference in evenness between genotypes wild and susceptible, as well as a difference in phylogenetic diversity.
2. Based on the group significance test, there is a difference in phylogenetic diversity by genotype, as well as donor.

Beta Diversity

1. Yes, I can find a separation between the data across the y-axis. This could be caused by the genotype factor in the metadata. The weighted graph also has some slight separation across the x-axis.
2. There does seem to be some clustering by cage.
3. There is a significant effect based on the donor, with each donor type grouping together.

4. C31 and C35 seem to be grouping together, so there does not seem to be a significant difference in the microbial communities. C31 and C43 however do have a significant difference with each clustering to completely different groups. The results do look the way they expect given the boxplots for the donor.
5. There is a significant variance for the cages with C31, C35 and C42 grouping together, while C43, C44 and C49 group together.
6. The effect of genotype will still remain significant. 4.15% of variation explains the genotype.

Taxonomic classification

1. The sequence 07f183edd4e4d8aef1dcb2ab24dd7745 is a bacteria part of the Christensenellaceae family, with a confidence of 0.9836881157645692, however it is slightly unclassified as it could not confidently classify at a genus level.
2. Two features are classified as g__Akkermansia.
3. I did end up getting the same taxonomic classification

Taxonomy barchart

1. I can see a slight consistent difference between the donors. This does not surprise me as I figured that there would be an association between donor and microbial community.

Differential abundance with ANCOM-BC

1. There seems to be a lot more differentially abundant features between donors, this makes sense given that in the beta diversity analysis we saw grouping based on donor more than genotype.
2. There is one feature that is differentially abundant in donor and genotype, 6a1747edbe461433bdd084f731e.
3. The combined formula has more differentially depleted features, while genotype has more enriched, and donor has an equal amount between the two. There are more differentially abundant features among the individual plots.

Taxonomic classification again

1. Yes, the enriched ASVs have more specific taxonomic resolutions in the dada2_rep_set_multi_taxonomy.qzv
2. The dada2_rep_set_multi_taxonomy.qzv provided better resolution
3. This is what we would expect as we are training the classifier on what the sample should look like before attempting classification, so it has a reference to base it off of.

PCoA-based analyses

1. Yes, I do observe a temporal trend based on the PCoA, they are all displaying a specific directionality per cluster.
2. Yes technically we could visualize change over time without animation using different types of plots like line plots or heatmaps. If we color the plot by day_post_transplant we

would see the change in microbial composition over time, seeing the difference between each day.

3. Along axis 1, three cage groups seemed to be clustered together along each side of the y-axis in a more linear fashion. Axis 2 shows a more negative regression line with no particular clustering, while axis 3 also shows a linear relationship with no clustering.

Distance-based analysis

1. The hc_1 donor seems to be changing more over time than pd_1. Susceptible genotype seems to change a bit more over time compared to wild type, where it slightly plateaus after day 21. Cage C49 seems to have the most drastic change among all the other cage groups, with a sharp decrease on day 20.
2. There does not seem to be a significant association between genotype and temporal change.
3. Susceptible genotype is more stable.
4. There is a temporal change associated with the donor, as expected based on the volatility plot.
5. There is an interaction between the donor and genotype, as the p-value 0.003 indicates that the interaction between the genotype and donor is unlikely to be due to random chance.

1.