QIIME2:
Go through the Parkinson's Mouse model data set and answer the questions in the tutorial. Turn in a PDF (still on github) that has the answers to the questions in the green boxes in the tutorial.
Label each section of questions with the sub header from the tutorial (e.g. "Importing data into QIIME 2", "Sequence quality control and feature table, etc

Importing data into QIIME 2
1.After demultiplexing, which sample has the lowest sequencing depth?
4237

2.What is the median sequence length?
5101.5

3.What is the median quality score at position 125?
38

4.If you are working on this tutorial alongside someone else, why does your plot look slightly different from your neighbors? If you aren't working alongside someone else, try running this command a few times and compare the results.
Random sampling is used for the large data sets that is the reason for the slightly different plot.

Sequence quality control and feature table

1.How many total features remain after denoising?
287

2.Which sample has the highest total count of features? How many sequences did that sample have prior to DADA2 denoising?
04c8be5a3a6ba2d70446812e99318905 25,050 47
Maximum frequency 4,996.0
3.How many samples have fewer than 4250 total features?
0
4.Which features are observed in at least 47 samples?

04c8be5a3a6ba2d70446812e99318905	25,050	47
ea2b0e4a93c24c6c3661cbe347f93b74	23,895	47

1ad289cd8f44e109fd95de0382c5b252

5.Which sample has the fewest features? How many does it have?

recip.460.WT.HC3.D49	347

Generating a phylogenetic tree for diversity analysis
Alpha Rarefaction and Selecting a Rarefaction Depth

Alpha Rarefaction and Selecting a Rarefaction DepthStart by opening the alpha rarefaction visualization.

1.Are all metadata columns represented in the visualization? If not, which columns were excluded and why?
The following metadata columns have been omitted because they didn't contain categorical data, or the column consisted only of missing values: days_post_transplant

2.Which metric shows saturation and stabilization of the diversity?
Shannon diversity shows saturation and stabilization of the diversity

3.Which mouse genetic background has higher diversity, based on the curve? Which has shallower sampling depth?
457 mouse id has higher diversity

Now, let’s check the feature table summary.

4.What percentage of samples are lost if we set the rarefaction depth to 2500 sequences per sample?
43.89
Which mice did the missing samples come from?
457

Diversity analysis

Where did we get the value 2000 from? Why did we pick that?

Alpha diversity

1.Is there a difference in evenness between genotype? Is there a difference in phylogenetic diversity between genotype?
Yes there is difference between difference in phylogenetic diversity between genotype

2.Based on the group significance test, is there a difference in phylogenetic diversity by genotype? Is there a difference based on the donor?
Yes ,there is difference between difference in phylogenetic diversity between genotype

Beta diversity

1.Open the unweighted UniFrac emperor plot (core-metrics-results/unweighted_unifrac_emperor.qzv) first. Can you find separation in the data?
Yes i can find the separation in the data
2. If so, can you find a metadata factor that reflects the separation? What if you used weighted UniFrac distance (core-metrics-results/weighted_unifrac_emperor.qzv)?
Donor

3.One of the major concerns in mouse studies is that sometimes differences in communities are due to natural variation in cages. Do you see clustering by cage?
There is clustering by cage but is not significant

4.Is there a significant effect on donors?
Yes there is significant effect on donors because of unweighted (p=0.001) and weighted (p=0.001)

From the metadata, we know that cage C31, C35, and C42 all house mice transplanted from one donor, and that cages C43, C44, and C49 are from the other. Is there a significant difference in the microbial communities between samples collected in cage C31 and C35? How about between C31 and C43? Do the results look the way you expect, based on the boxplots for donor?
When looking at the pairwise comparisons between individual cages, there are significant differences between some cages (e.g. C31 vs C35, p=0.001), but not others (e.g. C31 vs C43, p=0.064). This matches the expectation based on the donor groupings.

Is there a significant difference in variance for any of the cages?
no significant differences in dispersion between the cages (p=0.193)

If you adjust for donor in the adonis model, do you retain an effect of genotype? What percentage of the variation does genotype explain?
genotype still has a significant effect (p=0.001). Genotype explains about 8.5% of the variation in the unweighted UniFrac distances.
Taxonomic classification

1.Find the feature, 07f183edd4e4d8aef1dcb2ab24dd7745. What is the taxonomic classification of this sequence? What's the confidence for the assignment?
k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Christensenellaceae; g__; s__
0.9836881157645692

2.How many features are classified as g__Akkermansia?
2

3.Use the tabulated representative sequences to look up these features. If you blast them against NCBI, do you get the same taxonomic identifier as you obtained with q2-feature-classifier?
Yes i have got same results with blast also

Taxonomy barchart

1.Visualize the data at level 2 (phylum level) and sort the samples by donor, then by genotype. Can you observe a consistent difference in phylum between the donors? Does this surprise you? Why or why not?
k__Bacteria;p__Firmicutes had a significant difference. It does not surprise me.

Differential abundance with ANCOM-BC

1.Are there more differentially abundant features between the donors or the mouse genotype? Did you expect this result based on the beta diversity?
There are more features in donors

2.Are there any features that are differentially abundant in both the donors and by genotype?
Yes there is 3017f87a3b0f5200ed54eca17eef3cbb is for genotype.

3.How do the bar plots for the combined formula ('donor + genotype') compare with the individual donor and mouse genotype bar plots? Are there more differentially abundant features in the individual plots or the combined?
There are significant differences in individual and combined formula as there are only 7 features in the combined formula

Taxonomic classification again

1.Examine the enriched ASVs in the da_barplot_donor.qzv visualization. Are there any of these enriched ASVs that have differing taxonomic resolution in the dada2_rep_set_multi_taxonomy.qzv visualization?
There are a total 12 enriched ASVs in the da_barplot_donor.qzv visualization.

2.If so, which taxonomy provided better resolution?

Taxonomy.qza

3.Is this what we expect, based on what we learned about taxonomic classification, accuracy, and re-training earlier in the tutorial?
No, bespoke_taxonomy.qza should be more accurate

Longitudinal analysis

PCoA-based analyses

1.Open the unweighted UniFrac emperor plot and color the samples by mouse id. Click on the “animations” tab and animate using the day_post_transplant as your gradient and mouse_id as your trajectory. Do you observe any clear temporal trends based on the PCoA?
I observed clear trend of negative regression after certain days of post transplant

2.Can we visualize change over time without an animation? What happens if you color the plot by day_post_transplant? Do you see a difference based on the day? Hint: Try changing the colormap to a sequential colormap like viridis
Yes we can visualize change over time without an animation,

3.Using the controls, look at variation in the cage along PCs 1, 2, and 3. What kind of patterns do you see with time along each axis?
PCs 1 apparies to be overlapping lines , PCs 2 regression lines first shows positive progress then goes in negative direction, PCs 3 most of them is straight lines .

Distance-based analysis

1.Based on the volatility plot, does one donor change more over time than the other? What about by genotype? Cage?
Yes hc_1 donor changes more over time compared to pd_1 donor ,when it comes to genotype wild type shows positive regression and susceptible shows negative regression, cage c31 shows significant change.

2.Is there a significant association between the genotype and temporal change?
Yes there is significant association between the genotype and temporal change

3.Which genotype is more stable (has lower variation)?
Wild type is more stable

4.Is there a temporal change associated with the donor? Did you expect or not expect this based on the volatility plot results?
There is a significant temporal change associated with the donor pd_1, particularly noticeable when considering the interaction with days post-transplant.

5.Can you find an interaction between the donor and genotype?
A significant interaction exists between genotype and donor, especially notable in changes over time, suggesting complex dynamics dependent on both genetic and donor microbiota factors.

Machine-learning classifiers for predicting sample characteristics

1.How did we do? Just for fun, try predicting some of the other metadata columns to see how easily cage_id and other columns can be predicted**.**
Model accuracy is 0.9

2.What features appear to differentiate genotypes? What about donors? Are any ASVs specific to a single sample group?
Yes there are 4 ASVs specific to a wild type and healthy as well as susceptible and healthy has 3 ASVs