Olivia Baldwin

Qiime2 – Mouse Tutorial Answers

Importing Data Into Qiime
1) The lowest sequencing depth is sample recip.460.WT.HC3.D14
2) The median in 5101.1
3) The median is 38
4) The plot was generated by a random sample of the reads, so their random sampling may have been different than mine, causing slightly different plots.

Sequence quality control and feature table
1) 287 total features
2) recip.460.WT.HC3.D49 had the highest number and after denoising, had the lowest (347)
3) 23 samples have fewer than 4250 features
4) 04c8be5a3a6ba2d70446812e99318905, ea2b0e4a93c24c6c3661cbe347f93b74, and 1ad289cd8f44e109fd95de0382c5b252 are all in at least 47 samples
5) recip.460.WT.HC3.D49 has the fewest number, 347

Alpha Rarefaction and Selecting Rarefaction Depth
1) days_post_translation were omitted for not having categorical data
2) The Shannon metric shows saturation/stability of diversity.
3) Mouse 435 has highest diversity and mouse 469 has lowest sequencing depth.
4) Lost 4 out of 48 samples (9%).
5) These samples came from mice 457, 469, 537, and 538.

Diversity Analysis
1) The 2000 is the sequencing depth. It was selected because it allows us to discard only one sample that had a depth far lower than the others (below 1000).

Alpha Diversity
1) There is no significant difference in evenness or phylogenetic diversity between the genotypes.
2) Based on the group significance test, there is not a difference in phylogenetic diversity between the genotypes, but there is a difference based on donor.

Beta Diversity
1) The unweighted unifrac has separation into two groups. This seems to be split by donor. The weighted unifrac does not have a space between the groups, but by donor there is still separation.
2) In both graphs, cages 31, 35, and 42 seem to be on one side and 43, 44, and 49 are on the other. However, this clustering matches that the first three cages have a different donor than the second three.

1) Based on the graphs and p-value, there is a significant effect on donor.

2) In the weighted data, there is not a significant difference between 31 and 35, but there is between 31 and 43. This is not surprising because 31 and 35 have the same donor, but 31 and 43 do not, and we see that the variable donor shows significant differences.

1) There are significant differences between 49, 43, and 44 and the other three cages.

1) Genotype goes down to about 4%.

Taxonomic Classification
1) The classification is k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__Christensenellaceae; g__; s__ with 98% certainty.
2) Two features are classified as g__Akkermansia.
3) When blasting the first sequence, I received some different taxonomic information.

Taxonomy Bar chart
1) One donor only has 3 phyla, while the other has up to 7 in the samples. This is not surprising, as we saw significant differences between donor samples in our diversity metrics.

Differential Abundance with ANCOM
1) There are more differentially abundant features between donors, which is expected.
2) No there are not any that overlap between the genotype and donor graphs.
3) The combined plot has far fewer than the solo donor plot, but more than the solo genotype plot.

Taxonomic Classification Again
1) Yes, features 04195686f2b70585790ec75320de0d6f and 54f7ee881a58ad84fe3f81d76968b072
2) The taxonomy went down to the species level and recognized E. coli and A. masseliensis,
3) Yes, we would expect the taxonomy to either stay or get more accurate.

Longitudinal Analysis
PCoA Analysis
1) After setting the animation, many of the lines were moving downward, towards axis one.
2) I see the same pattern of having the later days generally lower on the graph.
3) Axis 1 seems flat (slope of close to zero). Axis 2 and axis 3, by around day 20 cage 43, 44, and 49 are sloping up, and the other three (with a different donor) are sloping down.

Distance Based Analysis
1) For donors, it looks like they change at a similar rate, but in opposite directions. This pattern is similar in genotype. For cages, cage C42 showed the most change, with a drastic drop in the beginning, followed by a slow rise.
2) There does not seem to be an association between genotype and temporal change.

3) The susceptible seems to have lower variation.
4) Yes and yes I did expect this based on volatility.
5) Yes, there was a significant interaction between donor pd and genotype wildtype shown.

Machine-learning classifiers for predicting sample characteristics
1) Many features on the far left of the table only seem to show color for the healthy samples, while many on the far right only show for the pd samples. The features beginning with 092 and 301 show heat only for wild type and not susceptible, and none are completely absent from the wild type, but present in both susceptible samples. There are 17 features that only show color for one section, meaning they only apply to that sample group.