1. Include a screenshot of your interactive quality plot. Based on this plot, what values would you choose for --p-trunc-len and --p-trim-left for both the forward and reverse reads? Why have you chosen those numbers?

For the reverse reads, I would choose 0 for trim-left and 160 for trunc-len because the values in the reverse reads on the 3’ end are pretty good until they start to drop around 160, which would be the 5’ end. For forward reads I would choose 0 for trim-left and 190 for trunc-len for the same reason as the reverse reads.

A graph of a graph of a number of data

Description automatically generated with medium confidence

1. How would you modify the code above to truncate and trim in your desired way?

  --p-trim-left-f 0 \  
  --p-trunc-len-f 190 \  
  --p-trim-left-r 0 \  
  --p-trunc-len-r 160 \

Including the values where I want to trim at 0 for both forward and reverse and truncate at 190 for forward and 160 for reverse.

1. In the tutorial, you had to mv the files to rename them to just rep-seqs.qza, table.qza, and stats.qza. How could you modify the above code to skip that step? How do you need to modify qiime metadata tabulate in order to account for the renamed files being generated?

I would modify the code by removing “dada2” from the code. For example, instead of “--o-table table-dada2.qza \” I would instead type “--o-table table.qza \”

For the qiime metadata tabulate, I would remove the “-dada2” from the code, for example:

qiime metadata tabulate \

--m-input-file stats.qza \

--o-visualization stats.qzv

4) Your metadata file has a different name than that in the tutorial. How do you adjust your code in order to use the metadata file you have been given? I will remove “sample-metadata.tsv” and change it to “metadata.txt”, for example;

“--m-sample-metadata-file metadata.txt”. My metadata file was only working under a regular .txt file.

5) Include a screenshot of the table summary from visualizing your table and a screenshot of the sequence length statistics from the rep-seqs file.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a graph

Description automatically generated

6) Jump down to taxonomy. Once you have generated your taxonomy visualization, sort it by confidence. What are your top hits? A screenshot of a computer

Description automatically generated

7) What do you think this code is doing? Why do you think this is a necessary or important step? It will create a table based on the abundance of different microbial characteristics. It’s excluding mitochondrial and chloroplast DNA that are not of interest in microbiome analysis. This is important because it will filter out everything we tell it too but still leaving the annotations for all other taxonomic groups.

8) Re-do your table visualization and re-do your taxonomy commands. Do you have any differences now in the hits with the highest confidence? Why or why not? Really think about what the code is doing.

The highest confidence is now showing only bacteria, with 0.999 instead of the mitochondria. This is because the code filters the table to exclude mitochondria and chloroplast to the very end of the table and leaving the highest confidence that it’s not mitochondria.

9) Looking at taxa bar plots, what are your top 2 phyla? Include a screenshot. What are the top 5 most abundant classes? Include a screenshot. A screenshot of a graph

Description automatically generatedA screenshot of a computer

Description automatically generated

10) What is the difference between alpha and beta diversity? You will have to read outside resources to answer this question. Your response should be in your own words.

Alpha diversity will help identify the different types of species and the number of individuals of each species within one community to see if all species are even throughout while beta diversity will measure the difference between two more communities to compare them and identify any patterns of diversity. This can be used for a macro and micro analysis.

11) Before you calculate your diversity metrics, you have to choose a sampling depth. What file previously generated will you use to help you determine what to choose? Defend your choice of sampling depth. How many samples do you retain and how many do you lose?

I will be using table.qzv to help me determine the sampling depth, this can be found under “frequency”. The sampling depth I chose is 8545 because it’s closest to the highest values and bigger than the smallest values underneath it, the next value would be a big jump to a smaller range. I will be retaining 29 samples and losing 5 of them.

12) For alpha diversity, you need to create visualizations for Shannon diversity and Observed features. This will require you to modify the alpha-group-significance code. For which metadata values were graphs generated? Were any of those comparisons significant? How do you know whether they were or were not significant? Briefly describe what Shannon diversity and Observed features are measuring (less than 1 paragraph).

The metadata values used to generate the graphs were flock, sex and population. None of the comparisons were significant because all of the p-values were above 0.05. The Shannon diversity is measuring the diversity of species in the sex, population and flock of the community by noting the abundance and evenness of the species present, calculating the species distribution. Observed features will count the number of different features in the community but without taking into account the abundance or distribution.

13) For beta diversity, you will need to create visualizations for Bray Curtis dissimilarity. This will require your to modify the beta-group-significance code. You should have one visualization for sex, one for population, and one for flock. Include a screenshot of each visualization. Is there any significance? Regardless of significance, how can you interpret these results (hint: what is beta diversity looking at?) Beta Diversity is looking at the species diversity and evenness in two different communities. For the population comparison, there was a significance with a p-value of 0.035 but the visualization for sex and flock did not have any significance. In this case we are comparing the sex, population and flock for resident and migratory birds.

A screenshot of a computer

Description automatically generatedSex

A screenshot of a computer

Description automatically generatedPopulation

A screenshot of a computer

Description automatically generatedFlock

14) The core-metrics-phylogeny command generates a file called bray-curtis-emporer.qzv. Include 3 screenshots total (1 where the points are colored based on sex, one on population, one on flock). How do these results help you make sense of the results you got from question 13?

The emperor graphs help you analyze the significance graphs by analyzing how different or similar the different characteristics given to us are. Overall, it gives a better visualization of each population.

A screen shot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated