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?

A screenshot of a computer screen

Description automatically generated

* Based on the plot I would choose --p-trim-left-f 15, --p-trunc-len-f 192, --p-trim-left-r 29, --p-trunc-len-r 190. These numbers were chosen because the quality of the graph began to degrade on the right of the grahp for the forward read at 192 and for the reverse read at 190. The left of the graph had the best quality at the start of 15 for the forward read and 29 for the reverse read. Based on these obersavtions was how I determined the graph trimming numbers.

For questions 2 and 3: Because these are paired-end reads, you will have to modify the dada2 code in order to perform the quality trimming on both the forward and reverse reads. **You will not do the deblur.** You will need to adjust this code to account for --p-trunc-len and --p-trim-left for both the forward and reverse reads. The basics of the code you need to change are here.

qiime dada2 denoise-paired \

--i-demultiplexed-seqs demux.qza \

--p-trim-left-f 15 \

--p-trunc-len-f 192\

--p-trim-left-r 29 \

--p-trunc-len-r 190 \

--o-representative-sequences rep-seqs-dada2.qza \

--o-table table-dada2.qza \

--o-denoising-stats stats-dada2.qza

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

The above code would be modified as follows in blue

qiime dada2 denoise-paired \

--i-demultiplexed-seqs demux.qza \

--p-trim-left-f 15 \

--p-trunc-len-f 192\

--p-trim-left-r 29 \

--p-trunc-len-r 190 \

--o-representative-sequences rep-seqs-dada2.qza \

--o-table table-dada2.qza \

--o-denoising-stats stats-dada2.qza

* This will give me the desired data table, feature table and sample data with my adjusted code.

3) 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?

Modified Code to:

qiime dada2 denoise-paired \

--i-demultiplexed-seqs demux.qza \

--p-trim-left-f 15 \

--p-trunc-len-f 192\

--p-trim-left-r 29 \

--p-trunc-len-r 190 \

--o-representative-sequences rep-seqs.qza \

--o-table table.qza \

--o-denoising-stats stats.qza

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?

qiime feature-table summarize \

  --i-table table.qza \

  --o-visualization table.qzv \

  --m-sample-metadata-file metadata.txt

qiime feature-table tabulate-seqs \

  --i-data rep-seqs.qza \

  --o-visualization rep-seqs.qzv

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 graph and chart of a graph

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

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

* The taxonomy visualization when sorted by confidence outputs yields hits that included k\_\_Bacteria; p\_\_Proteobacteria; c\_\_Alphaproteobacteria; o\_\_Rickettsiales; f\_\_mitochondria with a confidence yield of 0.9999999999999984.

For question 7: Run this code

qiime taxa filter-table \

--i-table table.qza \

--i-taxonomy taxonomy.qza \

--p-exclude mitochondria,chloroplast \

--o-filtered-table table.qza

7) What do you think this code is doing? Why do you think this is a necessary or important step?

qiime taxa filter-table \

--i-table table-no-mitochondria.qza \

--i-taxonomy taxonomy.qza \

--p-exclude mitochondria,chloroplast \

--o-filtered-table table.qza

* The code that is entered is trying to filter the output of samples that include mitochondria and chloroplast.

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 confidence and the hits do not change because the rep-seqs that we use to generate the taxnomy are not filtered so the taxonomy visualization and confidence would be the same.

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.

* The top two phlya is k\_\_Bacteria;p\_\_Proteobacteria and k\_\_Bacteria;p\_\_Actinobacteria.
* Top 5 classes are:
  + k\_\_Bacteria;p\_\_Proteobacteria;c\_\_Alphaproteobacteria
  + k\_\_Bacteria;p\_\_Actinobacteria;c\_\_Actinobacteria
  + k\_\_Bacteria;p\_\_Cyanobacteria;c\_\_Chloroplast
  + k\_\_Bacteria;p\_\_Firmicutes;c\_\_Bacilli
  + k\_\_Bacteria;\_\_;\_\_

A 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 is the measure of diversity of the species in one ecosystem or habitat taking measurements of species richness and evenness in a particular part of the community.
* While Beta diversity, measures variation in species composition of various ecosystems or habitat. Beta diversity is also sometime used to measure similarity and dissimilarity of two different sites.

11) Before you calculate your diversity metrics, you have to choose a sampling depth (use TABLE. 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?

* The previous file that I have generated was the table.qzv. This had a sample depth of 1070 samples. So in the code the sample depth would be changed to 1070. I choice 1070 as my sample depth because it will only excluded three samples. I had a total sample size of 24 but after excluding 3 samples I would have 21 remaining samples.
* The edited code would be *–p sampling-depth 1070*

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 values that were created from the shannon-group-significance were sex, flock and population. Of the three measured values there were not any significance. I know this because each one yielded a q-value that was greater than 0.05.
* The Shannon diversity and observed features is measuring the relative abundance of a species in its habitat and the overall richness of the species that habitat represents. The observed features is how many different species are in an environment.

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?)

* There is no significance is the Bray Curtis dissimilarity the q-value for each of the visualizations of sex, population and flock. The beta diversity, which is also called species turnover, explains the change in species as you move across a habitat, communities or ecosystem.

**Flock:**

A screenshot of a computer

Description automatically generated

**Population:**

A screenshot of a computer

Description automatically generated

**sex:**

A screenshot of a computer

Description automatically generated

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 plots that were generated helps visualize the data set that was generated in question 13. Question 13 includes data on a theoretical image (what it should look like) and the plots help understand and see what is being generated in question 13 with an overall visualization of the data set.
* Question 13 gives you the data set and statistics and question 14 takes the dta and statistics and draws a map.

**Sex:**

A screenshot of a computer

Description automatically generated

**Population:**

A screenshot of a video game

Description automatically generated

**Flock:**

A screenshot of a video game

Description automatically generated