

APPENDIX D

QIIME2 Simplified Workflow: 16S rRNA Sequence Processing

Before you start:

1.) Check and record the QIIME program and version: qiime info

- Copy-paste all information from the terminal after this command, report this in your methods and bioinformatics notebook

2.) Activate the conda environment if your version of QIIME requires this (i.e., after a recent update)

- In my case: conda activate qiime2-2021.11

3.) Set your working directory (the file you will be working from)

- cd <drag and drop your directory file here>

4.) Make sure that your manifest file is in the correct format, location, and that the file paths are all correct

- Importing the data via the manifest file was the fussiest step I encountered, and I think it had to do with using a PC and a Virtual Box to use QIIME, so proceed with caution here

5.) Include your project background at the beginning of your bioinformatics notebook

- Brief description, objectives/hypotheses

Part I – Sequence data import and preparations (importing data, filtering and trimming sequences, aligning sequences, importing mapping file for later analyses)

1.) Import your sequence data into QIIME

- This is often the most difficult step that requires the most troubleshooting
- Gives you **file-name_single-end_demux.qza**

2.) Visualize the imported sequence data in QIIME2 View: <https://view.qiime2.org/>

- Use the **file-name_single-end_demux.qza** from the import step (#1) in QIIME2 View, convert it to a qzv file (**file-name_single-end_demux.qzv**) and use that in QIIME2 View
- Gives you an idea of how high quality your samples are
- Record the following information from your interactive quality plot and the tables in the overview tab:
 - Total number of sequences
 - Average number of reads per sequence
 - Sample with the highest number of reads
 - Sample with the lowest number of reads
 - The sample with the first big jump in sequence read number
- Determine where to trim your sequences for future steps:
 - By using figaro (command line installation and code if you can get it to work):
 - You will need various information on your primers for these commands
 - Or visually via the interactive quality plot: where does the sequence quality begin to drop off? Somewhat subjective, so give it careful thought.

3.) Filter your sequence data

- Using deblur filter by quality score, takes about 40 minutes to run
- Gives you **file_name_demux-filtered.qza** and **file_name_demux-filter-stats.qza**

4.) Visualize your filtered sequences in QIIME2 View

- Convert the **file-name_demux-filtered.qza** to **file-name_demux-filtered.qzv**, use qzv file in QIIME2 View
- Record the same information that you did for your raw unfiltered sequences (#2)
- Compare the filtered and unfiltered stats

- How many sequences were lost in the filtering step? What is this value as a percentage?
- If a huge number of sequences were lost, adjust the parameters to find the balance between good quality sequences and having enough sequences for later analyses. I just used the default parameters.
- If the filtered data looks good, use the demux-filtered.qza for future steps from now on

5.) Visualize the filtered stats in QIIME2 View

- Convert **file-name_demux-filter-stats.qzv** to **file-name_demux-filter-stats.qzv**, use qzv in QIIME2 View
- Not a make-or-break step, use this jointly with your visual interpretation of the filtered sequences quality plot (#4)
 - Allows you know how the filtering step worked, but I found it more useful to just look at the quality plot in #4

6.) Run deblur (for single-end sequence data only)

- Use **file-name_demux-filtered.qza** from deblur filter by quality score step (#3)
- This took roughly 24 hours to complete, because your computer is aligning millions of sequences
- Deblur groups and aligns based on sequence similarities and an algorithm, determines if a sequence is a real biological sequence or noise (based on how present that sequence is across the whole dataset)
 - Useful deblur info:
https://awbrooks19.github.io/vmi_microbiome_bootcamp/rst/3_sequences_to_composition.html
- Deblur will align your sequences and trim them (if you specified that option)
- Gives you **file-name_rep-seqs-deblur.qza**, **file-name_deblur-stats.qza**, and **file-name_table-deblur.qza**

7.) Visualize the deblur stats from the deblur step – what did deblur do to my sequence data?

- Use the `file-name_deblur-stats.qza` from the deblur step (#6) and convert it to `file-name_deblur-stats.qzv` using the provided code
- Visualize the `file-name_deblur-stats.qzv` table in QIIME2 View
 - Look at the unique reads column – that acts as a pre-indicator of how many unique bacterial species you will most likely have

8.) Visualize the representative sequences from the deblur step

- Checks to see if deblur did what it was supposed to do (i.e., trimming)
- Use the `file-name_rep-seqs-deblur.qza` from deblur (#6) and convert it to `file-name_rep-seqs-deblur.qzv`
- Visualize `file-name_rep-seqs-deblur.qzv` in QIIME2 View
 - Look at sequence length to see if everything was trimmed correctly
 - Check a sequence to make sure the primer sequences were cut out

9.) Visualize and summarize the table from the deblur step

- Allows you to match your sample IDs and other information to the bacterial sequences for that sample. You will run your data analyses on this in later steps
- Use the `file-name_table-deblur.qza` from deblur (#6) and your `metadata-file.tsv`, the code will convert these to `file-name_table-deblur.qzv`
 - Your metadata file (mapping file) must be in a tab separated values (.tsv) file format, otherwise it won't work. I used Google Sheets to export my mapping file as a .tsv
- Use the `file-name_table-deblur.qzv` in QIIME2 View, record the following:
 - How many samples are in the dataset?
 - How many features (bacterial “species”) are in the dataset?
 - What is the total frequency (total number of DNA sequences in the dataset)?

- What is the frequency per sample?
- What is the mean frequency per feature (the mean number of sequences assigned to a feature, a.k.a., bacterial “species”)?

Part II – Analyses on your sequence data

a.) Preparing your sequences and assigning taxonomy to your sequence data with Silva

1.) Training feature classifier:

- This step creates a new pared-down reference database that QIIME will compare your sequences to.
 - The Silva database has the full 16S rRNA gene (with a lot of sequences and assigned identities), but we only need to look at the V4-V5 region, which is the region specified by our primers (515f, 926r). The database is pared down in the classifier step, so QIIME isn’t trying to search the entire database.
- In our case, the training feature classifier step had to be custom-made by Dr. Walke and Shelby Fettig because of the primers we are using (515f, 926r), as QIIME only has the 806r primer pre-installed.
- The series of steps here will give you a qiime classifier file (i.e., the **Shelby Fettig - silva_99_138.1_qiime_classifier.qza** I used) that you will need to assign taxonomy to your sequences (#2).
- **Caution!** → the classifier must be trained using the same version of QIIME that you have installed, otherwise it won’t work (because you are using an older version of the classifier that does not match the latest version of the software).
 - Also, the **qiime feature-classifier fit-classifier-naive-bayes** step generally can’t be run on a Virtual Box because there isn’t enough RAM on it. Dr. Walke ended up running this code on her Mac to get around this. There are ways to upgrade the memory on your Virtual Box, but I didn’t want to mess around with that.

2.) Assign taxonomy to your sequences

- Identifies the bacterial “species” found in your samples

- Uses the **file-name_Silva138_515_926_classifier.qza** from the training feature classifier step (#1) and the **file-name_rep-seqs-deblur.qza** from the deblur step (#6, part I).
- This step gives you a **file-name_taxonomy.qza** to use in later steps.

3.) Visualize the taxonomy of your sequences:

- Use **file-name_taxonomy.qza** from taxonomy step (#2), converts it to a **file-name_taxonomy.qzv**, visualize in QIIME2 View
- Will be an interactive table, very useful!
- Use this to determine what you need to filter out in the filtering steps (i.e. chloroplasts, mitochondria, unassigned sequences, etc.) and how to format that in your code.
- Also useful for getting an idea of what types of bacteria are in your sample.

4.) Filter out mitochondria and chloroplasts from your table:

- Removes mitochondria and chloroplasts from your sequence table
- Uses the **file-name_table-deblur.qza** from the deblur step (#6, part I) and the **file-name_taxonomy.qza** from the taxonomy assignment step (#2)
- This step gives you a **file-name_filtered-table.qza** to use in later steps

5.) Check that the mitochondria and chloroplasts were filtered out from your table

- Uses **file-name_filtered-table.qza** the from the table filtering step (#3) and your **mapping/metadata file.tsv** that you used in the deblur visualization (#9, part I) converts it to a **file-name_filtered-table.qzv**
- Upload the **file-name_filtered-table.qzv** into QIIME2 View to visualize your data table, compare it with your unfiltered data table (the **file-name_table-deblur.qzv** from the deblur step):
 - Compare the number of features (ASV or bacterial “species”), make sure that there are fewer in the filtered table

6.) Filter the mitochondria and chloroplasts out of your sequences:

- Removes mitochondria and chloroplasts from your sequences themselves
- Use the **file-name_rep-seqs-deblur.qza** from the deblur step (#6, part I) and the **file-name_taxonomy.qza** from the taxonomy assignment step (#2)
- Gives you **file-name_filtered-rep-seqs.qza** to use in later steps

7.) Remove any contaminants from your sequences

- Examine controls and suspicious taxa, use similar code to #6, see notebook for details

8.) Remove control samples from your dataset

- Use **qiime feature-table filter-samples**, see notebook for details

9.) Visualize the microbiome using taxa bar plots

- Allows you to look at the relative abundance of bacteria on each sample, more useful than the table in #6
- Use the **file-name_filtered-table.qza** from the mitochondria and chloroplast filter table step (#3), the **file-name_taxonomy.qza** from the taxonomy step (#2) and the **mapping/metadata file.tsv** you uploaded earlier
- Gives you **file-name_taxa-bar-plots.qzv**, look at this in QIIME2 View

b.) Generate a multiple sequence alignment (MSA) using MAFFT (Multiple Alignment using Fast Fourier Transform), build phylogenetic tree using fasttree, rarefy sequence data

1.) Make a phylogenetic tree for phylogenetic diversity analyses later on

- Aligns your filtered deblur sequences with MAFFT into a phylogenetic tree, will be used in alpha and beta diversity analyses later on. This code gives you many outputs!
- Use **file-name_filtered-rep-seqs.qza** from the mitochondria/chloroplast sequence filter step (#5,a)
- Gives you: **file-name_aligned-rep-seqs.qza**, **file-name_masked-aligned-rep-seqs.qza**, **file-name_unrooted-tree.qza**, and **file-name_rooted-tree.qza**

2.) Rarefy your filtered data and phylogenetic tree

- Rarefaction is a way to standardize your tests so the number of sequences does not affect the results. This step will determine the sequence depth in later analyses (i.e., in calculating alpha and beta diversity)
- Use the **file-name_filtered-table.qza** from table filter step (#4,a), the **file-name_rooted-tree.qza** from the phylogenetic tree step (#1) and the **mapping/metadata file.tsv** you've been using

- Gives you **file-name_alpha-rarefaction.qzv**, examine in QIIME2 View and determine the sequence depth that you will use in later analyses
- **Caution!** → this step is extremely fussy about the formatting of your mapping file. Remove all weird characters and messy labels (see notebook for more details).

c.) Actual data analyses

1.) Calculate alpha and beta diversity of your samples

- Use **file-name_rooted-tree.qza** from phylogenetic tree step (#), the **file-name_filtered-table.qza** from table filter step (#1,b), and the **mapping/metadata file.tsv** that you've been using.
 - If you edited your mapping file for the rarefaction step, continue to use that edited mapping file.
- Gives you **file-name_rarefied-table.qza**, view this in QIIME2 View
- Also gives you a **file-name_core-metrics-results folder** (fancy!)
 - Output folder contains alpha diversity metrics (by default: Shannon's diversity index, observed OTU's, Faith's phylogenetic diversity, evenness) and beta diversity metrics (Jaccard, Bray-Curtis, unweighted UniFrac distances, weighted UniFrac distances)

2.) Calculate alpha diversity statistics with nonparametric Kruskal-Wallis tests for categorical variables:

- **Shannon's diversity:** number of species present (richness) and abundance of each species
 - Use **file-name_core-metrics-results/shannon_vector.qza** from diversity metrics folder (#1) and the **mapping/metadata file.tsv** that you've been using
 - Gives you **file-name_core-metrics-results/shannon_significance.qzv**, view in QIIME2 View
- **Observed Features:** number of different ASVs (bacterial "species") in a sample, the richness of different bacterial taxa present

- Use `file-name_core-metrics-results/faith_pd_vector.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
- Gives you `file-name_core-metrics-results/observed_ottus_significance.qzv`, view in QIIME2 View
- **Faith's Phylogenetic Diversity:** relatedness of bacterial taxa present
 - Use `file-name_core-metrics-results/faith_pd_vector.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
 - Gives you `file-name_core-metrics-results/faith_pd_significance.qzv`, view in QIIME2 View
- **Evenness:** the abundances of each species present
 - Use `file-name_core-metrics-results/evenness_vector.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
 - Gives you `file-name_core-metrics-results/evenness_significance.qzv`, view in QIIME2 View

3.) Calculate alpha diversity statistics using Spearman's rank correlations for continuous variables:

- Use the alpha diversity vector file (*i.e., Bat_2021_MYVO_only_core-metrics-results/shannon_vector.qza*) from diversity core metrics folder)
 - Repeat for all diversity metrics (**Shannon, Faith's phylogenetic diversity, Evenness, Observed Features**)
- Gives you `Bat_2021_MYVO_only_core-metrics-results/shannon_correlation_Spearman.qzv`, visualize with QIIME2 View
- Output includes all continuous variables (*i.e., weight, wing damage*)

4.) Calculate beta diversity statistics using nonparametric PERMANOVAs for categorical variables:

Bray-Curtis dissimilarity: abundances of bacterial taxa

- Use `file-name_core-metrics-results/bray_curtis_distance_matrix.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
- Specify which column in your metadata file that you want to examine – the “treatment” that you are investigating (i.e., differences in bacterial diversity between site, Bd status, etc.). In my case, ectoparasite presence (the “Ectoparasites” column)
- Gives you `file-name_core-metrics-results/bray_curtis_site_significance.qzv`, view in QIIME2 View as the Principal Coordinates Ordination Plot (PCoA), each point is a sample and the closer the points are to each other, the more similar their microbiomes.

Jaccard similarity: presence or absence of bacterial taxa, comparing microbial composition

- Use `core-metrics-results/jaccard_distance_matrix.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
- Specify which column in your metadata file that you want to examine – in my case, ectoparasite presence (the “Ectoparasites(Y/N)” column)
- Gives you `file-name_core-metrics-results/jaccard_site_significance.qzv`, view in QIIME2 View

Unweighted UniFrac Distances: qualitative measure of bacterial presence/absence

- Use `file-name_core-metrics-results/unweighted_unifrac_distance_matrix.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
- Specify which column in your metadata file that you want to examine – in my case, ectoparasite presence (the “Ectoparasites(Y/N)” column)
- Gives you `file-name_core-metrics-results/unweighted_unifrac_site_significance.qzv`, view in QIIME2 View

Weighted UniFrac Distances: quantitative, more comprehensive, abundance and relatedness of bacterial taxa

- Use `file-name_core-metrics-results/weighted_unifrac_distance_matrix.qza` from diversity metrics folder (#1) and the `mapping/metadata file.tsv` that you've been using
- Specify which column in your metadata file that you want to examine – in my case, ectoparasite presence (the “Ectoparasites(Y/N)” column)
 - Need to repeat for other variables you want to examine

- Gives you **file-name_core-metrics-results/weighted_unifrac_site_significance.qzv**, view in QIIME2 View

5.) Calculate beta diversity statistics using Mantel tests for continuous variables

- Uses beta diversity distance matrix files from core-metrics results folder (i.e., **Bat_2021_core-metrics-results/bray_curtis_distance_matrix.qza**)
- Specify what column in your metadata file you want to examine (i.e., weight, wing damage)
- Gives you **Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qzv**, view in QIIME2 View
 - Repeat for all beta diversity metrics (**Bray-Curtis dissimilarity, Jaccard similarity, unweighted Unifrac distance, weighted UniFrac distance**)
- Visualize with the emperor ordination plots made for the beta diversity analyses
 - Weighted UniFrac plots are the most comprehensive overall, use weighed UniFrac distance matrix to make the ordination plots
 - Can make prettier ordination plots in R, see R notebook for details

6.) Run an indic species analysis or linear discriminant analysis to determine if the relative abundances of bacterial taxa differ between your variables

- I did indic species in R: see QIIME notebook and R code notebook for details
- If you can get the website to work, do an LDA at the Harvard website (<http://huttenhower.sph.harvard.edu/galaxy/>), defaults are OK)
- Visualize with grouped taxa bar plots, or if abundances are really small, make individual bar plots or boxplots

7.) Make grouped taxa bar plots to go along with the indic species analyses

- Use **qiime feature-table group** and **--p-mode 'mean-ceiling'** to get the average abundances of taxa across all samples, grouped by whatever variable you're interested in (i.e., ectoparasites, bat species, etc.), see notebook for details

APPENDIX E

QIIME2 16S rRNA Data Processing Bioinformatics Notebook

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington state bats

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats to compare their relative abundances sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the most abundant culturable bacterial isolates from western bats will differ from the most abundant bacteria in our sequencing data.

This sequencing data comes from maternity colonies sampled across Washington state from April – June 2021 with the WDFW.

QIIME2 Program and Version: qiime info

```
(qiime2-2021.11) qiime2@qiime2core2021-2:~$ qiime info
```

System versions

Python version: 3.8.12

QIIME 2 release: 2021.11

QIIME 2 version: 2021.11.0

q2cli version: 2021.11.0

Installed plugins

alignment: 2021.11.0

composition: 2021.11.0

cutadapt: 2021.11.0

dada2: 2021.11.0

deblur: 2021.11.0

demux: 2021.11.0

diversity: 2021.11.0

diversity-lib: 2021.11.0

emperor: 2021.11.0

feature-classifier: 2021.11.0

feature-table: 2021.11.0

fragment-insertion: 2021.11.0

gneiss: 2021.11.0

longitudinal: 2021.11.0

metadata: 2021.11.0

phylogeny: 2021.11.0

quality-control: 2021.11.0

quality-filter: 2021.11.0

sample-classifier: 2021.11.0

taxa: 2021.11.0

types: 2021.11.0

vsearch: 2021.11.0

Application config directory

/home/qiime2/miniconda/envs/qiime2-2021.11/var/q2cli

Getting help

To get help with QIIME 2, visit <https://qiime2.org>

To activate the latest version of QIIME before you start:

```
conda activate qiime2-2021.11
```

2/11/2022: I created my manifest file for our single-end sequencing data and tried to import my data into QIIME2.

Creating your manifest file:

- The headers must be the way they are below, with “sample-id” and “absolute-filepath”
 - Also, QIIME doesn’t like underscores (_), dashes (-) or spaces in sample IDs, so I replaced all of those with periods
 - To get the absolute file path, I dragged and dropped my zipped fasta files into the QIIME2 terminal, and then copy-pasted those into my manifest for the corresponding sample
 - ***The manifest file I am using is in the directory I am working from (my “QIIME” folder) not a subfolder within that directory folder – this is important or else it won’t work for the import step!***
 - My manifest file was saved as a .csv, as this seems to work well for PCs. It seems that Mac users have some more wiggle room in regards to what file type they save their manifest file as (i.e., .txt), but I’ve had the best success with a .csv.

Getting started: it's QIIME for QIIME!

1. Activate your conda environment as needed: conda activate qiime2-2021.11
2. Set your working directory: cd <drag and drop your directory folder to the terminal>

Part I – Sequence data import and preparations (importing data, filtering and trimming sequences, aligning sequences, importing mapping file for later analyses)

1.) Importing Sequence Data into QIIME:

For single-end sequence data:

```
qiime tools import \
--type 'SampleData[SequencesWithQuality]' \
--input-path <dragged and dropped manifest file filepath> \
--output-path <file_name_single-end-demux.qza> \
--input-format SingleEndFastqManifestPhred33V2
```

The code I used:

```
qiime tools import \
--type 'SampleData[SequencesWithQuality]' \
--input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv' \
--output-path Bat_2021_single-end-demux.qza \
--input-format SingleEndFastqManifestPhred33V2
```

- QIIME didn't like the .csv. For some reason it was reading the headers as one long smushed word instead of as separate columns.
- Saving the manifest as different file types (text file, tab-delimited) straight from excel did not work. Dr. Magori thinks that when excel exports from a PC, some things are lost in translation and the Linux system on QIIME doesn't like that.

Code and error message:

```
(qiime2-2021.11) qiime2@qiime2core2021-2:/media/sf_QIIME/QIIME$ qiime tools import --type
'SampleData[SequencesWithQuality]' --input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv'
--output-path Bat_2021_single-end-demux.qza --input-format SingleEndFastqManifestPhred33V2
```

There was a problem importing /media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv:

/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest.csv is not a(n) SingleEndFastqManifestPhred33V2 file:

Found unrecognized ID column name 'sample-id,absolute-filepath' while searching for header. The first column name in the header defines the ID column, and must be one of these values:

Case-insensitive: 'feature id', 'feature-id', 'featureid', 'id', 'sample id', 'sample-id', 'sampleid'

Case-sensitive: '#OTU ID', '#OTUID', '#Sample ID', '#SampleID', 'sample_name'

NOTE: Metadata files must contain tab-separated values.

There may be more errors present in the metadata file. To get a full report, sample/feature metadata files can be validated with Keemei: <https://keemei.qiime2.org>

Find details on QIIME 2 metadata requirements here:

<https://docs.qiime2.org/2021.11/tutorials/metadata/>

2/15/2022: Dr. Magori and I troubleshooted the manifest file problem in R, got the sequence data imported, and created the .qzv file for the visualization step.

Troubleshooting in R – the code we used:

```
#Load in the dataset without the 'filepath', use data without the single quotations
```

```
bat3=read.csv("DC_Bat_2021_Manifest_No_Quote.csv",header=T)
```

#Fix the column name so they have the dashes

```
names(bat3)[1] = "sample-id"
```

```
names(bat3)[2] = "absolute_filepath"
```

```
#Save the edited file as a tab delimited file from R
```

```
write.table(bat3,file="DC_Bat_2021_Manifest_Tab_Fixed.txt",sep="\t",row.names=F,quote=F)
```

For some reason, when exporting the manifest from a PC to the Linux system (the QIIME system), it messes up the formatting and takes away the dashes in the file names. We used R to edit the manifest file and saved it to a tab-delimited file.

- First, we had to remove all single quotations (') around the file path names in excel. We think that these are added during the transfer between the two systems. So, the 'file paths' were changed to just file paths with no quotations
 - Then, we edited the column names in R to fix the transfer problem
 - We saved the manifest as a tab-delimited file. This file format seems to save the changes to the headers, so QIIME doesn't have a problem reading them

The manifest that worked:

- We removed the single quotations from the copy-pasted file paths, and then fixed the file in R and saved it as a tab-delimited file
- My manifest was saved in the directory file I was working from, not a subfolder within that directory

1.) Import the Bat 2021 sequence data into QIIME:

```
qiime tools import \
--type 'SampleData[SequencesWithQuality]' \
--input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt' \
--output-path Bat_2021_single-end-demux.qza \
--input-format SingleEndFastqManifestPhred33V2
```

```
qiime2@qiime2core2021-2: /media/sf_QIIME/QIIME
File Edit View Search Terminal Help
rt --type 'SampleData[SequencesWithQuality]' --input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt' --output-path Bat_2021_single-end-demux.qza
--input-format SingleEndFastqManifestPhred33V2
There was a problem importing /media/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt:
/nmedia/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt is not a(n) SingleEndFastqManifestPhred33V2 file:
Filepath on line 2 and column "absolute-filename" could not be found ('/media/sf_QIIME/QIIME/Bat_2021_Sequence_Data/20220203_RRL11_JW9664_S2_R1_001.fastq.gz') for sample "RR.L.11".
(qiime2-2021.11) qiime2@qiime2core2021-2:/media/sf_QIIME/QIIME$ qiime tools import \
--type 'SampleData[SequencesWithQuality]' --input-path '/media/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt' --output-path Bat_2021_single-end-demux.qza
--input-format SingleEndFastqManifestPhred33V2
Imported /media/sf_QIIME/QIIME/DC_Bat_2021_Manifest_Tab_Fixed.txt as SingleEndFastqManifestPhred33V2 to Bat_2021_single-end-demux.qza
(qiime2-2021.11) qiime2@qiime2core2021-2:/media/sf_QIIME/QIIME$ qiime demux summarize \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_single-end-demux.qza' \
--o-visualization Bat_2021_single-end_demux.qzv
Saved Visualization to: Bat_2021_single-end_demux.qzv
(qiime2-2021.11) qiime2@qiime2core2021-2:/media/sf_QIIME/QIIME$
```

Success!

2.) Visualize the sequence data in QIIME: use the .qza file from the import step here

For single-end reads:

```
qiime demux summarize \
--i-data <drag and drop your single-end demux.qza file from import step here> \
--o-visualization <file_name_single-end_demux.qzv>
```

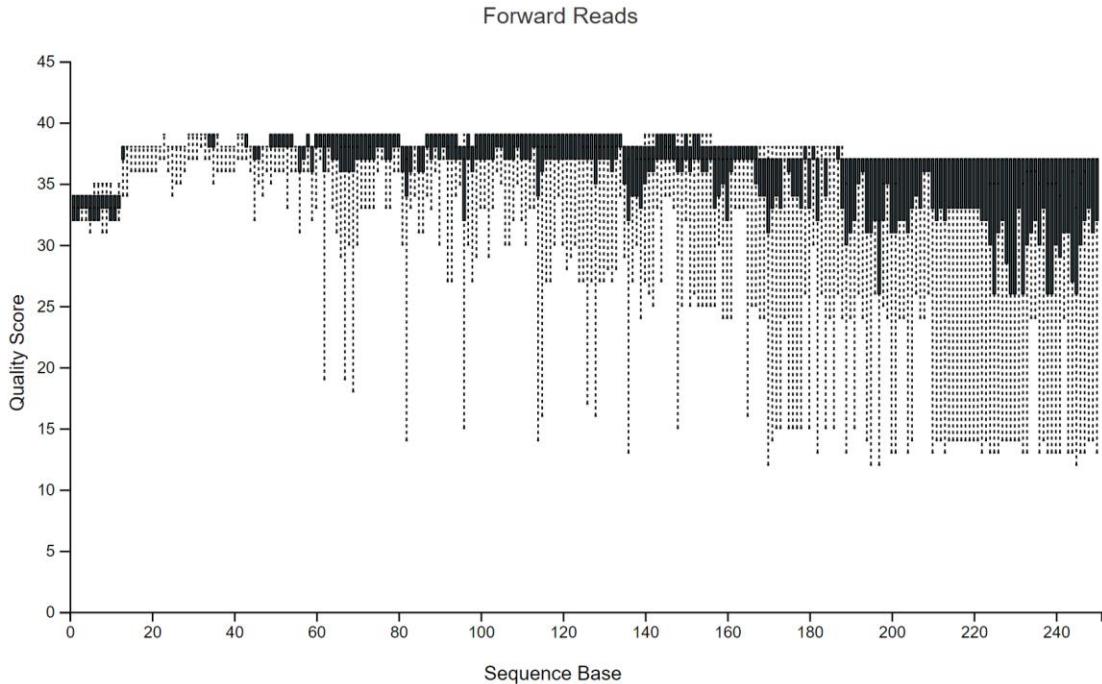
The code I used to visualize my single-end sequence data:

```
qiime demux summarize \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_single-end-demux.qza' \
--o-visualization Bat_2021_single-end_demux.qzv
```

Use the .qzv file from this step in qiime2view to visualize the sequence quality

3.) QZV File visualization via qiime2view: <https://view.qiime2.org/>

Quality plot: what a beauty!



- The sequence data quality drops off initially around 140 bases, although it drops off even more around 190 bases
- Total number of sequences: 8,374,027 sequences
- Number of reads per sequence: average of 87,229.45
- Sample with the highest number of reads: OR LS-18, with 154,469 reads
- Sample with the lowest number of reads: LC C-5, with 9,144 reads
 - 2nd lowest: E-C with 11,691 reads
 - 3rd lowest: S_13-19 with 26,453 reads, ~1,500 more sequences than E-C

Where should I trim the sequences for the deblur step?

- **Visually:** left (5') = 18, right (3') = 220
- **Figaro:** <https://github.com/Zymo-Research/figaro> → use command line installation and code
 - Amplicon length = 250 (250 bp long total)
 - Forward primer length (515f) = 19, have barcode tag as well (Parada)
 - Reverse primer length (926r) = 20 (Quince)
 - Path to output → drag and drop a folder for a location, give it a unique name, specify location
 - Figaro ended up not working so we didn't bother with it
- **Quality score below 25? Look at dark lines going down**

What is deblur and what does it do?

Deblur is a quality-control step used to trim and filter single end sequence data by their quality. Sequences that fall below a certain quality threshold are discarded, and low-quality reads are trimmed out.

2/22/2022: we gave up trying to mess with Figaro to figure out where to trim the sequences, so we decided to run the quality control step to see how many sequences were filtered out vs. how many we started with in the original qzv file from the import step.

After this, we will trim the sequences by sight (either no trimming at all because the quality of these sequences is not that bad, or after 220). By sight, we looked at the dark bands coming down from the curve. Most people use a quality score of 25 as the cut-off (which most of ours are above), although using 30 if possible is even better.

I updated my mapping file with all of the information we collected from bat sampling (species, sex, forearm length, etc.) and fixed any formatting issues (i.e., making sure to include periods in the sample ID names instead of spaces or underscores) by checking it with **Keemei** (<https://keemei.qiime2.org/>). The metadata file was supposedly good when I checked it with Keemei (make sure you copy-paste the contents of the excel sheet, Keemei won't be there unless it's a Google Sheet), but we'll see.

Useful info about deblur and QIIME processing workflow in general:

https://awbrooks19.github.io/vmi_microbiome_bootcamp/rst/3_sequences_to_composition.html

1.) Quality control using Deblur Filter by quality score: gets rid of poor quality sequences based on default parameters (*do this first!*). You can also change up these parameters as needed too. If the default removed too many sequences, go ahead and adjust the parameters (good quality sequences vs. enough sequences for analysis)

Note: the filtering step took around 40 minutes before it was completed

```
qiime quality-filter q-score \
--i-demux <drag and drop file-name_single-end-demux.qza from import data step> \
--o-filtered-sequences <file_name_demux-filtered.qza> \
--o-filter-stats <file_name_demux-filter-stats.qza>
```

The code I used for deblur filter quality control:

```
qiime quality-filter q-score \
--i-demux Bat_2021_single-end-demux.qza \
--o-filtered-sequences Bat_2021_demux-filtered.qza \
--o-filter-stats Bat_2021_demux-filter-stats.qza
```

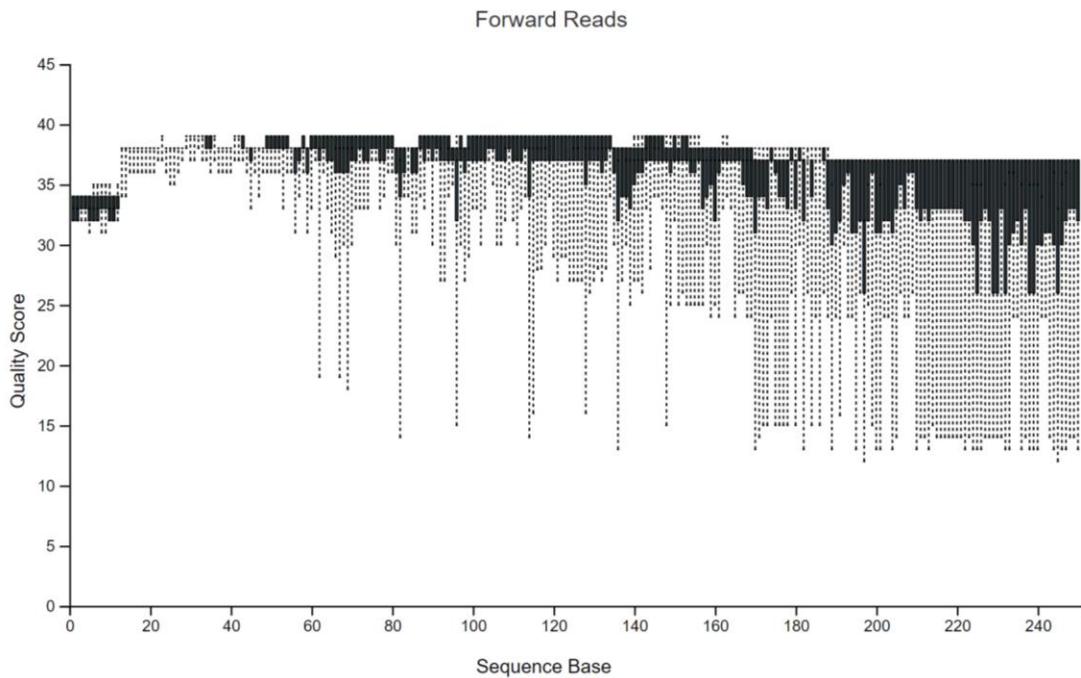
2.) Visualize the output from the filter quality control step:

```
qiime demux summarize \
--i-data <drag and drop file-name_demux-filtered.qza from deblur filter quality control step> \
--o-visualization <file-name_demux-filtered.qzv >
```

The code I used to visualize the demux-filtered qza output file: → never use the unfiltered qza file, only use filtered qza as the input for analyses from now on!

```
qiime demux summarize \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_demux-filtered.qza' \
--o-visualization Bat_2021_demux-filtered.qzv
```

Filtered sequences (Bat_2021_demux-filtered.qzv) QZV visualization in qiime2 view:



- Total number of sequences: 8,372,516 reads (vs. 8,374,027 unfiltered)
 - The filter step filtered out 1,511 sequences (lost $\sim 1.8 \times 10^{-4}$ % of total sequences, kept 99.9% of sequences)
- Number of reads per sequence: average of 87,213.71 reads (vs. 87,229.45 unfiltered)
- Sample with the highest number of reads: OR LS-18, with 154,445 reads (vs. 154,469 unfiltered)
- Sample with the lowest number of reads: LC C-5, with 9,144 reads (same as unfiltered)
 - 2nd lowest: E-C with 11,686 reads (vs. 11,691 unfiltered)
 - 3rd lowest: S_13-19 with 26,449 reads (vs. 26,453 unfiltered), $\sim 1,500$ more sequences than E-C

No crazy amounts of sequences were removed, so it's OK to stick with using the filtered data made with the default settings (the demux-filtered.qza) from here on out

3.) Visualize the filtered stats - visualize the sample list, how quality filtering step worked (use this in addition to the quality plot comparisons)

```
qiime metadata tabulate \
--m-input-file <file-name_demux-filter-stats.qza> \
--o-visualization <demux-filter-stats.qzv> \
```

Note: QIIME doesn't like it when you try to turn the filter stats into a qzv file – it will not work

Quality filtering: total # of sequences that were filtered out (total # or as a percentage)

How to look up a command to see the options (parameters) and info:

<The command you're interested in> --help

To look up info on qiime deblur denoise:

qiime deblur denoise-16S --help

2/25/2022: I ran deblur on my sequence data, trimming out the first 12 bases to cut out that lower quality section in the beginning, which may be due to residual primers [not actual bat sequences]. I also visualized the qza output from deblur, checking to see that the areas we wanted trimmed actually got trimmed out.

Note: I started deblur around 7:30 PM, was still not done when I checked again around 10:30 PM

Update: I don't know if QIIME pauses when the computer sleeps or not, but regardless this process wasn't finished even after overnight. In the future, remember that this process will take a very long time so plan accordingly.

Update on the update: the deblur process finished by 1:15 PM the following day (2/26/2022), so don't hold your breath for future deblur runs.

4.) Run the deblur process: check sequence length summary in QIIME2 view

Deblur groups based on sequence similarity and an algorithm (figures out if sequence is real biological sequence vs. an error based on how present it is across the whole dataset).

```
qiime deblur denoise-16S \
--i-demultiplexed-seqs <drag and drop your demux-filtered qza file> \
--p-trim-length 250 \
--p-left-trim-len 12 \ → this code is optional, trims out the first 12 bases (5' end, the left side)
--o-representative-sequences <file-name_rep-seqs-deblur.qza> \
--o-table <file-name_table-deblur.qza> \
--p-sample-stats \
--o-stats <file-name_deblur-stats.qza>
```

The code I used to run deblur on my filtered sequence data:

```
qiime deblur denoise-16S \
--i-demultiplexed-seqs '/media/sf_QIIME/QIIME/Bat_2021_demux-filtered.qza' \
--p-trim-length 250 \
--p-left-trim-len 12 \
--o-representative-sequences Bat_2021_rep-seqs-deblur.qza \
--o-table Bat_2021_table-deblur.qza \
--p-sample-stats \
--o-stats Bat_2021_deblur-stats.qza
```

Look at summary and quality plot again to verify that it did what we wanted it to do, extra primers on the left on explaining that drop?

3/1/2022: I visualized the deblur stats to see what the deblur step did to my sequence data via the deblur stats table. I also ran the qiime metadata tabulate on the filtered data just because. I imported my mapping (metadata) file after some troubleshooting. Turns out the secret lies with Google Sheets.

5.) Visualize the deblur stats – what did the deblur step do to my sequence data?

```
qiime deblur visualize-stats \
--i-deblur-stats <drag and drop the deblur-stats.qza file from previous step> \
--o-visualization <matched-file-name_deblur-stats.qzv>
```

The code I used to visualize the deblur stats:

```
qiime deblur visualize-stats \
--i-deblur-stats '/media/sf_QIIME/QIIME/Bat_2021_deblur-stats.qza' \
--o-visualization Bat_2021_deblur-stats.qzv
```

Deblur stats table in qiime2 view: visualize with the deblur-stats.qzv

	sample-id	reads-raw	fraction-artifact-with-minsize	fraction-artifact	fraction-missed-reference	unique-reads-derep	reads-derep	unique-reads-deblur	reads-deblur	unique-reads-hit-artifact	reads-hit-artifact	unique-reads-chimeric	reads-chimeric	unique-reads-hit-reference	reads-hit-reference	unique-reads-missed-reference	reads-missed-reference
0	S.13.19	26449	0.531135	0.000076	0.019319	1644	12403	213	7873	1	2	3	5	120	7545	7	152
1	S.S.7	48805	0.471673	0.000082	0.000145	2344	25789	175	13902	2	4	17	103	99	13682	1	2
2	E.C	11686	0.412117	0.000000	0.000843	807	6870	157	4759	0	0	7	13	93	4619	1	4
3	HH.M.14	56198	0.376188	0.000000	0.000511	2548	35057	179	19745	0	0	25	190	120	19460	1	10
4	WB.T.17	75594	0.375757	0.000079	0.003247	3722	47195	298	26439	3	6	36	259	148	25870	2	85
5	LC.C.6	93765	0.374937	0.000128	0.001641	4938	58621	291	29547	6	12	46	294	111	28952	3	48
6	S.S.16	61199	0.370513	0.000098	0.000743	2915	38530	180	21678	3	6	18	149	100	21389	1	16
7	CL.KL.4	37433	0.370128	0.000294	0.002069	2286	23589	220	13305	5	11	45	255	75	12850	1	27
8	POM.PO.5	69020	0.366517	0.000116	0.000000	3602	43731	200	23653	4	8	23	185	94	23287	0	0
9	CL.KL.5	94526	0.366174	0.000254	0.001929	4740	59937	305	31484	12	24	50	372	118	30784	3	60
10	CC.O.8	83761	0.365994	0.000143	0.002574	4021	53117	324	29515	6	12	52	377	115	28751	3	75
11	TL.KL.1	95882	0.365293	0.000188	0.000415	4737	60875	314	31784	9	18	65	427	126	31121	2	13
12	POM.PO.20	93994	0.365034	0.000319	0.000505	4775	59713	338	31947	15	30	38	266	133	31316	2	16
13	S.S.9	100995	0.364582	0.000139	0.007923	5011	64188	479	36552	7	14	82	455	176	35307	2	286
14	HH.M.8	86884	0.363312	0.000219	0.002559	4807	55337	336	28434	9	19	37	295	133	27765	4	72

- The unique sequences can act as preliminary indicators of how many unique bacterial species each sample will probably have
- Chimeric sequences are sequence hybrids
- Deblur uses the quality info to do the deblur step, so you can't see the quality plot after deblur runs, you only have your straight sequences without the quality information

6.) Visualize the representative sequences from the deblur step: will check and see if deblur did what it was supposed to do

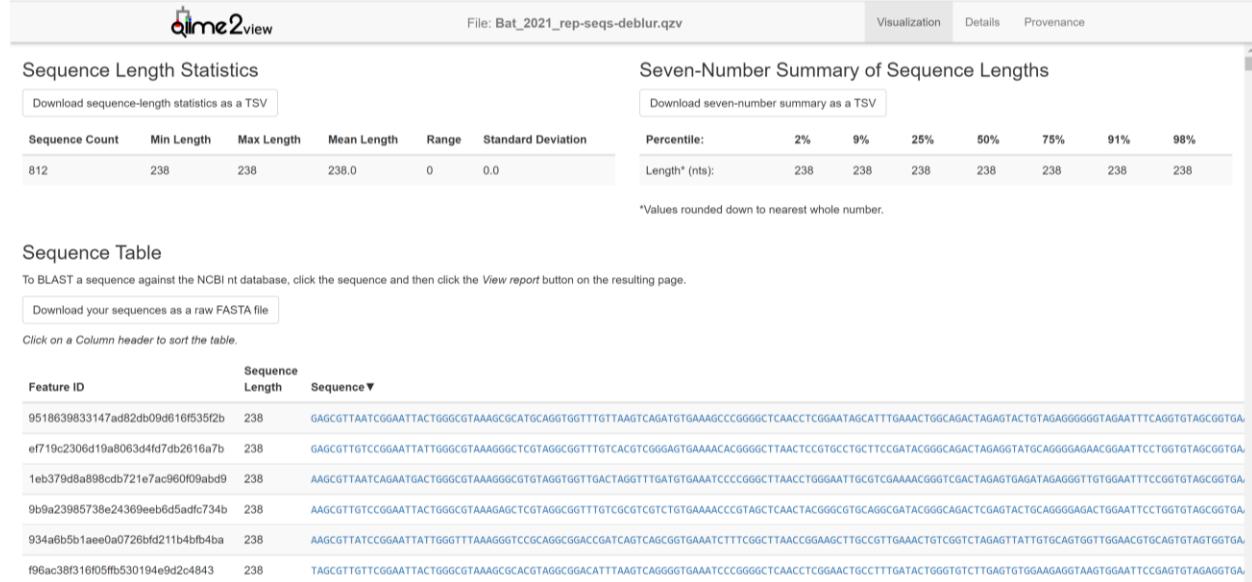
```
qiime feature-table tabulate-seqs \
--i-data <drag and drop the rep-seqs-deblur.qza from deblur step> \
--o-visualization <file-name_rep-seqs-deblur.qzv>
```

The code I used to visualize the representative sequences:

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
```

--o-visualization Bat_2021_rep-seqs-deblur.qzv

Deblur representative sequences qza visualization:



Sequence Table

To BLAST a sequence against the NCBI nt database, click the sequence and then click the View report button on the resulting page.

Download your sequences as a raw FASTA file

Click on a Column header to sort the table.

Feature ID	Sequence Length	Sequence▼
9518639833147ad82db09d616f53f2b	238	GAGCGTTAACGGAATTACTGGCGTAAAGCGCATGCAGGTGTTTAAGTCAGATGTGAAAGCCGGGCTCACCTCGGAATAGCATTGAAACTGGCAGACTAGACTGTAGAGGGGGTAGAATTTCAGGGTAGCGGTGA
e1f19c2306d19a8063d4fd7db2616a7b	238	GAGCGTTGCGGAATTATGGCGTAAAGGGCTCGTAGGGCGGTTGTCAGCTGGAGTGAAAACACGGGCTTAACCTCGTGCCTGCTCCGATAACGGCAGACTAGAGGTATGCAAGGGAGAACGGAATTCCGGTAGCGGTGA
1eb379d8a898cdb721e7ac960f09abd9	238	AAAGCGTTAACGAAATGACTGGCGTAAAGGGCGTGTAGGGTTGACTAGGTTGATGTGAAATCCCCTGGCTAACCTGGAAATTGCGTCAAACGGGGTGACTAGGTGAGATAGAGGGTTGGAATTCCGGTAGCGGTGA
9b9a23985738e24369eeb6d5adf734b	238	AAAGCGTTGCGGAATTACTGGCGTAAAGAGCTCGTAGGGCGGTTGCGCGTGTGAAACCCCGTAGCTCAACTACGGCGTGTGCAAGGGCAGACTGAGTACTGCAGGGAGACTGGAATTCCGGTAGCGGTGA
934a6bb5b1aee0a0726bf211b4bf4ba	238	AAAGCGTTATCCGGAAATTATGGGTTAAAGGGTCCGAGGGCGACGTCAGCTGGATAACCTTCCGGCTAACCGGAAGCTTCCGGTGAACACTGCGGTCTAGAGTTATGTGCACTGGTTGGAACGTGCGTGTAGGGTGA
f96ac38f316f05fb530194e9d2c4843	238	TAGCGTTGCGGAATTACTGGCGTAAAGCGACGTCAGGGCGACCTTAAGTCAGGGTGAAATCCCCTGGCTAACCTCGGAACCTGCTTGTACTGGGTCTTGAGTGTGAGAGGTAAAGGAATTCCGAGGTAGAGGTGA

- Looking at the sequence length statistics, we see that deblur did indeed trim out the 12 base pairs at the 5' end like we wanted ($350 - 12 = 238$)
- We copy-pasted a sequence into a new word document to make sure that the following primer sequences were not included in the final sequence (via “find”):
 - **Illumina 5' adapter:** AATGATACGGCGACCACCGAGATCTACACGCT
 - **Forward primer pad:** TATGGTAATT
 - **515F forward primer (Parada):** GTGYCAGCMGCCGCGTAA → Klaatu *parada* nikto

7.) Visualize and summarize the table from the deblur step: your metadata file (mapping file) will be imported in this step! This is what you will run your analyses on in later steps.

```
qiime feature-table summarize \
--i-table <drag and drop the table-deblur.qza file from the deblur step> \
--o-visualization <matched-file-name_table-deblur.qzv> \
--m-sample-metadata-file <drag and drop your metadata file in .tsv file format>
```

Note: your metadata file needs to be in a .tsv file format. Do this using Google Sheets

The code I used to (try and) visualize and summarize the deblur table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--o-visualization Bat_2021_table-deblur.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_DC_Mapping-File_TSV.txt'
```

Error code: something is wrong with our mapping file format

There was an issue with loading the file /media/sf_QIIME/QIIME/Updated_DC_Mapping-File_TSV.txt as metadata:

There was an issue with loading the metadata file:

Detected empty metadata ID. IDs must consist of at least one character.

There may be more errors present in the metadata file. To get a full report, sample/feature metadata files can be validated with Keemei: <https://keemei.qiime2.org>

Find details on QIIME 2 metadata requirements here:

<https://docs.qiime2.org/2021.11/tutorials/metadata/>

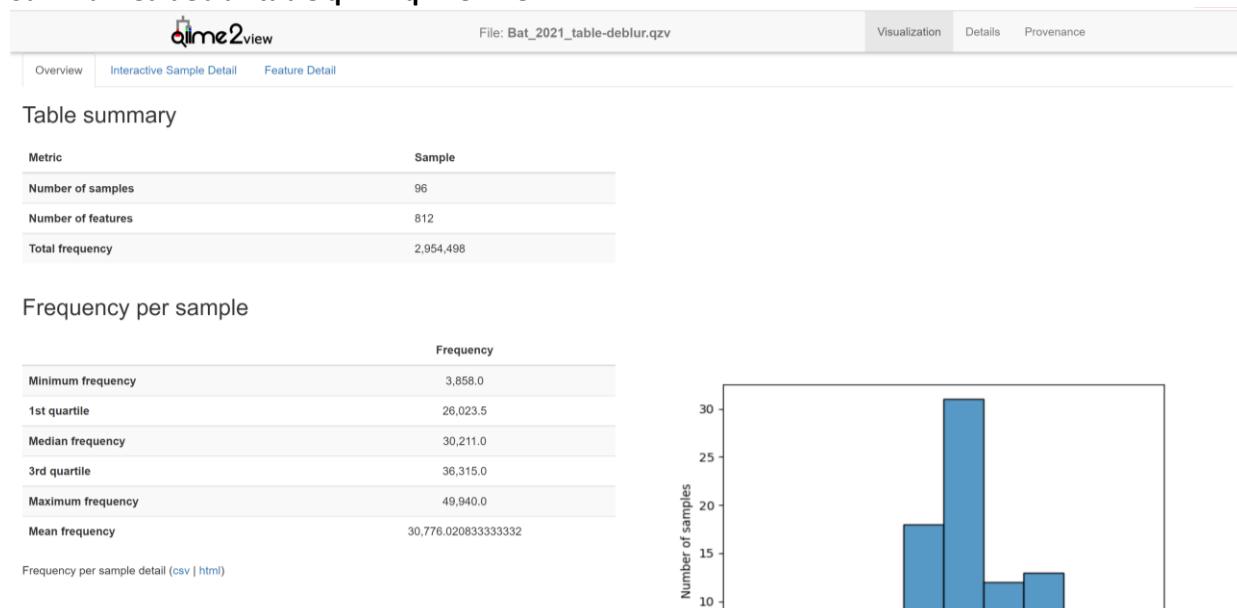
Update: I fixed the problem with my mapping file format (in your face QIIME!):

- Open a new Google Sheets spreadsheet
- Copy-paste all content from your master mapping file (your original excel sheet) into the Google Sheet
- Check it again with the Keemei extension
- File → download → as .tsv (tab separated values) → save to your QIIME folder

The code that actually worked for the visualize/summarize deblur table step:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--o-visualization Bat_2021_table-deblur.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'
```

Summarized deblur-table qzv in qiime2 view:



- How many samples are in the dataset? → **96 samples**
- How many features (bacterial “species”) are in the dataset? → **812**
- What is the total frequency (total number of DNA sequences in the dataset)? → **2,954,498 total sequences in the dataset**
- What is the frequency per sample? → **[mean] 30,776 per individual bat**

- What is the mean frequency per feature [scroll down!] (the mean number of sequences assigned to a feature, a.k.a., bacterial “species”)? → **3,638.54 per bacterial “species”**

Part II – Analyses on your sequence data

3/4/2022: I (tried to) assign taxonomy to my samples using the training feature classifier (qza file) made by Dr. Walke and Shelby Fettig.

a.) Preparing your sequences and assigning taxonomy to your sequence data

Note: the training feature classifier step tells QIIME where to search within the Silva database when assigning taxonomy to your samples, instead of trying to search throughout the whole database. In our case, the training feature classifier step had to be custom-made by Dr. Walke and Shelby Fettig because of the primers we are using (515f, 926r), as QIIME only has the 806r primer pre-installed.

Luckily, we were able to skip the training feature classifier steps since the qiime classifier qza was already made the **Shelby Fettig - silva_99_138.1_qiime_classifier.qza** file. I've included the training feature classifier code anyways in orange below for future reference, but I did not run through this code first-hand.

Update: just kidding, we weren't lucky. We ended up having to run through the training feature classifier steps because the old classifier file was out of date.

The reference database that we will compare our sequences to (the classifier file we uploaded into the taxonomy step) has been trimmed for our primers (515f, 926r). The Silva database has the full 16S rRNA gene (with a lot of sequences and assigned identities), but we only need to look at the V4-V5 region, which is the region specified by our primers (515f, 926r). The database is pared down in the classifier step so QIIME isn't trying to search the entire database.

1.) Training feature classifier:

```
qiime tools import \
--type 'FeatureData[Sequence]' \
--input-path silva_132_99_16S.fna \
--output-path silva_132_99_16S.qza
```

```
qiime tools import \
--type 'FeatureData[Taxonomy]' \
--input-format HeaderlessTSVTaxonomyFormat \
--input-path taxonomy_7_levels.txt \
--output-path <file-name_ref-taxonomy.qza>
```

```
qiime feature-classifier extract-reads \
--i-sequences silva_132_99_16S.qza \
--p-f-primer GTGYCAGCMGCCGCGGTAA \ → 515f primer sequence: GTGYCAGCMGCCGCGGTAA
--p-r-primer CCGYCAATTYMTTTRAGTT \ → 926r primer sequence: CCGYCAATTYMTTTRAGTT
--p-trunc-len 250 \
```

```
--p-min-length 250 \
--p-max-length 500 \
--o-reads <file-name_ref-seqs.qza>

qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads <drag and drop the ref-seqs.qza from the previous step> \
--i-reference-taxonomy <drag and drop the ref-taxonomy.qza> \
--o-classifier silva-132-515-926-nb-classifier-PC.qza → Shelby's qza file
```

2.) Assign taxonomy to your sequences with Silva database:

```
qiime feature-classifier classify-sklearn \
--i-classifier <silva-138-515-926-nb-classifier-PC.qza from training feature classifier steps> \
--i-reads <file-name_rep-seqs-deblur.qza from deblur step> \
--o-classification <file-name_taxonomy.qza>
```

The code I (tried) to use to assign taxonomy to my samples:

```
qiime feature-classifier classify-sklearn \
--i-classifier '/media/sf_QIIME/QIIME/Shelby Fettig - silva_99_138.1_qiime_classifier.qza' \
--i-reads '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--o-classification Bat_2021_taxonomy.qza
```

Error code: the pre-made classifier (Shelby's qza) appears to be outdated and therefore incompatible with my most current version of QIIME2. Looking around the QIIME2 forum, people either downloaded the most recent version of a ready-made classifier to use, or they had to re-classify to generate a most recent classifier file.

Plugin error from feature-classifier:

The scikit-learn version (0.23.1) used to generate this artifact does not match the current version of scikit-learn installed (0.24.1). Please retrain your classifier for your current deployment to prevent data-corruption errors.

Debug info has been saved to /tmp/qiime2-q2cli-err-otdrlyg.log

3/11/2022: I reclassified the training feature to generate an updated qza file classifier in the Silva database, using the code chunks previously in orange, starting from the extract reads step. I had to download the silva sequences qza file (**silva-138-99-sequences.qza**) and the reference taxonomy file (**silva-138-99-tax.qza**) from the QIIME2 website (<https://docs.qiime2.org/2021.11/data-resources/>).

Note: the extract reads step takes a long time. This step took just under 5 hours to run.

References for downloaded files from QIIME2:

For the sequence reference database:

Michael S Robeson II, Devon R O'Rourke, Benjamin D Kaehler, Michal Ziemska, Matthew R Dillon, Jeffrey T Foster, Nicholas A Bokulich. RESCRIPT: Reproducible sequence taxonomy reference database management for the masses. bioRxiv 2020.10.05.326504;
doi: <https://doi.org/10.1101/2020.10.05.326504>

For SILVA in general:

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* 41: D590 – D560

For the taxonomic framework:

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucl. Acids Res.* 42: D643 – D648

Training feature classifier: We were able to skip the first few chunks of code since QIIME already had the input files we needed. There were a few tricky parts regarding where to trim the sequences in the qiime feature-classifier extract-reads step. Visit the training feature classifier pages on the QIIME2 website (<https://docs.qiime2.org/2022.2/tutorials/feature-classifier/>) for more details if you will be training your own feature classifier in the future. The main takeaways though:

- Be very careful with this step!
- The most conservative option was to cut out the --p-trunc-len parameter and set --p-min-length and --p-max-length at 0 to turn them off. This is what I did, to avoid cutting out any good sequences.
- The --p-min-length and --p-max-length exclude amplicons that are far outside of the length you expect with the primers you used, because those are likely non-target sequences and should be excluded. When altering these codes, make sure to select settings that are appropriate for your marker genes!
- The --p-trunc-len parameter should only be used to trim the reference sequences if the search sequences are trimmed to the same length or shorter. If you don't know enough about this, do not mess with it!
 - It's tricky because single end reads may be variable in length.
- For untrimmed single-end reads, QIIME2 recommends training a classifier on sequences that have been extracted at the corresponding primer sites (in our case, 515f and 926r) but not trimmed.

```
qiime feature-classifier extract-reads \
--i-sequences <drag and drop the silva_132_99_16S.qza from Silva website> \
--p-f-primer GTGYCAGCMGCCGCGGTAA \ → 515f primer sequence
--p-r-primer CCGYCAATTYMTTTRAGTT \ → 926r primer sequence
--p-min-length 0 \ → 0 turns this off, --p-trunc-len code line removed too
--p-max-length 0 \ → 0 turns this off
--o-reads <file-name_ref-seqs.qza>
```

```
qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads <drag and drop the file-name_ref-seqs.qza from the previous extract reads step> \
--i-reference-taxonomy <drag and drop the ref-taxonomy.qza from Silva website> \ → Note: this is the
reference taxonomy file. The new taxonomy file you create in the taxonomy assignment step will be used in your later analyses
--o-classifier <file-name_silva-132-515-926-nb-classifier-PC.qza>
```

The code I used to train the feature classifier starting from the extract reads step:

```
qiime feature-classifier extract-reads \
--i-sequences '/media/sf_QIIME/QIIME/silva-138-99-seqs.qza' \
```

```
--p-f-primer GTGYCAGCMGCCGCGTAA \
--p-r-primer CCGYCAATTYMTTTRAGTT \
--p-min-length 0 \
--p-max-length 0 \
--o-reads Extract_Reads_ref-seqs.qza → this step was successful
```

```
qiime feature-classifier fit-classifier-naive-bayes \
--i-reference-reads '/media/sf_QIIME/QIIME/Extract_Reads_ref-seqs.qza' \
--i-reference-taxonomy '/media/sf_QIIME/QIIME/silva-138-99-tax.qza' \
--o-classifier Updated_2022_silva-132-515-926-nb-classifier-PC.qza → Unsuccessful, not enough RAM
on virtual box to run this
```

Error message for qiime feature-classifier fit-classifier-naive-bayes step:

Plugin error from feature-classifier:

Unable to allocate 1.00 GiB for an array with shape (134217728,) and data type float64

It looks like there's not enough RAM in my virtual box to complete this step. The QIIME forum suggested using a pre-trained classifier or running the code on another computer (i.e., not a virtual box) with more memory. I don't know how easy it would be to upgrade the virtual box memory, so it seems like having someone else with a Mac run this chunk with the input files ([Extract_Reads_ref-seqs.qza](#) and [silva-138-99-tax.qza](#)) would be simplest. Sigh.

3/13/2022: Dr. Walke ran the qiime feature-classifier fit-classifier-naive-bayes on her Mac using the [Extract_Reads_ref-seqs.qza](#) that I made earlier and the [silva-138-99-tax.qza](#) from the QIIME website. The process took 5 hours total to run. Using the updated classifier file from Dr. Walke's run ([Updated_Mar2022_Silva138_515_926_classifier.qza](#)), I tried to assign taxonomy to my sequence data.

Note: the classifier must be trained using the [exact same version](#) of QIIME that you have installed, otherwise it won't work in your assign taxonomy step (because you are using an older version of the classifier that does not match the latest version of the software). Dr. Walke made the classifier on her 2020 version of QIIME (I have the 2022 version) and it was not compatible with mine (the most recent 2022 version):

Plugin error from feature-classifier:

The scikit-learn version (0.22.1) used to generate this artifact does not match the current version of scikit-learn installed (0.24.1). Please retrain your classifier for your current deployment to prevent data-corruption errors.

Debug info has been saved to /tmp/qiime2-q2cli-err-q1ineznd.log

3/14/2022: Dr. Walke assigned taxonomy to my samples by running Shelby's classifier on her matching version of QIIME (my version was too recent to match). I filtered the mitochondria and chloroplasts from my table and converted it to a qzv.

The code Dr. Walke used to assign taxonomy (see #2 for more detail):

```
qiime feature-classifier classify-sklearn \
--i-classifier /Users/jwalke/Desktop/Dana\ QIIME\
13Mar22/ShelbyFettig_Silva_99_138.1_qiime_classifier.qza \
--i-reads /Users/jwalke/Desktop/Dana\ QIIME\ 13Mar22/Bat_2021_rep-seqs-deblur.qza \
--o-classification Bat_2021_taxonomy.qza
```

3.) Visualize the taxonomy.qzv file before you try to filter out the mitochondria and chloroplasts. See #6 for the code and the entry note on 3/15/2022 for details on why you should do this.

3.) Filter mitochondria and chloroplasts out of your table:

```
qiime taxa filter-table \
--i-table <drag and drop the table-deblur.qza from deblur step> \
--i-taxonomy <drag and drop the file-name_taxonomy.qza from taxonomy assignment step> \
--p-exclude D_4__Mitochondria,D_3__Chloroplasts,D_3__Chloroplast,Unassigned \
--o-filtered-table <file-name_filtered-table.qza> → make sure that this works! Verify that these were removed
```

The code I tried to use to filter mitochondria and chloroplasts from my deblur table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude <Whatever needs to be filtered out from your data, i.e., mitochondria, chloroplasts, unassigned, etc.> \
--o-filtered-table Bat_2021_filtered-table.qza
```

Open output table from that step (filtered table), convert to qzv and open it to check taxonomy
 Compare # of features in filtered steps, vs. deblur filtered table → make sure that there are fewer, 1-5 different features (ASV, bacterial “spp”) usually matched chloroplasts, mitochondria, unassigned (things that are not bacteria or archaea but made it into the sequence data anyways)

4.) Check that the mitochondria and chloroplasts were filtered out from your table:

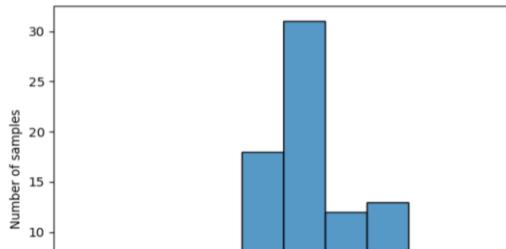
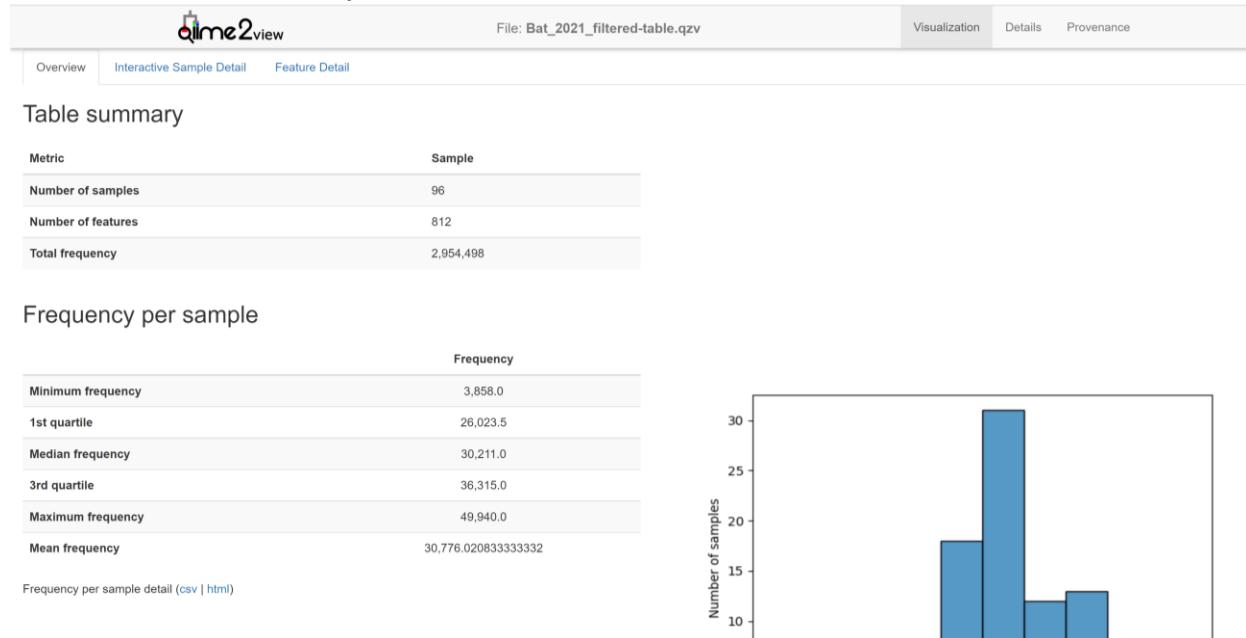
```
qiime feature-table summarize \
→ gives qzv file that you can look at, can see your sample list, bacteria, abundance
--i-table <file-name_filtered-table.qza from step #3> \
--o-visualization <matched-file-name_filtered-table.qzv> \
--m-sample-metadata-file <the mapping/metadata file you used earlier.tsv>
```

The code I tried to use to check that the mitochondria and chloroplasts were filtered from my deblur table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_filtered-table.qza' \
--o-visualization Bat_2021_filtered-table.qzv \
```

--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'

Summarized filtered-table.qzv in QIIME2 View:



- How many samples are in the dataset? → 96 samples → 96 samples
- How many features (bacterial “species”) are in the dataset? → 812 → 812
- What is the total frequency (total number of DNA sequences in the dataset)? → 2,954,498 total sequences in the dataset → 2,954,498
- What is the frequency per sample? → [mean] 30,776 per individual bat → 30,776
- What is the mean frequency per feature [scroll down!] (the mean number of sequences assigned to a feature, a.k.a., bacterial “species”)? → 3,638.54 per bacterial “species” → 3,638.54

The filtering step did not work this first time around. All values are identical to the unfiltered sequences and tables, so something needs to be adjusted.

Update: the parameters I used in the --p-exclude line were specifically for Phillip’s data set since I used his code as a template, so these parameters were not in my dataset. The --p-exclude parameters are specific to your dataset, so look at your taxonomy.qza file first to see what you will need to filter out.

Otherwise, if something is not present the default is that QIIME filters nothing, explaining why nothing happened this first time around.

5.) Filter mitochondria and chloroplasts out of your sequences:

```
qiime taxa filter-seqs \
--i-sequences <file-name_rep-seqs-deblur.qza from deblur step> \
--i-taxonomy <file-name_taxonomy.qza from taxonomy step> \
--p-exclude <Whatever needs to be filtered out from your data, i.e., mitochondria, chloroplasts, unassigned, etc.> \
--o-filtered-sequences <file-name_filtered-rep-seqs.qza>
```

The code I tried to use to filter the mitochondria and chloroplasts from my sequences:

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude D_4__Mitochondria,D_3__Chloroplasts,D_3__Chloroplast,Unassigned \
--o-filtered-sequences Bat_2021_filtered-rep-seqs.qza → unsuccessful, re-run
```

The code I used to visualize the representative sequences:

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--o-visualization Bat_2021_rep-seqs-deblur.qzv
```

3/15/2022: Dr. Walke and I troubleshooted the mitochondria/chloroplast filtering for the table and sequences, so we got those correctly filtered out. I visualized the taxonomy of my sequences in both the table form (#6) and as taxa bar plots (#7), in addition to making the phylogenetic tree and (finally) getting through the alpha rarefaction step.

When filtering out the mitochondria and chloroplasts, you need to **look in your taxonomy.qza (#6) first to determine what needs to be removed and match the formatting exactly in --p-exclude**. However, when we did this, QIIME didn't like our matched formatting (i.e., g__Chloroplast), probably because of the double underscore. We ended up having to do this as a two-step process for filtering the chloroplasts and mitochondria from the table and sequences, because QIIME didn't like it when we tried to do this all at once (i.e., Mitochondria,Chloroplast), probably because QIIME wanted a semicolon(;) not a comma. I'm not sure if using a semicolon would allow you to do it all in one fell swoop.

Mitochondria are usually from the host (in our case, from the bats) while the chloroplasts are generally from the plants in the environment (i.e., bats roosting in trees). Both are not part of the microbiome and need to be removed before any analyses.

The code I used to [actually] filter mitochondria and chloroplasts from my deblur table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_table-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Mitochondria \
--o-filtered-table Bat_2021_Real_filtered-table.qza
```

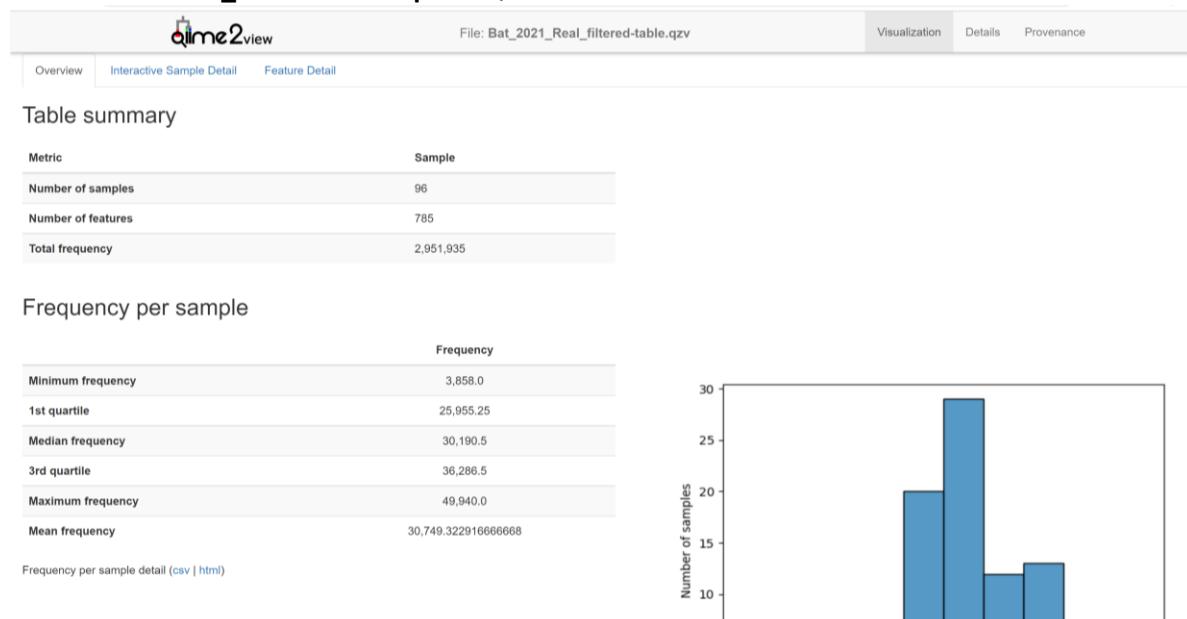
```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Chloroplast \
--o-filtered-table Bat_2021_Real_filtered-table.qza
```

→ Use **Bat_2021_Real_filtered-table.qza** from now on

The code I used to check that the mitochondria and chloroplasts were filtered from my deblur table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--o-visualization Bat_2021_Real_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete Mapping File.tsv'
```

Summarized Real_filtered-table.qzv in QIIME2 View: this worked!



- How many samples are in the dataset? → **96 samples** → **96 samples**
- How many features (bacterial “species”) are in the dataset? → **812** → **785**
- What is the total frequency (total number of DNA sequences in the dataset)? → **2,954,498 total sequences in the dataset** → **2,951,935**
- What is the frequency per sample? → [mean] **30,776 per individual bat** → **30,749.32**
- What is the mean frequency per feature [scroll down!] (the mean number of sequences assigned to a feature, a.k.a., bacterial “species”)? → **3,638.54 per bacterial “species”** → **3,760.42**

The code I used to [actually] filter the mitochondria and chloroplasts from my sequences:

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Mitochondria \
```

3/17/2022: Looks like I used the wrong input file for this step. I should have used the Real_filtered from the first chunk like I did for the table. See the code on 3/17/2022 for

```
--o-filtered-sequences Bat_2021_Real_filtered-rep-seqs.qza
```

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_rep-seqs-deblur.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Chloroplast \
--o-filtered-sequences Bat_2021_Real_filtered-rep-seqs.qza
```

The code I used to visualize my Bat_2021_Real_filtered-rep-seqs.qza:

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-rep-seqs.qza' \
--o-visualization Bat_2021_Real_filtered-rep-seqs.qzv
```

→ Use Bat_2021_Real_filtered-rep-seqs.qza from now on

6.) Visualize the taxonomy of your sequences: *Do this before the mitochondria/chloroplast filtering step in the future.*

```
qiime metadata tabulate \
--m-input-file <file-name_taxonomy.qza> \
--o-visualization <matched-file-name_taxonomy.qzv>
```

The code I used to visualize taxonomy of my sequences:

```
qiime metadata tabulate \
--m-input-file '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--o-visualization Bat_2021_taxonomy.qzv
```

Taxonomy visualization in QIIME2 View: very useful file to refer to!

Feature ID #q2-types	Taxon	Confidence
	categorical	categorical
00226c6c55ac0ff79fa62454779c7b05	d_Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Micrococcales; f__Micrococcaceae	0.9997112830316928
00559626b6cc8e90b491e52ada0c8b4f	d_Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Caulobacteraceae; g__Brevundimonas	0.9999226862658391
0091c580b2ea1fbbaed91f2c37a7ae	d_Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae	0.996056082347
00c70932b5b33a1b0f57574566d8d404	d_Bacteria; p__Actinobacteriota; c__Actinobacteria	0.9667167904091691
00d71527af4c51a79a34f999e9c1fd94	d_Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rhizobiales; f__Beijerinckiaceae; g__1174-901-12	0.9997696663761819
0116c1f55934b0812fec333f0b92be65	d_Bacteria; p__Bacteroidota; c__Bacteroidia; o__Sphingobacteriales; f__Sphingobacteriaceae	0.9999992264961638
017fbdd33cf8407bf60b7547bfe8473	d_Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Xanthomonadales; f__Xanthomonadaceae; g__Stenotrophomonas	0.9978054969922782
02139f4d0f7e7b5b8177d9d8aa4233	d_Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Micrococcales; f__Microbacteriaceae; g__Leucobacter	0.8468665752358634
02571eb270b90bfc92ccf0872b78120	d_Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Corynebacteriales; f__Dietziaceae; g__Dietzia	0.9663668426426844
025a0ac19040f054f3e79797c08a5ba5	d_Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae	0.8371990007848171
028701138db649b04390028d42be3dbc	d_Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Tannerellaceae; g__Macelilbacteroides	0.9998981166011289
02963118ec968a339e9c0137467cce329	d_Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Bacillaceae	0.8255020588966776
02a3c23986d33b79bf077811a244bd9b	d_Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Acetobacteriales; f__Acetobacteraceae; g__Endobacter; s__Acetobacteraceae_bacterium	0.7284123953466564
02c54d0cf51f5ff190be0c045f85fd7e	d_Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Corynebacteriales; f__Corynebacteriaceae	0.9989216325038046

- Sift through your data and determine what needs to be filtered out:
 - We had one mitochondrion and ~30 chloroplasts, and no unassigned sequences. The fewer number of features in our Bat_2021_Real_filtered-table.qzv confirms that everything was filtered out correctly.
- Use the search bar at the top of the page to search for anything you need to remove, or fun and wiley bacteria to focus on in your discussion. From a quick glance through this table, we found:

- Family Chitinophagaceae: antifungal bacteria that break down chitin, the primary component of fungal cell walls!
- *Rhodococcus* spp: antifungal bacteria, show up a lot in our lit and widely studied as a naturally-occurring bat skin bacterium being developed as a probiotic for Pd. We appear to have a lot of *Rhodococcus*.
- *Pseudomonas* spp: antifungal bacteria, show up a lot in our lit and widely studied as a naturally-occurring bat skin bacterium being developed as a probiotic for Pd. We appear to have a few *Pseudomonas*.
- *Vibrio* spp: Well-known contaminant of DNA extraction kits. Generally kit contamination is even more pronounced in cases where the amount of host bacteria are low (as in our case with the bat skin bacteria).

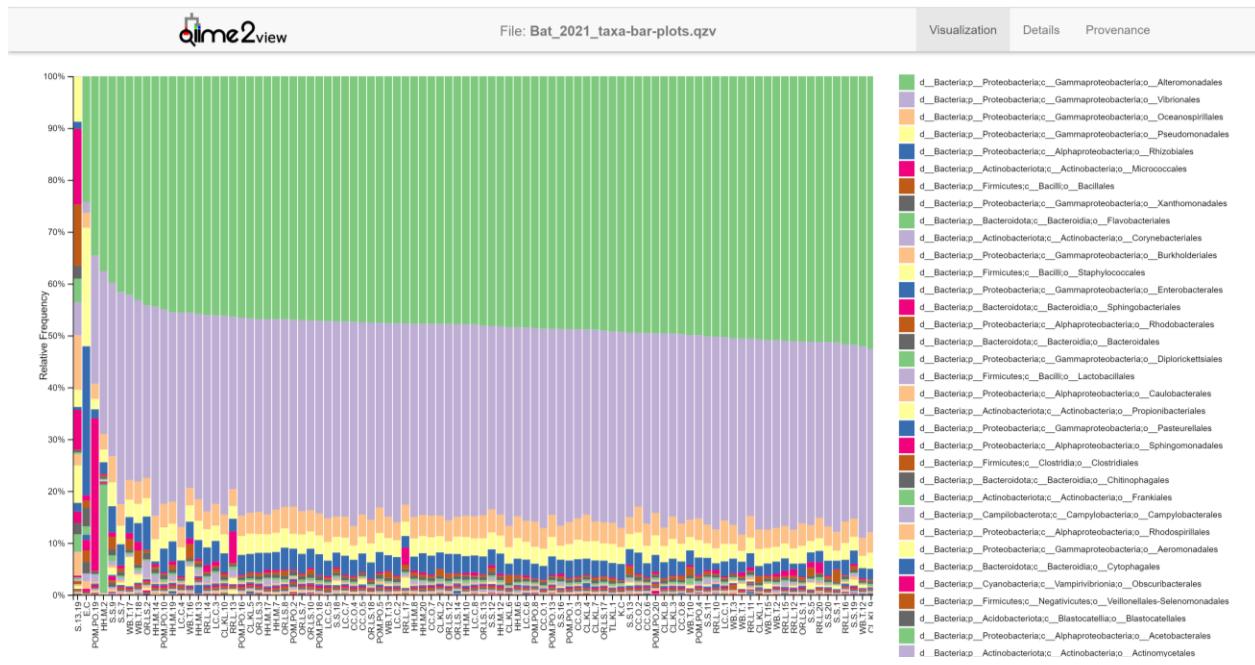
7.) Visualize the microbiome using taxa bar plots:

```
qiime taxa barplot \
--i-table <file-name_filtered-table.qza from filter table step> \
--i-taxonomy <file-name_taxonomy.qza from taxonomy step> \
--m-metadata-file < the mapping/metadata file you used earlier.tsv > \
--o-visualization <file-name_taxa-bar-plots.qzv>
```

The code I used to make taxa bar plots:

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021 Complete \
Mapping File.tsv' \
--o-visualization Bat_2021_taxa-bar-plots.qzv
```

Taxa bar plots in QIIME2 View:



We think the orders Alteromonadales (light green) and Vibionales (lavender) are attributed to contamination from our Qiagen DNEasy extraction kits. Apparently kit contamination straight from the factory is not uncommon.

b.) Create phylogenetic tree and rarefy sequence data

1.) Generate a tree for phylogenetic diversity analysis

```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences <file-name_filtered-rep-seqs.qza> from sequence filter step> \
--o-alignment <file-name_aligned-rep-seqs.qza> \
--o-masked-alignment <file-name_masked-aligned-rep-seqs.qza> \
--o-tree <file-name_unrooted-tree.qza> \
--o-rooted-tree <file-name_rooted-tree.qza>
```

The code I used to build a phylogenetic tree:

```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_unrooted-tree.qza \
--o-rooted-tree Bat_2021_rooted-tree.qza
```

2.) Alpha rarefaction plotting → check plots together

```
qiime diversity alpha-rarefaction \
--i-table <file-name_filtered-table.qza> from table filter step> \
--i-phylogeny <file-name_rooted-tree.qza> from phylogen. tree step> \
--p-max-depth <The median frequency per sample in your filtered-table.qza> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_alpha-rarefaction.qzv>
```

The code I used to plot alpha rarefaction, after much frustration:

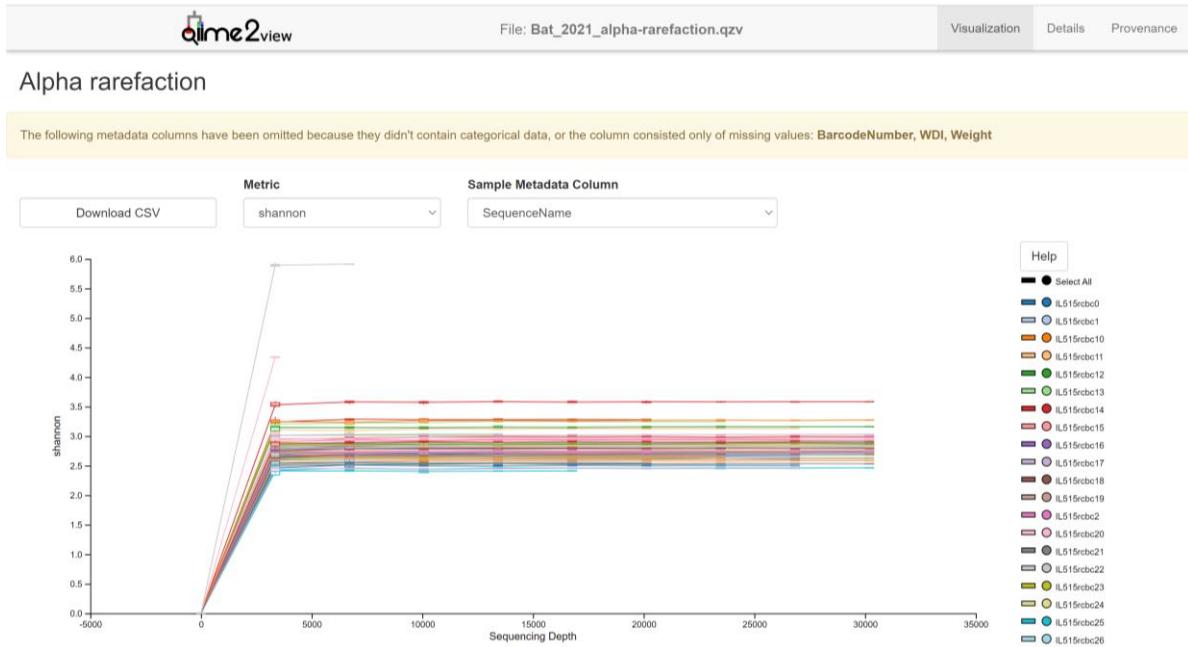
```
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_rooted-tree.qza' \
--p-max-depth 30191 \ → was 30,190.5, so I rounded to 30191
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_alpha-rarefaction.qzv
```

Word of warning on this sucker: this step was extremely fussy. For whatever reason QIIME did not like the forward slashes (/) or periods in my header names in my mapping file, even though using that same exact file was fine up to this point. What I ended up having to do:

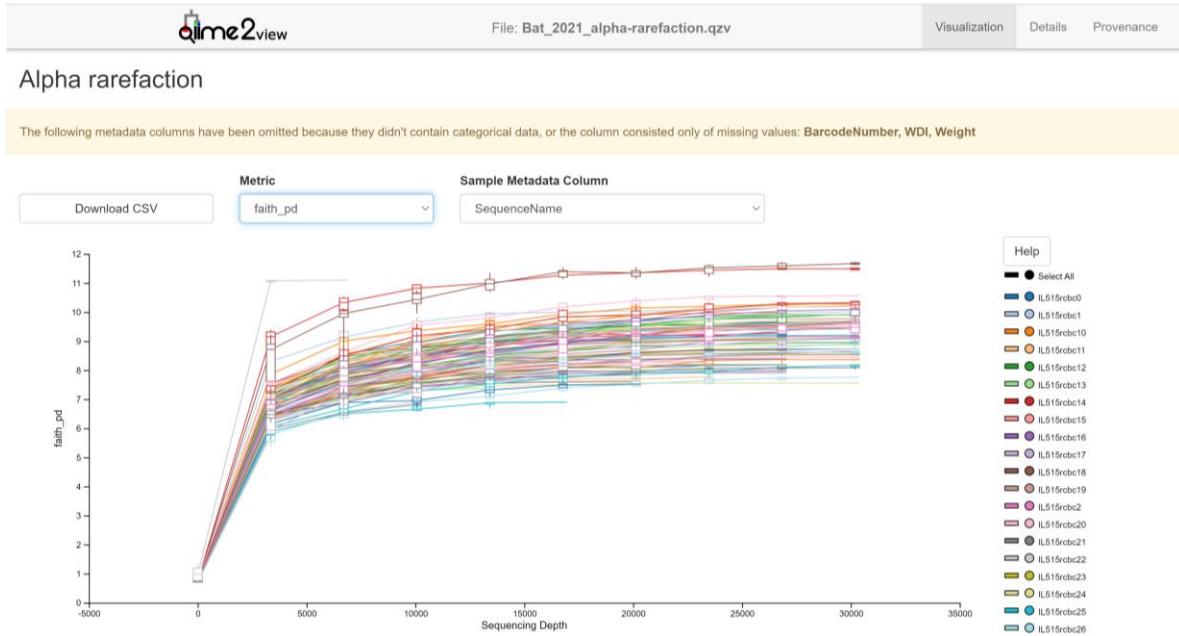
- I made another Google Sheets copy of my original mapping file Google Sheet that I used to make the .tsv (step #7, part I). It was too difficult to edit the .tsv mapping file directly without messing up the entire file formatting.
- In the copy, I removed all forward slashes and periods from my header titles. I also cleaned them up by removing all the information in parentheses, since that would show up on the figures. The information for each column can be looked up on the master excel sheet mapping file.
- I then exported this as a .tsv and dragged the file from my “downloads” folder into my “QIIME” folder so that the .tsv format would stay the same. If you try to open the .tsv after downloading and then “save as” to your directory sometimes it changes it to a text file, rendering it absolutely useless.

→ Use the Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv mapping file from now on

Alpha rarefaction plots in QIIME2 View: Shannon's



Alpha rarefaction plots in QIIME2 View: Faith's phylogenetic diversity



Alpha rarefaction plots in QIIME2 View: Observed features



- From the three alpha rarefaction plots, we decided to rarefy our sequences at a sequencing depth of

3/17/2022: Dr. Walke and I examined my alpha rarefaction plots, the Real_filtered table list of sequences with the number of features (sequences) and the taxa bar plot (downloaded as a .csv) to determine where to rarefy our sequences. Surprise, there's more complications.

Looking at the alpha rarefaction plots:

- We looked at where the lines began to level off, where there were no more big jumps from sequence depth to sequence depth.
- Ideally, we'd want to rarefy our sequences at the highest depth, but in doing so we'd lose many of our samples. Looking for that sweet spot with the highest sequence depth possible but still have enough sample types represented to answer the study question is the tricky part.
- Across the faith's phylogenetic diversity and observed_features plots, 12,700 and 17,000 appear to be decent areas to clip everything at. The Shannon's is level regardless so it doesn't matter what we do there.

Looking at the Real_filtered-table.qzv:

- Rarefying at 12,700 would remove the LC.C.5, the environmental control (E.C.), and Shelby's control sample (S.13.19).
- Rarefying at 17,000 would remove the three samples above, plus S.S.7 and CL.KL.4.
 - I don't think it would be huge deal to increase our sequence depth to 17,000 and sacrifice two more samples. The Spokane samples were all pretty comparable (all yuma myotis, all females, all no ectoparasites), as were the Klickitat samples (all big brown bats, all females, most had ectoparasites) so removing these samples should not influence our ability to answer our study question.

File: Bat_2021_Real_filtered-table.qzv

S.S.1	24506
WB.T.2	24008
S.S.19	23967
S.S.11	23946
RR.L.14	23877
RR.L.11	23700
WB.T.18	23338
POM.PO.5	23287
CL.KL.9	21540
S.S.16	21389
RR.L.20	20350
HH.M.14	19460
WB.T.5	17277
S.S.7	13672
CL.KL.4	12850
S.13.19	7450
E.C	4601
LC.C.5	3858

17,000 sequence depth
would remove these samples

12,700 sequence depth
would remove these samples

Looking at the taxa bar plot:

- The three most abundant taxa are likely contaminants:
 - Pseudoalteromonas, Vibrio, Halomonas*
 - The first two that are listed are well-known contaminants, while the *Halomonas* is a salt-loving bacterium so it makes sense that it would be a contaminant in something salty like our buffer. The large proportions of these taxa are likely not real members of the bat skin microbiome.
 - We will filter out these contaminants, similar to the chloroplast/mitochondria filtering step. After that, we will look at the updated taxa bar plot and re-rarefied plots and make a decision.

The code I used to filter out the contaminants from our table: I had to do this step by step, I couldn't figure out how to do it all at once.

Remember: when you do the filtering one by one, make sure that for the second thing you filter out that you use the output from the previous step, not the original file (otherwise you are only filtering out one thing, not two). Don't do what I did the first time I tried to filter out mitochondria and chloroplasts from my sequences. I redid all my filtering for the sequences just to be sure everything got out.

1.) The code I used to filter out the contaminants from our table:

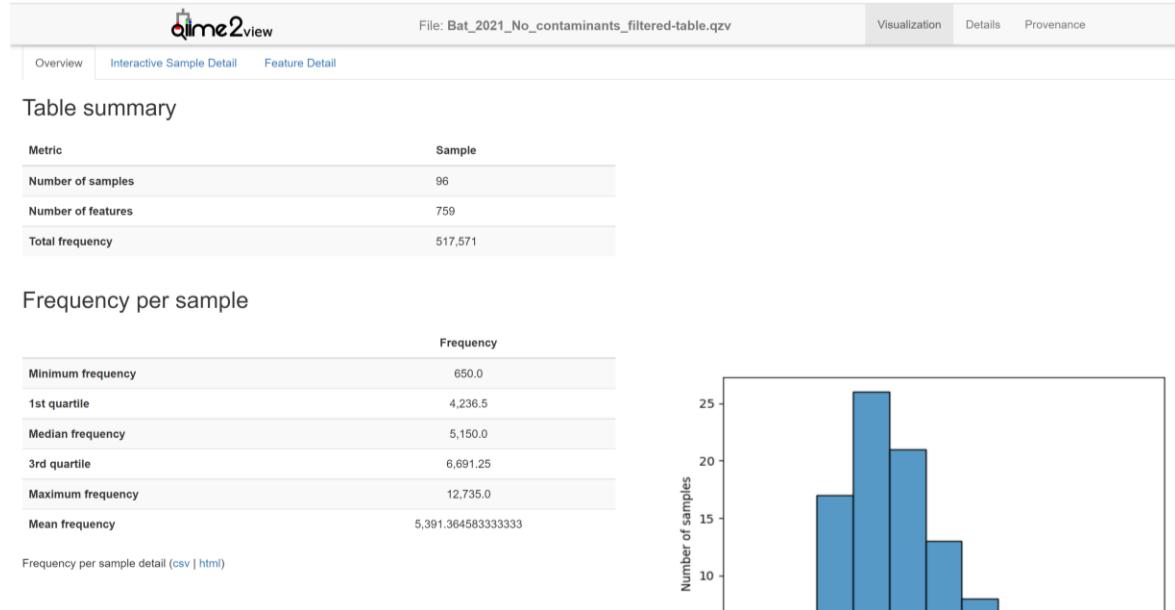
```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Vibrio \
--o-filtered-table Bat_2021_No_contaminants_filtered-table.qza
```

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \ ← output from
previous filter step
```

```
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Halomonas \
--o-filtered-table Bat_2021_No_contaminants_filtered-table.qza
```

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Pseudoalteromonas \
--o-filtered-table Bat_2021_No_contaminants_filtered-table.qza
```

→ **Bat_2021_No_contaminants_filtered-table.qza** is the most updated table at this point



2.) The code I used to check that the contaminants were filtered from the table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-table.qza' \
--o-visualization Bat_2021_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Updated_Bat_2021_Mapping_File - Bat 2021
Complete Mapping File.tsv'
```

3.) The code I used to filter out the contaminants from our sequences (I re-did mitochondria and chloroplasts too just to be sure they were removed):

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_Real_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Mitochondria \
```

```
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \ ← output
from previous filter step
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Chloroplast \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Vibrio \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Halomonas \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza

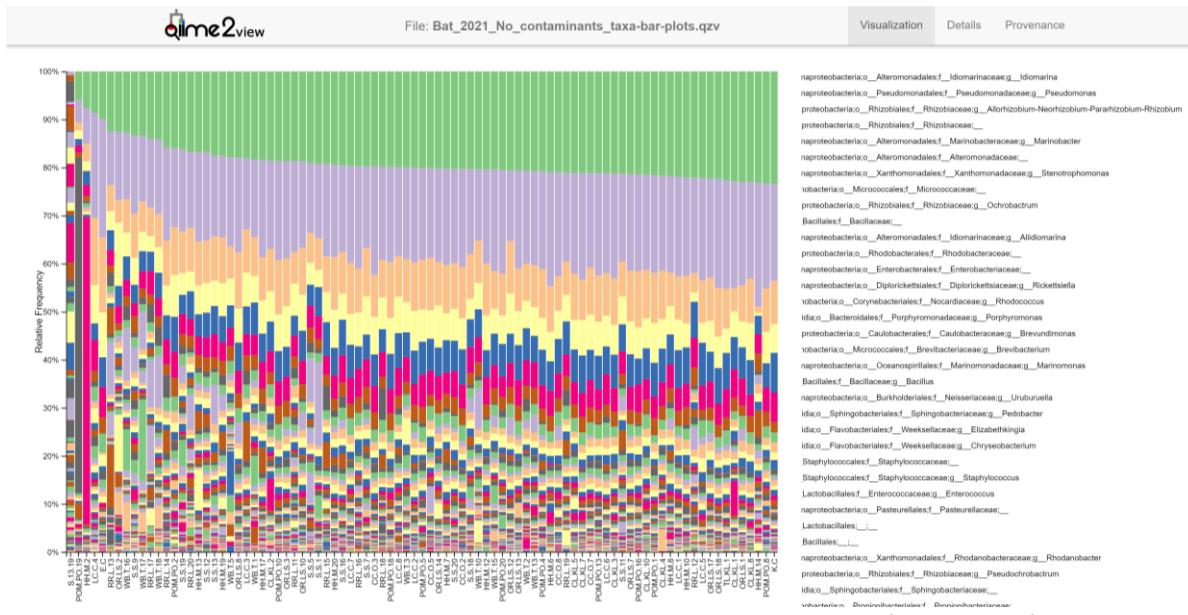
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Pseudoalteromonas \
--o-filtered-sequences Bat_2021_No_contaminants_filtered-rep-seqs.qza
```

4.) The code I used to visualize my Bat_2021_No_contaminants_filtered-rep-seqs.qza:

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--o-visualization Bat_2021_No_contaminants_filtered-rep-seqs.qzv
```

5.) The code I used to visualize the no contaminants microbiome using taxa bar plots:

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_contaminants_taxa-bar-plots.qzv
```



Good news, the genera *Pseudoalteromonas*, *Vibrio*, and *Halomonas* were filtered out from the table.

3/18/2022: I made a new phylogenetic tree using my no contaminants filtered sequences and replotted my alpha rarefaction plots using my no contaminants table. From this, Dr. Walke and I were able to determine where to rarefy the sequences for the data analyses.

1.) The code I used to build a phylogenetic tree using my no contaminants sequences:

```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_No_Contaminants_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_No_contaminants_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_No_contaminants_unrooted-tree.qza \
--o-rooted-tree Bat_2021_No_contaminants_rooted-tree.qza
```

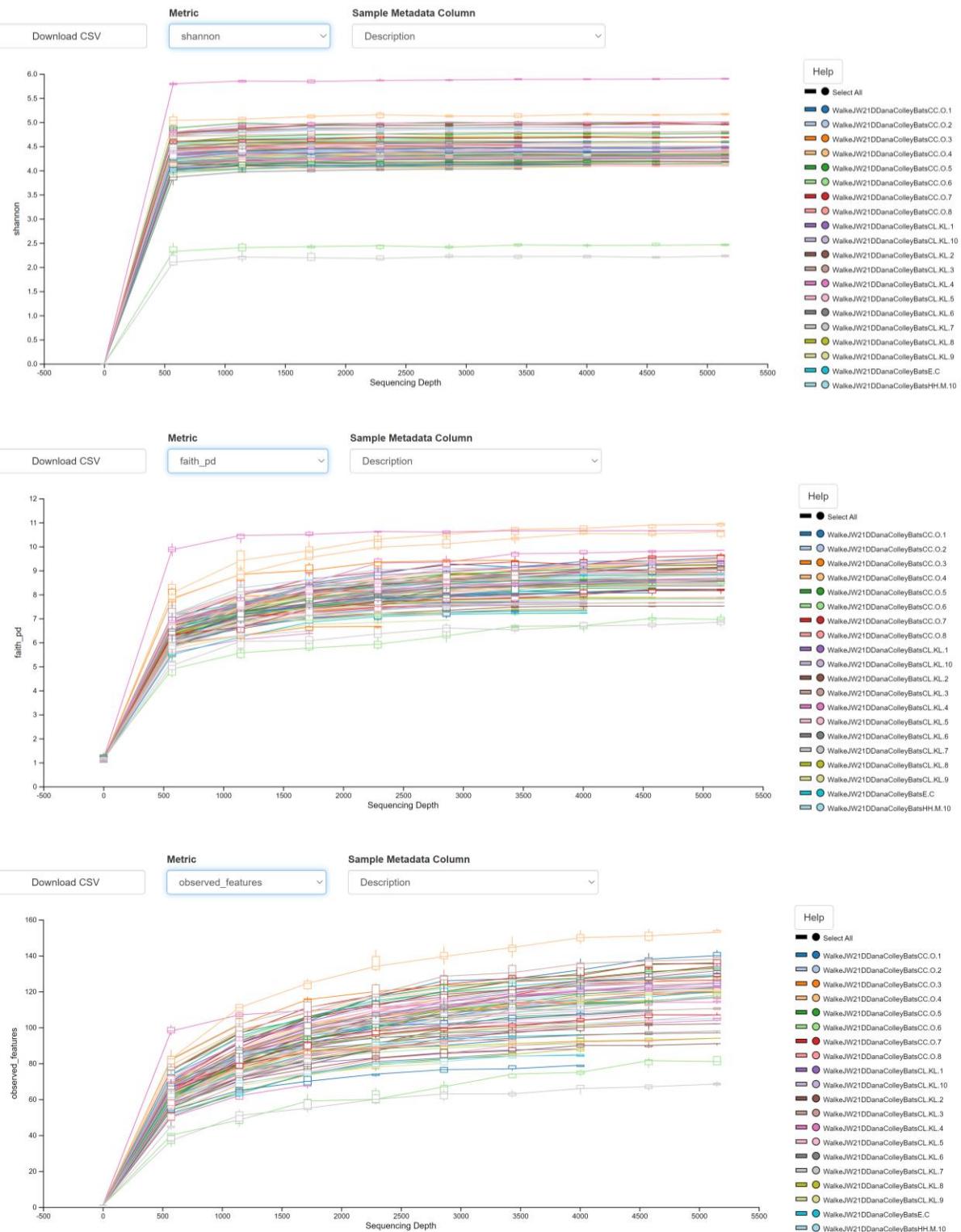
2.) The code I used to replot my alpha rarefaction using my no contaminants table

```
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_rooted-tree.qza' \
--p-max-depth 5150 \ ↪ new value from No_contaminants table
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_contaminants_alpha-rarefaction.qzv
```

Where should we rarefy our sequences?

The new no contamination alpha rarefaction plots appear to be more leveled out for the Faith's phylogenetic diversity metric. The Shannon metric is still the same (level) and so is the observed_features (increasing).

The no contaminants table with the list of samples with their number of sequences appears to be more leveled out as well. There are fewer large jumps between the lowest sequence samples.



Contaminant	Count
RR.L.12	3780
RR.L.11	3748
S.S.19	3678
RR.L.15	3643
WB.T.12	3627
S.S.11	3616
WB.T.2	3485
WB.T.1	3408
RR.L.19	3376
S.S.16	3180
HH.M.14	3177
RR.L.20	3123
S.S.1	3062
CL.KL.9	2949
S.S.7	2712
CL.KL.4	2096
WB.T.5	2077
L.C.C.5	650

- ~2,000 or 2,700 appears to be a good place to rarefy. However, the top hit after filtering (*Idiomarina*) appears to be another contaminant (see no contaminants taxa bar plot). It is a deep-sea halophilic bacterium, and it does not appear in other bat skin microbiome literature. It seems very unlikely that this is indeed a true member of the bat skin microbiome (and that it is the most abundant bacterium), so I will filter it out, and re-do all of the phylogenetic trees and rarefaction plot steps to see where we should trim everything.

3.) The code I used to filter *Idiomarina* from my table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Idiomarina \
--o-filtered-table Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza
```

4.) The code I used to check that the contaminants were filtered from the table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--o-visualization Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

5.) The code I used to filter *Idiomarina* from our sequences

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_contaminants_filtered-rep-seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Idiomarina \
--o-filtered-sequences Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington state bats – part 2

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats to compare their relative abundances sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the most abundant culturable bacterial isolates from western bats will differ from the most abundant bacteria in our sequencing data.

This sequencing data comes from maternity colonies sampled across Washington state from April – June 2021 with the WDFW.

Part II – Analyses on your sequence data – continued from part 1 of my bioinformatics notebook

a.) Preparing your sequences and assigning taxonomy to your sequence data – continued from part 1

3/19/2022: I visualized the no *Idiomarina* sequences, created an updated no *Idiomarina* taxa bar plot, and re-did the phylogenetic tree and alpha rarefaction plotting step with the no *Idiomarina* sequences and table.

1.) The code I used to visualize my Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza:
 qiime feature-table tabulate-seqs \
 --i-data '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
 --o-visualization Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qzv

2.) The code I used to visualize the Bat_2021_No_Idiomarina_No_contaminants microbiome using taxa bar plots:

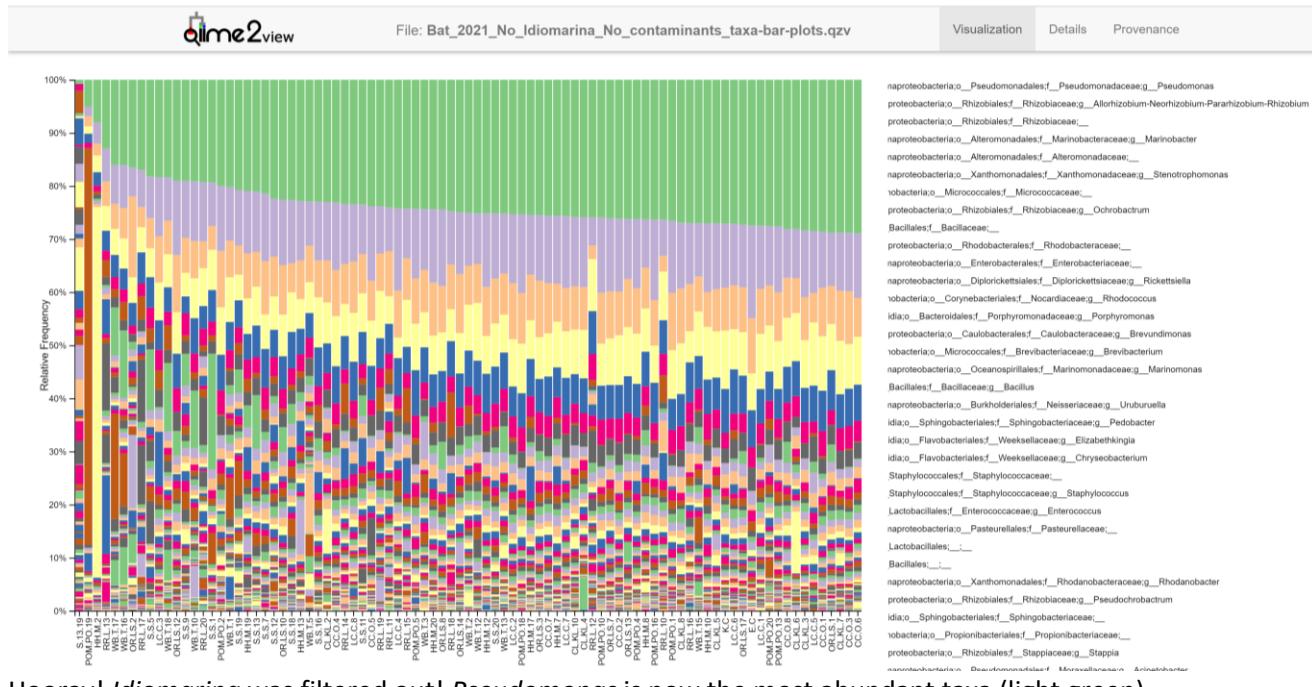
```
qiime taxa barplot \  

--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \  

--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \  

--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \  

--o-visualization Bat_2021_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv
```



Hooray! *Idiomarina* was filtered out! *Pseudomonas* is now the most abundant taxa (light green).

b.) Create phylogenetic tree and rarefy sequence data – continued from part 1

1.) The code I used to build a phylogenetic tree using my No_Idiomarina_No_contaminants sequences:

```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_No_Idiomarina_No_Contaminants_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_No_Idiomarina_No_contaminants_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_No_Idiomarina_No_contaminants_unrooted-tree.qza \
--o-rooted-tree Bat_2021_No_Idiomarina_No_contaminants_rooted-tree.qza
```

2.) The code I used to replot my alpha rarefaction using my No_Idiomarina_No_contaminants table:

```
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-phylogeny '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--p-max-depth 4032 \← new value from No_Idiomarina table \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_Idiomarina_No_contaminants_alpha-rarefaction.qzv
```

Where should we rarefy our sequences?

- From **No_Idiomarina alpha rarefaction plots**: 2,000 – 2,150 appears to be a decent cutoff point, with no huge jumps in these areas. There doesn't appear to be a huge difference between 2,000 or 2,150, so it would probably be good to go for the higher sequencing depth of **2,150** like we were thinking about for the other versions of the unfiltered alpha rarefaction plots.
 - It looks like the sequences continue to level out more with each filtering step. If we decided to cut off at 2,450 this would remove a sample of importance. As we continue

to filter these sequences, more samples are up for cutting as we figure out where to rarefy them.

- **Samples up for cutting from No_Idiomarina table:**

- **HH.M.14:** one of the common species and sex from that site (little brown, female), no ectoparasites. Probably OK to cut.
- **CL.KL.9:** we need to keep this one, has ectoparasites, ectoparasite presence not as common so we need to keep every ectoparasite bat we can (see note about CL.KL.4).
- **S.S7:** comparable to other samples from this site (all yuma myotis, all females, all no ectoparasites).
- **WB.T.5:** comparable to other samples from this site (most yuma myotis females, no ectoparasites).
- **CL.KL.4:** comparable to other samples from this site (all big brown bats, all females, most had ectoparasites), but ectoparasite bats in general were not as abundant overall so we should try to keep every ectoparasite sample that we can. However, it would probably be OK to cut just one ectoparasite bat to increase our sampling depth to 2,150 (otherwise we would be stuck at 1,500 which would not be desirable), but I would not want to remove any more ectoparasite samples.
- **LC.C.5:** comparable to other samples from this site (all female little brown bats with no ectoparasites).

Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qzv:

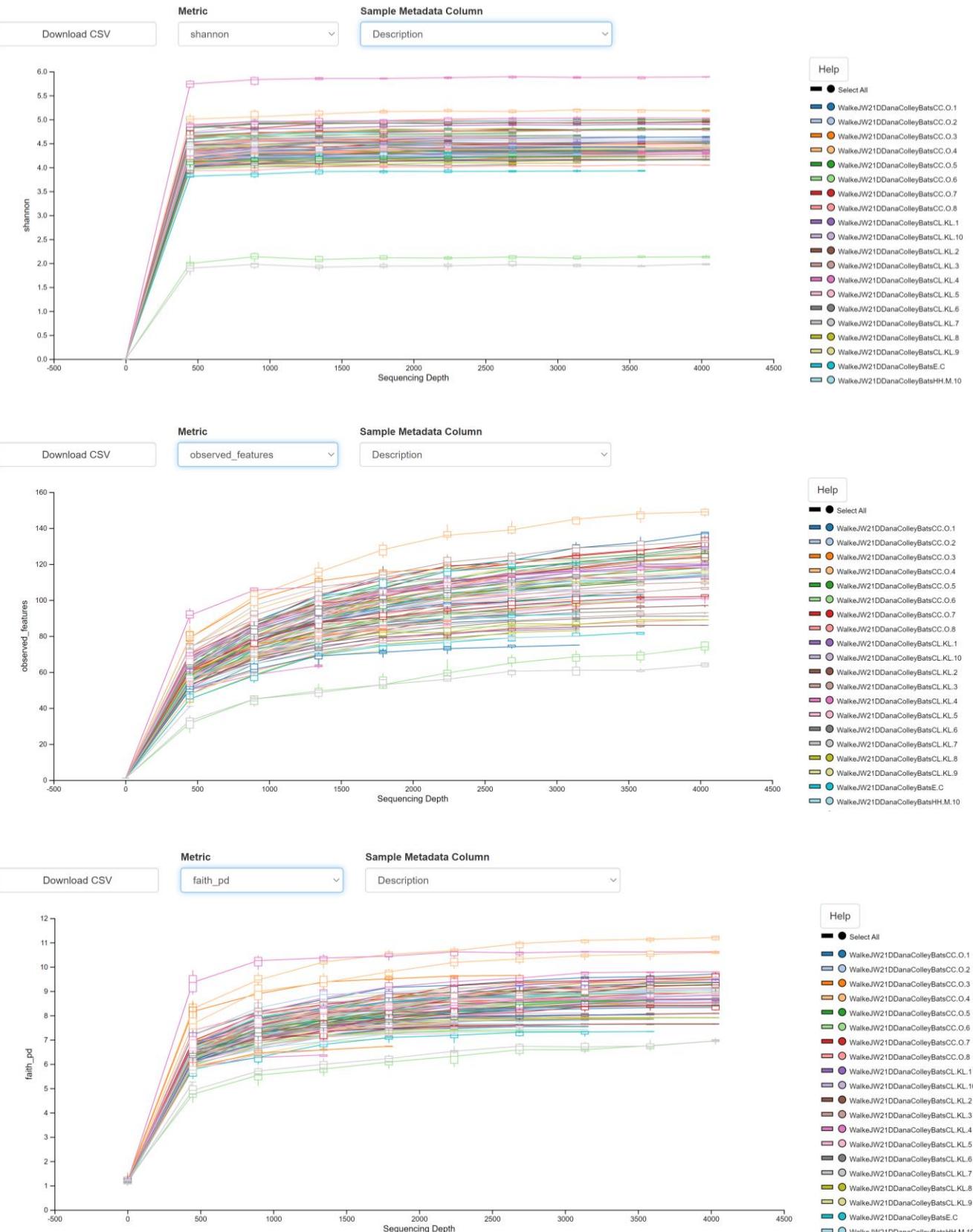
Qlime2 view		File: Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qzv
RR.L.15		2895
RR.L.12		2890
CL.KL.6		2856
CC.O.1		2845
S.S.11		2816
WB.T.1		2726
WB.T.2		2693
WB.T.12		2658
RR.L.19		2627
RR.L.20		2552
S.S.16		2495
S.S.1		2453
HH.M.14		2378
CL.KL.9		2156
S.S.7		2147
WB.T.5		1676
CL.KL.4		1603
LC.C.5		488

2,450 would
remove these
samples

2,150 would
remove these
samples

2,000 would
remove these
samples

From Bat_2021_No_Idiomarina_No_contaminants alpha rarefaction plots:



Update: last bout of filtering contaminants. I filtered out *Marinobacter*, *Marinomonas*, *Salinisphaera*, and *Salinarimonas* from the No_Idiomarina table and sequences. I remade the phylogenetic tree and alpha rarefaction plots using the updated sequences and table.

List of contaminants filtered out of our sequence data:

- *Pseudoalteromonas*
- *Vibrio*
- *Halomonas*
- *Idiomarina*
- *Marinobacter*
- *Marinomonas*
- *Salinisphaera*
- *Salinarimonas*

1.) The code I used to filter out the remaining contaminants from my table:

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinobacter \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \ 
← output from previous step
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinomonas \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinisphaera \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

```
qiime taxa filter-table \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinarimonas \
--o-filtered-table Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

2.) The code I used to visualize the final no contaminants table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
-o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv --m-sample-
metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

3.) The code I used to filter out the remaining contaminants from my sequences:

```
qiime taxa filter-seqs \
--i-sequences '/media/sf_QIIME/QIIME/Bat_2021_No_Idiomarina_No_contaminants_filtered-rep-
seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinobacter \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

```
qiime taxa filter-seqs \
--i-sequences
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-
seqs.qza' \ ← output from previous step
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Marinomonas \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

```
qiime taxa filter-seqs \
--i-sequences
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-
seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinisphaera \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

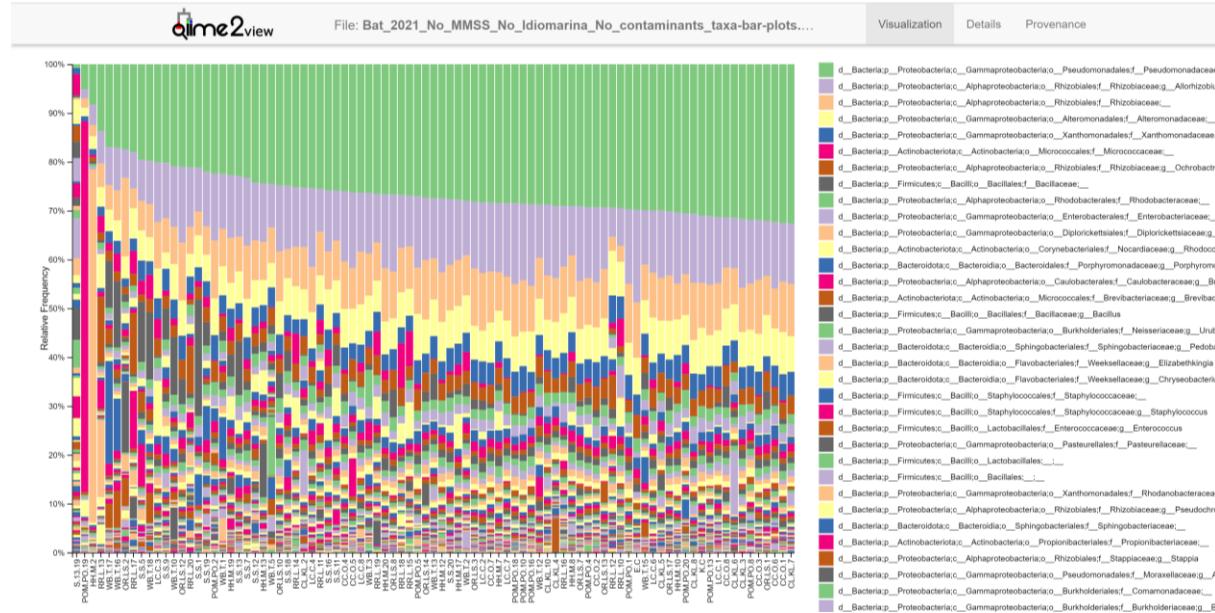
```
qiime taxa filter-seqs \
--i-sequences
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-
seqs.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-exclude Salinarimonas \
--o-filtered-sequences Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza
```

**4.) The code I used to visualize my Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
rep-seqs.qza:**

```
qiime feature-table tabulate-seqs \
--i-data '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-
seqs.qza' \
-o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qzv
```

5.) The code I used to visualize the Bat_2021No_MMSS_No_Idiomarina_No_contaminants microbiome using taxa bar plots:

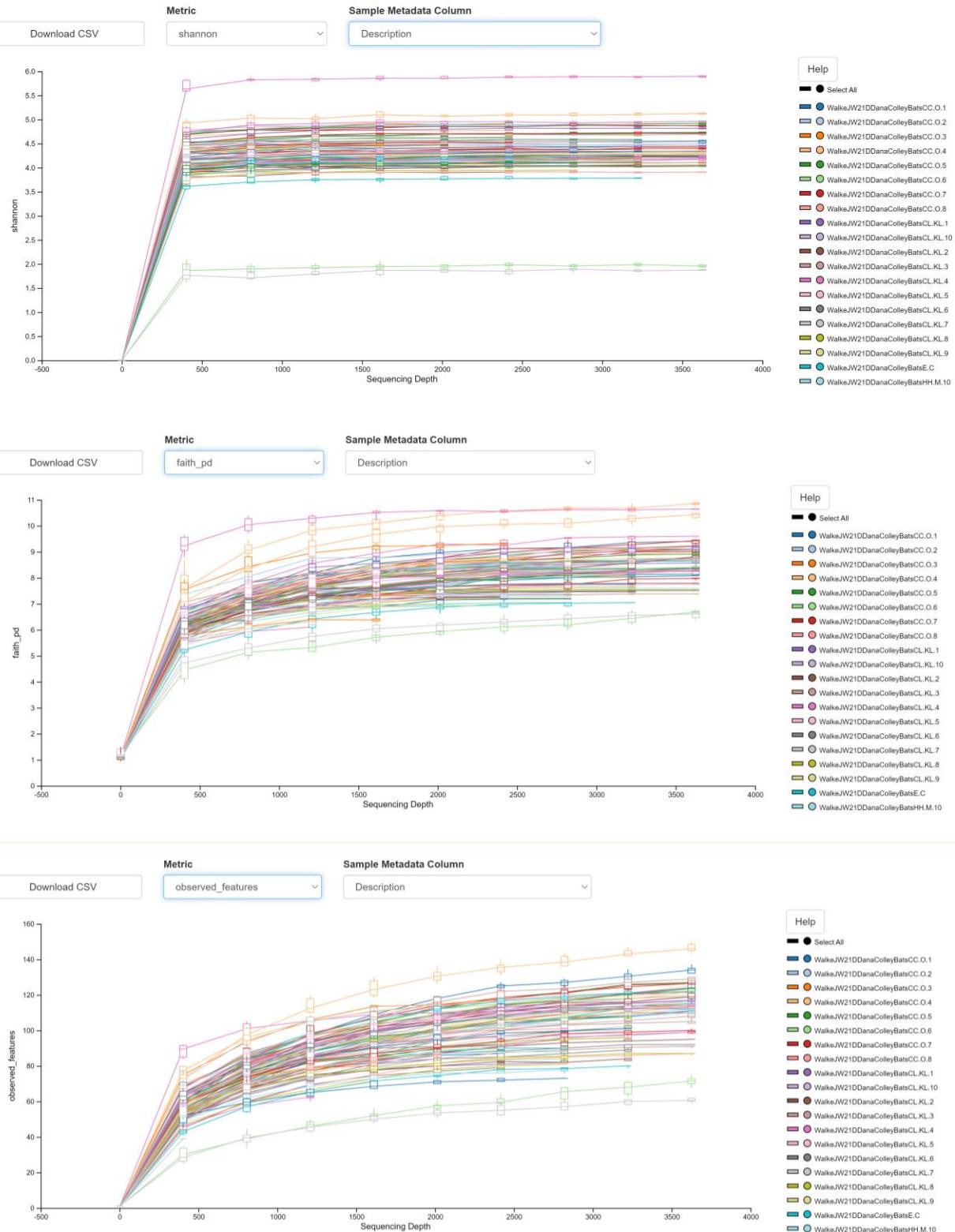
```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv
```



```
qiime phylogeny align-to-tree-mafft-fasttree \
--i-sequences \
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-rep-seqs.qza' \
--o-alignment Bat_2021_No_MMSS_No_Idiomarina_No_Contaminants_aligned-rep-seqs.qza \
--o-masked-alignment Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_masked-aligned-rep-seqs.qza \
--o-tree Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_unrooted-tree.qza \
--o-rooted-tree Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza
```

7.) The code I used to replot my alpha rarefaction using my No_MMSS table:

```
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--i-phylogeny \
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--p-max-depth 3626 \
--m-metadata-file '/media/sf_QIIME/QIIME/Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_alpha-rarefaction.qzv
```



qiime2view
File: Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv

RR.L.15	2608
S.S.11	2559
CL.KL.6	2553
RR.L.12	2553
CC.O.1	2505
WB.T.1	2450
WB.T.2	2410
RR.L.19	2357
RR.L.20	2334
WB.T.12	2325
S.S.1	2242
HH.M.14	2217
S.S.16	2216
S.S.7	1976
CL.KL.9	1888
WB.T.5	1569
CL.KL.4	1430
LC.C.5	425

1,750 or 1,800
would remove
these samples

3/20/2022: I made a new table, alpha rarefaction, and taxa bar plots using my updated metadata file with the species column included. I forgot to include species so these new plots have species included.

1.) With_spp table:

```
qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file
'/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

2.) With_spp taxa bar plots:

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv
```

3.) With_spp alpha rarefaction:

```
qiime diversity alpha-rarefaction \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--p-max-depth 3626 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - -
Sheet1.tsv' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_alpha-
rarefaction.qzv
```

4/5/2022: I removed the control samples (E.C, S.13.19, K.C) from the dataset and I ran the alpha and beta diversity analyses. You had to specify the column you wanted to analyze. In our case, all of the controls had "NA" listed for the Species, so we could remove those samples. I visualized all of the alpha diversity stats results.

Removing control samples from the table:

Should everything be lumped together

```
qiime feature-table filter-samples \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - -
Sheet1.tsv' \
--p-where "[Species]='NA'" \
--p-exclude-ids \
--o-filtered-table Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza
```

Visualizing the table without the controls: check to see that the samples were filtered out

```
qiime feature-table summarize \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
--o-visualization Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qzv \
--m-sample-metadata-file
'/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

→ Use No_control table from now on

c.) Actual data analyses – alpha and beta diversity

1.) Calculating alpha and beta diversity of your samples: → no controls

Filter data table by species, site level differences within a species

```
qiime diversity core-metrics-phylogenetic \
```

```
--i-phylogeny <file-name_rooted-tree.qza from phylog. tree step> \
--i-table <file-name_filtered-table.qza from table filter step> \
--p-sampling-depth <sequence depth number determined from alpha rarefaction step> \
--m-metadata-file <the mapping/metadata file you used earlier> \
--o-rarefied-table <file-name_rarefied-table.qza> \
--output-dir <file-name_core-metrics-results>
```

The code I used to calculate alpha and beta diversity:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_rarefied-table.qza --output-dir Bat_2021_core-metrics-results
```

2.) Calculate alpha diversity statistics with nonparametric Kruskal-Wallis tests:

Make sure to transfer all stat data to a single excel file for each metric.

Shannon diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity <file-name_core-metrics-results/shannon_vector.qza from diversity metrics folder in
step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv > \
--o-visualization <file-name_core-metrics-results/shannon_significance.qzv>
```

The code I used to visualize Shannon's diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/shannon_vector.qza' /
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/shannon_significance.qzv
```

→ Note: the output goes to the core-metrics-results folder

Observed OTU's:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity <file-name_core-metrics-results/observed_features_vector.qza from diversity metrics
folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_core-metrics-results/observed_features_vector_significance.qzv>
```

The code I used to visualize Observed OTU's:

```
qiime diversity alpha-group-significance \
```

```
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/observed_features_vectors_significance.qzv

Faith's Phylogenetic Diversity:
qiime diversity alpha-group-significance
--i-alpha-diversity <file-name_core-metrics-results/faith_pd_vector.qza from diversity metrics folder in
step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_core-metrics-results/faith_pd_significance.qzv>
```

The code I used to visualize Faith's Phylogenetic Diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance
--i-alpha-diversity <file-name_core-metrics-results/evenness_vector.qza from diversity metrics folder in
step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--o-visualization <file-name_core-metrics-results/evenness_significance.qzv>
```

The code I used to visualize Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/evenness_significance.qzv
```

Discrete: different test

3.) Calculate beta diversity statistics using nonparametric PERMANOVAs:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix <file-name_core-metrics-results/bray_curtis_distance_matrix.qza from diversity
metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever "treatment" you're comparing diversity values between> \
--o-visualization <file-name_core-metrics-results/bray_curtis_site_significance.qzv>
--p-pairwise
```

The code I used to visualize Bray-Curtis diversity: all variables included

```
qiime diversity beta-group-significance \
```

```
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix <file-name_core-metrics-results/jaccard_distance_matrix.qza from diversity metrics
folder in step #1>
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., ectoparasites> \
--o-visualization <file-name_core-metrics-results/jaccard_site_significance.qzv>
--p-pairwise
```

The code I used to visualize Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix <file-name_core-metrics-results/unweighted_unifrac_distance_matrix.qza from
diversity metrics folder in step #1>
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., ectoparasites> \
--o-visualization <file-name_core-metrics-results/unweighted_unifrac_site_significance.qzv>
--p-pairwise
```

The code I used to visualize unweighted UniFrac:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance
```

```
--i-distance-matrix <file-name_core-metrics-results/weighted_unifrac_distance_matrix.qza from
diversity metrics folder in step #1>
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., ectoparasites> \
--o-visualization <file-name_core-metrics-results/weighted_unifrac_site_significance.qzv>
--p-pairwise
```

The code I used to visualize weighted UniFrac:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

4/10/2022: I ran alpha and beta diversity analyses on data tables with just one type of species to see if ectoparasite presence does have an influence within the species level. Ectoparasite presence appears currently to not influence the microbiome diversity.

MYYU-only data table

```
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--p-where "[Species]='MYYU'" \
--o-filtered-table
Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

Visualize the MYYU-only table to check that only MYYU included

```
qiime feature-table summarize \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contamin-
ants_filtered-table.qza' \
--o-visualization
Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qzv \
--m-sample-metadata-file
'/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv'
```

MYLU-only data table

```
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza' \
```

```
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--p-where "[Species]='MYLU'" \
--o-filtered-table
Bat_2021_MYLU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

MYVO-only data table

```
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--p-where "[Species]='MYVO'" \
--o-filtered-table
Bat_2021_MYVO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

MYYU/LU-only data table

```
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--p-where "[Species]='MYYU/LU'" \
--o-filtered-table
Bat_2021_MYYU_LU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza
```

EPFU-only data table

```
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-
-table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--p-where "[Species]='EPFU'" \
--o-filtered-table
Bat_2021_EPFU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

Calculate alpha and beta diversity for MYYU-only data table:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
```

```
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYYU_only_rarefied-table.qza \
--output-dir Bat_2021_MYYU_only_core-metrics-results
```

Visualize MYYU-only alpha diversity results:

Shannon:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/shannon_vector.qza' / \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics- results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics- results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics- results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYYU-only beta diversity results:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics- results/bray_curtis_distance_matrix.qza' \
```

```
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-
results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_only_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

Calculate alpha and beta diversity for MYLU-only data table:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_No_controls_No_MMSS_No_Idiomarina_No_contamin-
ants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYLU_only_rarefied-table.qza \
```

```
--output-dir Bat_2021_MYLU_only_core-metrics-results
```

Visualize MYLU-only alpha diversity results:

Shannon:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-
results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYLU-only beta diversity results:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

Calculate alpha and beta diversity for MYVO-only data table:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_No_controls_No_MMSS_No_Idiomarina_No_contami-
nants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYVO_only_rarefied-table.qza \
--output-dir Bat_2021_MYVO_only_core-metrics-results
```

Visualize MYVO-only alpha diversity results:**Shannon:**

```
qiime diversity alpha-group-significance \
```

```
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-
results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYVO-only beta diversity results:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/jaccard_distance_matrix.qza' \
```

```
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

Calculate alpha and beta diversity for MYYU/LU-only data table:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_No_controls_No_MMSS_No_Idiomarina_No_conta
minants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_MYYU_LU_only_rarefied-table.qza \
--output-dir Bat_2021_MYYU_LU_only_core-metrics-results
```

Visualize MYYU/LU-only alpha diversity results:

Shannon:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-
results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYVO-only beta diversity results:**Bray-Curtis:**

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-
results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYYU_LU_only_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_MYYU_LU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

Calculate alpha and beta diversity for EPFU-only data table:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_No_controls_No_MMSS_No_Idiomarina_No_contamin-
ants_filtered-table.qza' \
--p-sampling-depth 1850 \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_EPFU_only_rarefied-table.qza \
--output-dir Bat_2021_EPFU_only_core-metrics-results
```

Visualize EPFU-only alpha diversity results:**Shannon:**

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/shannon_significance.qzv
```

Observed features:

```
qiime diversity alpha-group-significance \
```

```
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/observed_features_vectors_significance.qzv
```

Faith's phylogenetic diversity:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/faith_pd_significance.qzv
```

Evenness:

```
qiime diversity alpha-group-significance \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/evenness_significance.qzv
```

Visualize MYVO-only beta diversity results:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

Jaccard:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/jaccard_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/jaccard_site_significance.qzv \
--p-pairwise
```

Unweighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/unweighted_unifrac_distance_matrix.qza'
```

```
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/unweighted_unifrac_site_significance.qzv \
--p-pairwise
```

Weighted UniFrac Distances:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_EPFU_only_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_EPFU_only_core-metrics-results/weighted_unifrac_site_significance.qzv \
--p-pairwise
```

4/13/2022: I ran correlation tests on my continuous variables (weight, forearm length, ear length, tragus length) for both alpha diversity (Spearman's rank correlation) and beta diversity (Mantel test). I ran beta diversity analyses for species and site including all variables.

Beta diversity for Species:

Bray-Curtis:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Site \
--o-visualization Bat_2021_core-metrics-results/bray_curtis_site_significance.qzv \
--p-pairwise
```

4/19/2022: I ran Spearman's rank correlations for WDI and weight for the alpha diversity test.

1.) Alpha diversity for continuous variables: Spearman's rank correlation test

Shannon:

```
qiime diversity alpha-correlation \
--i-alpha-diversity <file_name_core-metrics-results/shannon_vector.qza> \
--m-metadata-file <the mapping/metadata file you've been using.tsv> \
--o-visualization <Bat_2021_core-metrics-results/shannon_correlation_Spearman.qzv>
```

```
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/shannon_correlation_Spearman.qzv
```

Faith's Phylogenetic Diversity:

```
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/faith_pd_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/faith_pd_correlation_Spearman.qzv
```

Observed features:

```
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/observed_features_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/observed_features_correlation_Spearman.qzv
```

Evenness:

```
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/evenness_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--o-visualization Bat_2021_core-metrics-results/evenness_correlation_Spearman.qzv
```

Repeat for species-only comparisons**2.) Calculate beta diversity statistics using Mantel tests:****Bray-Curtis:**

```
qiime diversity beta-correlation \
--i-distance-matrix <file-name_core-metrics-results/bray_curtis_distance_matrix.qza from diversity metrics folder in step #1> \
--m-metadata-file <the mapping/metadata file you used earlier.tsv> \
--m-metadata-column <whatever variable you're interested in, i.e., Ectoparasites> \
--p-intersect-ids \
--o-metadata-distance-matrix <file-name_core-metrics-results/bray_curtis_correlation.qza> \
--o-mantel-scatter-visualization <file-name_core-metrics-results/bray_curtis_correlation.qzv>
```

Repeat for Jaccard similarity, unweighted UniFrac Distances, weighted UniFrac Distances:**3.) Beta diversity PERMANOVA analyses:****For species:**

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Species.qzv \
--p-pairwise
```

Investigating how bat ectoparasites influence the skin microbiome diversity and composition in Washington state bats – part 3

The objectives of this study are 1.) to investigate whether ectoparasite infestation in bats influences the skin microbiome diversity of bats, and 2.) to isolate and identify culturable bacteria from the skin of bats to compare their relative abundances sequencing data.

We hypothesized that bats with ectoparasites will have decreased skin microbiome diversity and altered composition compared to bats without ectoparasites, placing bats at a higher risk of Pd infection. Since culture methods are highly selective and many bacteria in a microbiome are unable to be cultured, we also hypothesized that the most abundant culturable bacterial isolates from western bats will differ from the most abundant bacteria in our sequencing data.

This sequencing data comes from maternity colonies sampled across Washington state from April – June 2021 with the WDFW.

Part II – Analyses on your sequence data – continued from part 2 of my bioinformatics notebook

c.) Actual data analyses – alpha and beta diversity

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Ectoparasites.qzv \
--p-pairwise

qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_slash_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Species.qzv \
--p-pairwise

qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--o-visualization Bat_2021_core-metrics-results/bray-curtis_significance_Bat_Ectoparasites.qzv \
--p-pairwise
```

4/24/2022: I ran beta diversity analyses on all of the individual species comparisons. I ran the rest of the Spearman's rank correlations for the species-specific tables for alpha diversity.

1.) Calculate beta diversity for site within each species:

Beta diversity among site in MYLU only: repeat for other beta diversity metrics and other species

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_MYLU_only_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--m-metadata-column Site \
--o-visualization Bat_2021_MYLU_only_core-metrics-results/bray_curtis_site_significance_Site.qzv \
--p-pairwise
```

2.) Repeat for other alpha diversity metrics and other species

```
qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_MYVO_only_core-metrics-
results/shannon_vector.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_MYVO_only_core-metrics-results/shannon_correlation_Spearman.qzv
```

2.) Calculate beta diversity statistics using Mantel tests:

Bray-Curtis:

```
qiime diversity beta-correlation \
--i-distance-matrix <file-name_core-metrics-results/bray_curtis_distance_matrix.qza from diversity
metrics folder in step #1> \
--m-metadata-file <whatever variable you're interested in> \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-
results/bray_curtis_Ectoparasites_correlation.qzv
```

```
qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_Fil
e - Sheet1.tsv' \
--m-metadata-column Weight \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/bray_curtis_Weight_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/jaccard_distance_matrix.qza'
```

```
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_Fil
e - Sheet1.tsv' \
--m-metadata-column WDI \
--p-intersect-ids --o-metadata-distance-matrix Bat_2021_core-metrics-
results/jaccard_WDI_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/jaccard_WDI_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/unweighted_unifrac_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_Fil
e - Sheet1.tsv' \
--m-metadata-column WDI \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/unweighted_unifrac_WDI_correlation.qza \
/
--o-mantel-scatter-visualization Bat_2021_core-metrics-
results/unweighted_unifrac_WDI_correlation.qzv

qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_Fil
e - Sheet1.tsv' \
--m-metadata-column WDI \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_core-metrics-results/weighted_unifrac_WDI_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_core-metrics-results/weighted_unifrac_WDI_correlation.qzv
```

I had to remove the control samples from my metadata file, QIIME didn't like those samples because they had NAs for many of the values

Repeat for other continuous variables (WDI), for other beta diversity metrics, for species specific chunks

3.) Beta diversity PERMANOVA analyses:

For species:

For site:

```
qiime diversity beta-group-significance \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
```

```
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--o-visualization Bat_2021_core-metrics-results/weighted_unifrac_site_significance_Species.qzv \
--p-pairwise
```

5/4/2022: I ran LEfSe analyses (relative abundance comparisons) on the significant comparisons to see which bacterial taxa were actually driving the significant differences between the different groups (i.e., different sites, different species). I also ran a LEfSe analysis on ectoparasites just to see if any bacterial taxa differ from each other, even though the overall relationships were non-significant.

Update: LEfSe analysis did not work on the Harvard website, so I had to run a similar analysis (indicspecies) in R. See R Notebook for details.

LEfSe Step-by-Step:

1.) Calculate relative frequency for a collapsed table (genus in example at level 6, repeat for other levels like phylum, family, etc.):

```
qiime taxa collapse \\ → groups bacteria of particular taxonomy (i.e., genus)
--i-table <file_name_rarefied-table.qza from previous analyses, make sure that it's the rarefied table that was produced by the alpha/beta diversity analyses> \
--i-taxonomy <file_name_taxonomy.qza from previous analyses> \
--p-level 6 \\ → genus level here, refer back to the taxonomy bar plots for what the other levels are
--o-collapsed-table <file_name_filtered-table-l6.qza>
```

```
qiime feature-table relative-frequency \
--i-table <file_name_filtered-table-l6.qza from previous code chunk> \
--o-relative-frequency-table <file_name_frequency-table-l6.qza>
```

Exporting straight ASV table, non-collapsed, rarefied = do not include taxonomy header → total ASV, individual ASVs for the non-collapsed

The code I used to calculate relative frequency for a collapsed table:

```
qiime taxa collapse \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--p-level 6 \
--o-collapsed-table Bat_2021_filtered-table-l6.qza
```

```
qiime feature-table relative-frequency \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_filtered-table-l6.qza' \
--o-relative-frequency-table Bat_2021_frequency-table-l6.qza
```

2.) Export biom file from QIIME: variations of this code will work to export other data from QIIME as well for other uses besides LEfSe analyses

```
qiime tools export \
--input-path <file_name_frequency-table-l6.qza from previous step> \
--output-path <file_name_lefse-files>
```

```
qiime tools export \
--input-path '/media/sf_QIIME/QIIME/Bat_2021_frequency-table-l6.qza' \
--output-path Bat_2021_lefse-files
```

→ This exports a new folder (i.e., here the folder is titled “Bat_2021_lefse-files” in your directory

3.) Convert biom to text file (for LEfSe comparison): biom files from QIIME need to be converted to compatible file types for most other uses as well.

biom convert \ → for feature tables, asv table, species table, genus table → at different levels, genus/phylum/.../individual ASV is non-collapsed one

```
--input-fp <file_name_lefse-files/feature-table.biom> \
--output-fp <file_name_lefse-files/frequency-table-l6.txt> \
--header-key "taxonomy" / → works for general QIIME format, column with taxonomy indicated
--to-tsv
```

The code I used to convert the biom file to a text file:

```
biom convert \
--input-fp Bat_2021_lefse-files/feature-table.biom \
--output-fp Bat_2021_lefse-files/Bat_2021_frequency-table-l6.txt \
--header-key "taxonomy" \
--to-tsv
```

→ converted text file will be outputted in the folder you just made in your directory (i.e., “Bat_2021_lefse-files”), but the typical green “exported or saved” output after a successful code run does not show up in this step, so don’t panic. Just check the new folder you made for the new text file

```
qiime tools export --input-path --output-path --header-key "xx" --to-tsv
```

```
biom convert --input-fp --output-fp
```

4.) Edit the text file in Excel so you can use it for LEfSe:

How to open a text file in Excel and convert it to an Excel workbook:

- Open a new Excel workbook
- Data tab → “get and transform data” section on far left
- Select “from text/CSV” option in the get data section
- Find your text file → import → load
- Save your Excel workbook

Editing the text file in Excel for LEfSe analyses:

- Delete #Constructed from biom file row
- Add Ectoparasites as the new 1st row → whatever variable you are interested, check your mapping file to see what you want to do (i.e., Ectoparasites, Site, Species, Weight, etc.)
 - Your Ectoparasite row needs to have the data for each individual (i.e., ectoparasite present/absent for each bat), transpose data if necessary (number of rows needs to match the number of columns → paste special → more options → transpose)
- Replace #OTU ID with SampleID

→ This edited Excel file will be used in the next step.

Column1	Column2	Column3	Column4	Column5	Column6	Column7
Ectoparasites	Y	Y	N	Y	Y	Y
SampleID	POM.PO.1	WB.T.13	S.S.12	CC.O.7	WB.T.12	CL.KL5
d_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Bacillaceae;	0.0016216216216216215	0.055675675675675676	0.014594594594594595	0.008108108108108109	0.012972972972972972	0.001081081081081081
d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Flavobacteriaceae;0.004864864864864865	0.002162162162162162	0.0005405405405405405	0.005405405405405406	0.005945945945945946	0.005945945945945946	0.005945945945945946
d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Cytophagales;f_Hymenobacteraceae;0.0	0.0	0.0	0.0	0.0	0.0	0.0
d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;f_Caulobacterales;f_Caulobacteriaceae;0.01945945945945946	0.005945945945945946	0.016756756756756756	0.017297297297297298	0.02054054054054054	0.023243243243243242	0.023243243243243242
d_Bacteria;p_Firmicutes;c_Bacilli;o_Staphylococcales;f_Staphylococcaceae;g_S_0.0005405405405405405	0.004864864864864865	0.010810810810810811	0.0	0.04216216216216216	0.00216216216216216216	0.00216216216216216216
d_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Bacillaceae;g_Bacillus_0.016216216216216217	0.01351351351351352	0.008108108108108109	0.016216216216216217	0.012972972972972972	0.011891891891891892	0.011891891891891892
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Frankiales;f_Nakamurellaceae;0.0	0.0	0.001081081081081081	0.0	0.0	0.0	0.0
d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Devosiaeae;0.0005405405405405405	0.0005405405405405405	0.002162162162162162	0.003243243243243243	0.0005405405405405405	0.0005405405405405405	0.0005405405405405405
d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Nocardiidae;g_Nocardi;0.005945945945945946	0.0016216216216216215	0.03351351351351351	0.002702702702702703	0.0005405405405405405	0.0005405405405405405	0.0005405405405405405
d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Geodermatophilidae;g_Geodermatophil;0.0	0.0	0.01135135135135135	0.08216216216216216	0.08810810810810811	0.0	0.0
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Micrococcales;f_Dermabacterae;g_Dermabacter;0.0	0.0005405405405405405	0.002702702702703	0.0	0.002162162162162162	0.0	0.0
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Micrococcales;f_Microbacteridae;g_Microbacter;0.0	0.001081081081081081	0.0	0.0	0.002162162162162162	0.0	0.0
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Micrococcales;f_Micrococccae;g_Micrococcus;0.002162162162162162	0.002702702702703	0.015675675675675675	0.002162162162162162	0.004324324324324324	0.0	0.0
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Propionibacteriales;f_Propionibacteriidae;g_Propionibacter;0.008648648648648649	0.005405405405405406	0.0	0.0075675675675675675	0.005405405405405406	0.008108108108108109	0.008108108108108109
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Micrococcales;f_Dermatophytidae;g_Dermatophyt;0.0	0.0	0.0	0.0	0.0	0.0	0.0
d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Burkholderiales;f_Burkholderiales;g_Burkholderia;Com;0.005405405405405406	0.0037837837837837837	0.0037837837837837837	0.0037837837837837837	0.003243243243243243	0.004324324324324324	0.004324324324324324
d_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rhizobiales;f_Rhizoblace;g_Rhizoblace;0.001081081081081	0.0	0.0	0.0	0.0	0.0	0.0
d_Bacteria;p_Bacteroidota;c_Bacteroidia;o_Flavobacteriales;f_Weeksellaceae;g_Weeksellaceae;0.0	0.0	0.0	0.0	0.0	0.0	0.0005405405405405
d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomonadidae;f_Pseudomonad;0.004324324324324324	0.0037837837837837837	0.012972972972972972	0.01027027027027027	0.004864864864864865	0.00972972972972973	0.0
d_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Corynebacteriales;f_Corynebacteriidae;g_Corynebacter;0.0	0.0	0.0	0.0	0.0	0.0	0.0
d_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacteriales;f_Carnobacteriaceae;g_Carnobacter;0.003243243243243243	0.0005405405405405405	0.0016216216216216215	0.001081081081081081	0.001081081081081081	0.001081081081081081	0.001081081081081081
d_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Xanthomonadidae;f_Xo;0.0	0.0	0.0005405405405405405	0.0	0.0	0.0	0.0
d_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;f_Bacillaceae;g_Virgibacillus;0.0005405405405405405	0.0	0.0	0.0	0.001081081081081081	0.0	0.0

Running the LEfSe analyses:

→ Use the edited Excel file from the previous step:

- Follow instructions to use LEfSe on Harvard site (<http://huttenhower.sph.harvard.edu/galaxy/>)
- Defaults are OK
- For this project, there is only a class (ChytridResult) and no subclass.

LEfSe:

- Load in your data to the Galaxy/Hutlab website:
 - Load data in the “history” pane on the right side of the page
- Format data for LEfSe (part A)
 - Data listed in **columns**
 - All other defaults OK

3. Couldn't get the Lefse analysis to work, so we did an indic species analysis in R instead. See R code notebook for details.

5.) Create taxa bar plots for specific comparisons as necessary to go with the LEfSE analysis results qiime taxa barplot \

Remember, something like this:

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_filtered-
table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-visualization Bat_2021_With_spp_No_MMSS_No_Idiomarina_No_contaminants_taxa-bar-plots.qzv
```

How to export data files in QIIME:

```
qiime tools export \
--input-path '/media/sf_QIIME/QIIME/Bat_2021_core-metrics-
results/weighted_unifrac_distance_matrix.qza' \
--output-path Bat_2021_weighted_unifrac_distance_matrix-files
```

Making grouped taxa bar plots: we had to make fake metadata files for each grouped bar plot (see photo below). The fake metadata must include all group options (i.e., six different samples for bat species, each sample represents one of the six different bat species). Make sure you include the sequence name, barcode sequence, linker primer sequence, and description. They don't need to correspond to the correct samples, just make sure that all options are represented.

For ectoparasites:

```
qiime feature-table group \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--p-axis 'sample' \
```

```
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Ectoparasites \
--p-mode 'mean-ceiling' \
--o-grouped-table Bat_2021_Grouped_EP_Feature_Table

qiime feature-table summarize \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_EP_Feature_Table.qza' \
--o-visualization Bat_2021_Grouped_EP_Feature_Table.qzv

qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_EP_Feature_Table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/EP_Grouped_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet5.tsv' \
--o-visualization Bat_2021_Grouped_EP_taxa-bar-plots.qzv
```

#SampleID	SequenceName	BarcodeSequence	LinkerPrimerSeq	Description
N	IL515rcbc0	AGCCCTTGGTCGC	GTGYCAGCMGCC	WalkowitzW21DDanaColleyBatsRR.L.10
Y	IL515rcbc1	TCCATACCGGAA	GTGYCAGCMGCC	WalkowitzW21DDanaColleyBatsRR.L.11

For roost location:

```
qiime feature-table group \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--p-axis 'sample' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Site \
--p-mode 'mean-ceiling' \
```

```
--o-grouped-table Bat_2021_Grouped_Site_Feature_Table.qza
```

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_Site_Feature_Table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/Site_Grouped_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Site.tsv' \
--o-visualization Bat_2021_Grouped_Site_taxa-bar-plots.qzv
```

For bat species:

```
qiime feature-table group \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_rarefied-table.qza' \
--p-axis 'sample' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--m-metadata-column Species \
--p-mode 'mean-ceiling' \
--o-grouped-table Bat_2021_Grouped_Species_Feature_Table.qza
```

```
qiime taxa barplot \
--i-table '/media/sf_QIIME/QIIME/Bat_2021_Grouped_Species_Feature_Table.qza' \
--i-taxonomy '/media/sf_QIIME/QIIME/Bat_2021_taxonomy.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/Species_Grouped_Slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_File - Species.tsv' \
--o-visualization Bat_2021_Grouped_Species_taxa-bar-plots.qzv
```

Filter table by COTO:

```
qiime feature-table filter-samples \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered -table.qza' \
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File - Sheet1.tsv' \
--p-where "[Species]='COTO'" \
--o-filtered-table
Bat_2021_COTO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza
```

Calculate alpha and beta diversity metrics for COTO:

```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny
'/media/sf_QIIME/QIIME/Bat_2021_No_MMSS_No_Idiomarina_No_contaminants_rooted-tree.qza' \
--i-table
'/media/sf_QIIME/QIIME/Bat_2021_COTO_only_No_controls_No_MMSS_No_Idiomarina_No_contaminants_filtered-table.qza' \
--p-sampling-depth 1850 \
```

```
--m-metadata-file '/media/sf_QIIME/QIIME/With_Spp_Edited_Updated_Bat_2021_Mapping_File -
Sheet1.tsv' \
--o-rarefied-table Bat_2021_COTO_only_rarefied-table.qza \
--output-dir Bat_2021_COTO_only_Final_core-metrics-results

qiime diversity alpha-correlation \
--i-alpha-diversity '/media/sf_QIIME/QIIME/Bat_2021_COTO_only_Final_core-metrics-
results/shannon_vector.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_Fil
e - Sheet1.tsv' \
--o-visualization Bat_2021_COTO_only_Final_core-metrics-
results/shannon_correlation_Spearman_Weight.qzv
```

→ Repeat correlations for other alpha diversity metrics

Beta diversity correlation via Mantel tests

```
qiime diversity beta-correlation \
--i-distance-matrix '/media/sf_QIIME/QIIME/Bat_2021_COTO_only_Final_core-metrics-
results/bray_curtis_distance_matrix.qza' \
--m-metadata-file
'/media/sf_QIIME/QIIME/No_controls_No_slashes_With_Spp_Edited_Updated_Bat_2021_Mapping_Fil
e - Sheet1.tsv' \
--m-metadata-column Weight \
--p-intersect-ids \
--o-metadata-distance-matrix Bat_2021_COTO_only_Final_core-metrics-
results/bray_curtis_Weight_correlation.qza \
--o-mantel-scatter-visualization Bat_2021_COTO_only_Final_core-metrics-
results/bray_curtis_Weight_correlation.qzv
```