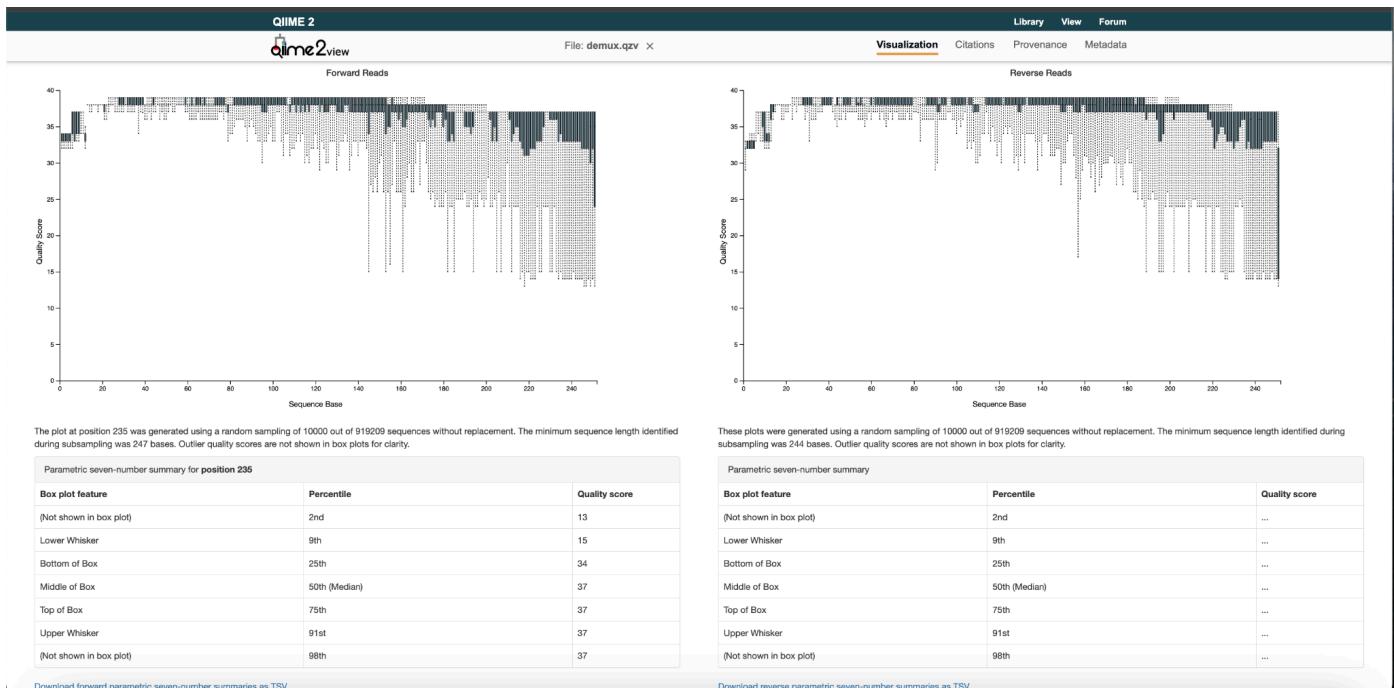


The actual numbers that I chose were too strict to demonstrate an adequate visualization. Consequently, I am utilizing the files sent by Dr. Van Laar.



When you filter the taxonomy.qzv file by taxon, you can see that there are various categories that the taxons can be classified into. These categories can include archaea and eukarya. This file is showing me the possible taxonomy that can be assigned to the taxonomy data. The taxonomy.qzv file is utilizing the sequences from the dataset and determining which organism each sequence belongs to. The table of the results is displaying the feature ID of the sequence, the taxon of the sequence, and the confidence of the identification.

I do not see that the file has reads that match things other than bacteria. Some of the reads do match unassigned sequences. Theoretically, the samples could match eukarya, archaea, and chloroplasts.

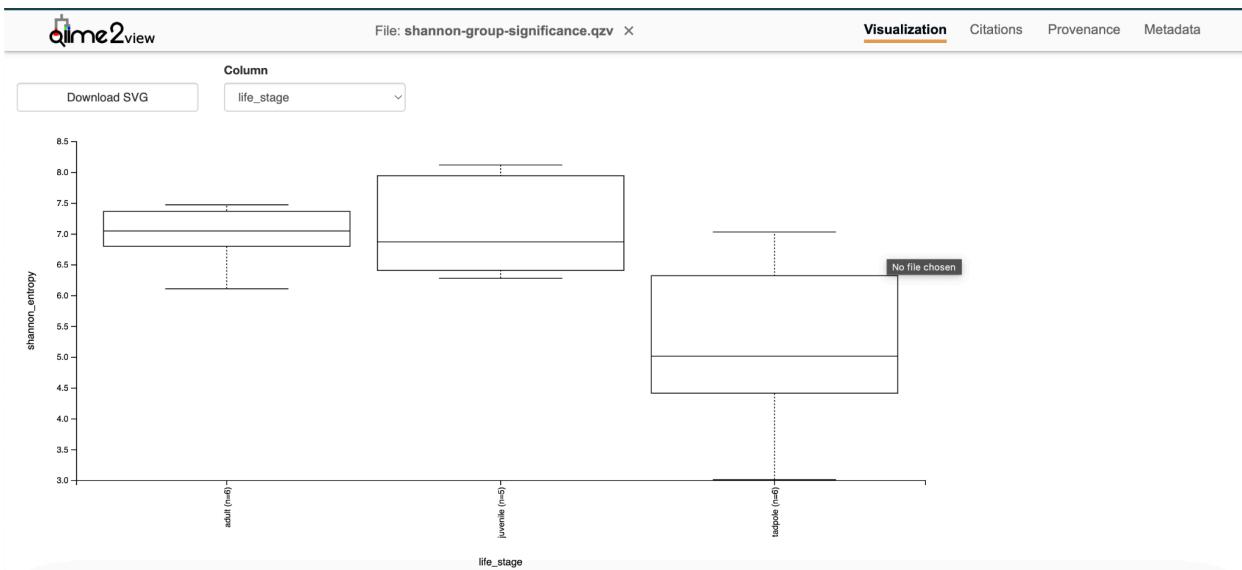
When you visualize the taxa-bar-plots.qzv file, I notice that the plot depicts the percent relative frequency for each sample. Each bar represents a different sample and each color represents a different taxa. The length of the bar depicts the relative abundance of each taxa. Changing the different levels changes the amount of taxa that are displayed in the plot. Level 1 has few colors, displaying the data on a broad taxonomic level. Contrastingly, level 7 has a multitude of colors, displaying the data on a specific taxonomic level. When you sort the data by site, the plot rearranges to display the different sites the samples were collected at, including Lagunillas and National Park. When you sort the data by life stage, the plot rearranges to display the different life stages the samples were collected at, including tadpole, adult, and juvenile.

Screenshot of observed-features-group-significance.qzv



The p-value is 0.024642862507849857, which is less than 0.05, meaning that it is significant. The only metric that was significant (that had a p-value over 0.05) was the adult juvenile sample, which had a p-value of 0.715001.

Screenshot of shannon-features-group-significance.qzv:



The p-value is 0.05452203610685121, which is greater than 0.05, meaning that there is a significant difference somewhere. The adult juvenile metric was significant, as it had a p-value of 0.855132; this is the most significant metric. Another significant metric was juvenile tadpole, which has a p-value of 0.067889.

The alpha diversity metrics are going to tell you a summary statistic for each individual sample in the table; it is going to focus on single samples. The alpha diversity provides information about the diversity in the sample. The beta diversity metrics are going to compare the samples in a pairwise fashion; it is going to focus on groups of samples. The beta diversity is going to be based on differences in microbial composition. The beta diversity provides information about how similar and different the samples are.

Both the Bray Curtis data and the Unweighted Unifrac data are beta diversity metrics. Both of these metrics are going to compare differences between microbial communities and samples. The Bray Curtis distance is going to tell you how different samples are based on the abundance of the shared sequence. The Unweighted Unifrac is going to depict how different the samples are based on the presence and absence of taxa and their evolutionary relationships. Unweighted Unifrac is going to utilize phylogenetic trees, while Bray Curtis is not going to utilize phylogenetic trees.

To change the commands to look at Unweighted Unifrac:

```
qiime diversity beta-group-significance \  
--i-distance-matrix  
diversity-metrics-results/unweighted_unifrac_distance_matrix.qza \  
--m-metadata-file metadata.txt \  
--m-metadata-column life_stage \  
--o-visualization  
diversity-metrics-results/unweighted-unifrac-life-stage-significance.  
qzv \  
--p-pairwise
```

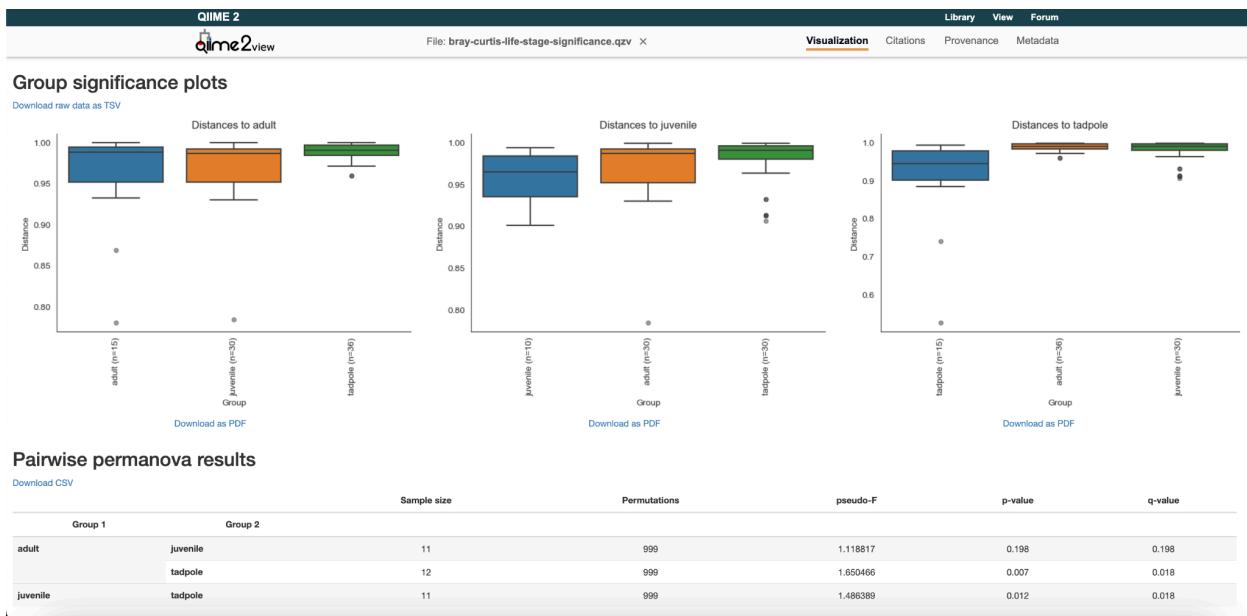
and

```
qiime diversity beta-group-significance \  
--i-distance-matrix  
diversity-metrics-results/unweighted_unifrac_distance_matrix.qza \  
--m-metadata-file metadata.txt \  
--m-metadata-column site \  
--o-visualization  
diversity-metrics-results/unweighted-unifrac-site-significance.qzv \  

```

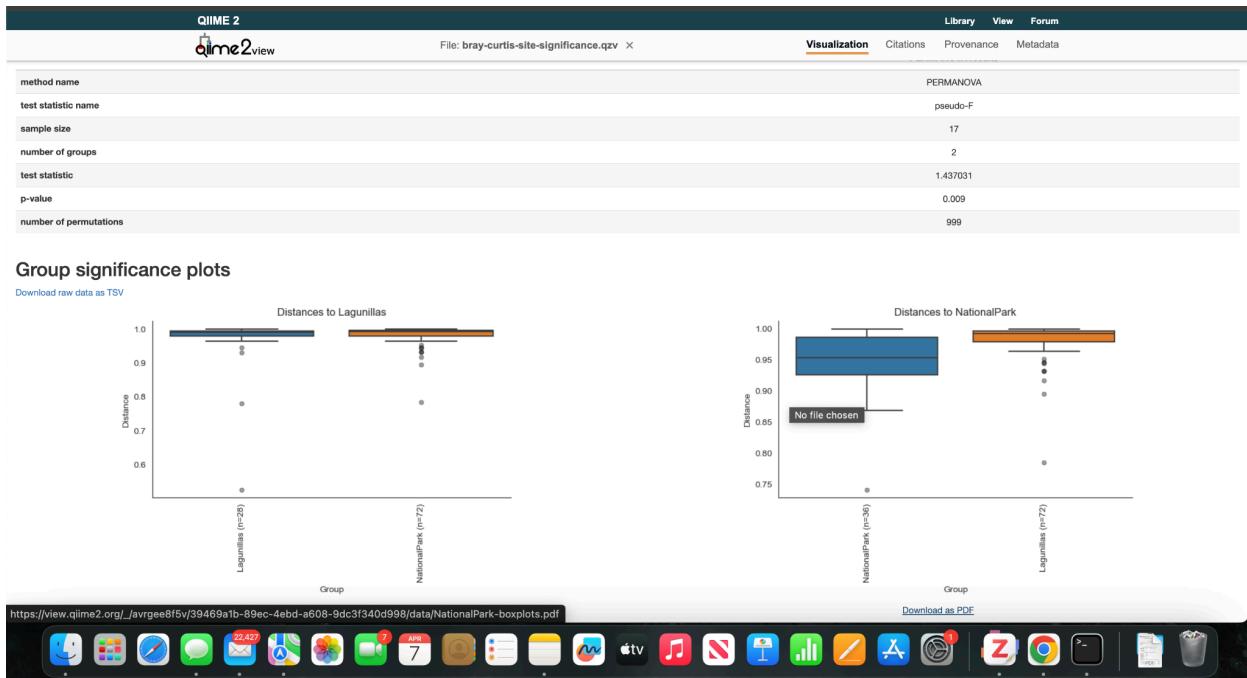
Bray Curtis pairwise comparisons:

Life stage significance:



The adult juvenile p-value is significantly higher than the adult tadpole and juvenile tadpole p-values, indicating that the sites do differ in community composition. The q-value also shows similar results to the p-value, indicating that it is significantly different compared to the adult tadpole and juvenile tadpole q-values. This implies that the life stages appear to have different community composition.

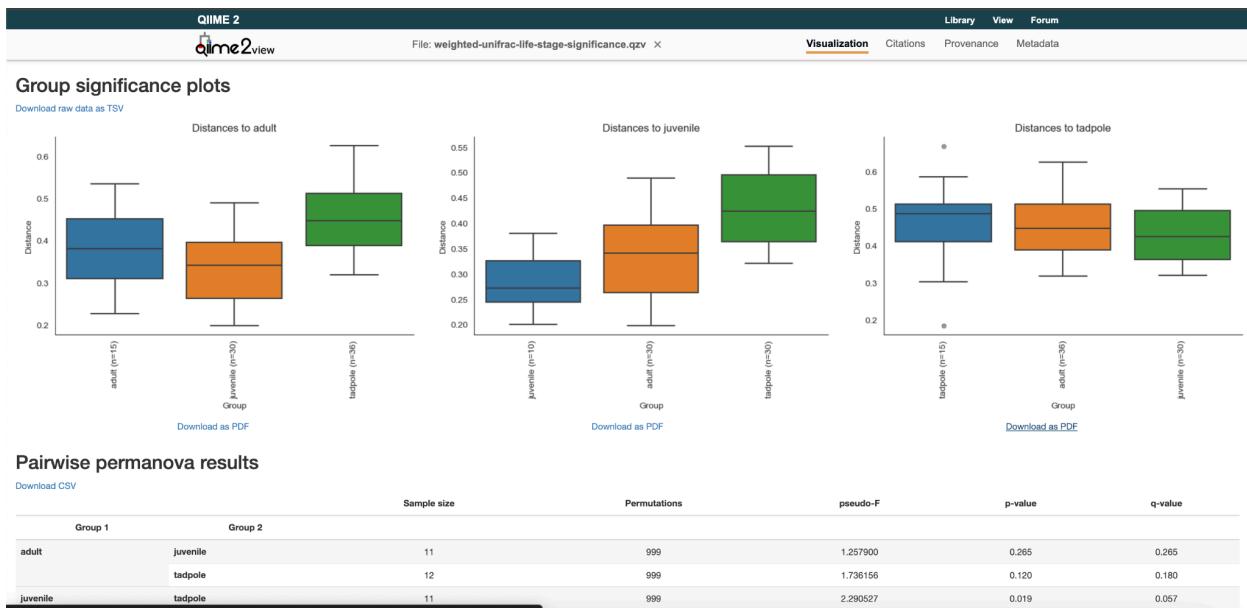
Site significance:



The p-value is 0.009, meaning that it is insignificant.

Weighted Unifrac:

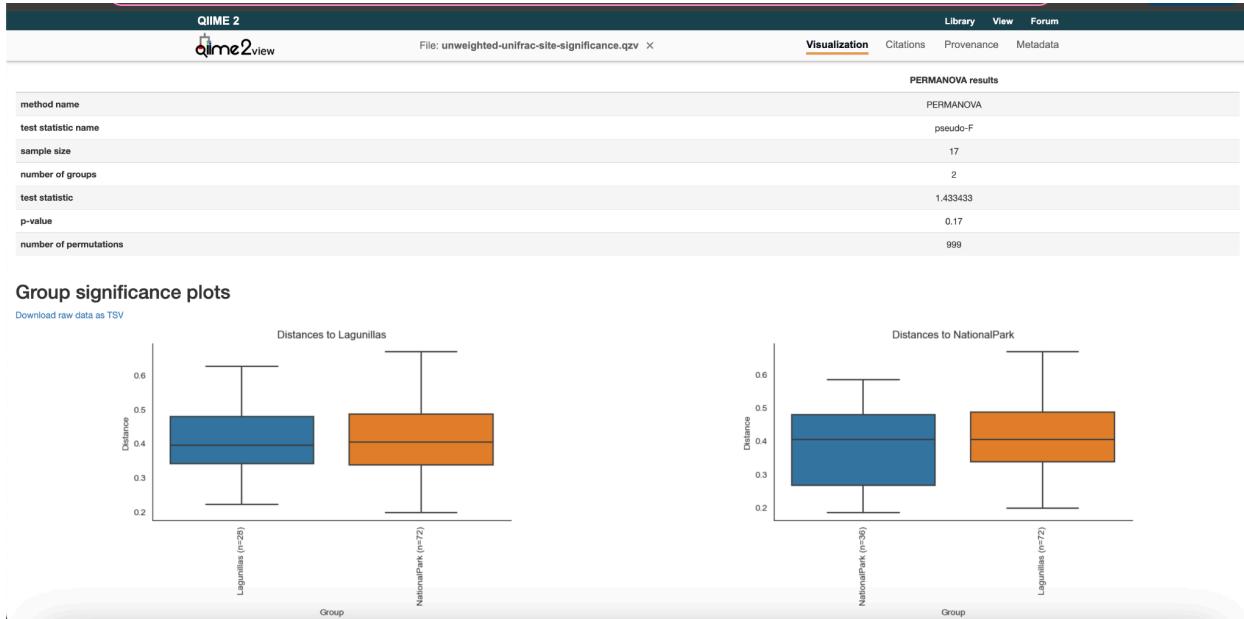
Life stage significance:



The juvenile tadpole p-value is significantly lower from the adult tadpole and adult juvenile p-values, indicating that the sites do differ in community composition. The q-value also shows that the juvenile tadpole q-value is significantly different compared to the adult tadpole and adult

juvenile tadpole q-values, indicating that the life stages appear to have different community composition. Both the p-value and the q-value for the adult juvenile is the highest metric.

Site significance:



The p-value is 0.17.

We did find taxa that were differentially expressed; we found 4 taxa.

The first taxa had a feature ID of 4890bc44aebcb26ff270bdd222f16667. This taxa that differentially expressed was in the kingdom bacteria, the phylum Gemmatimonadetes and the class Gemmatimonadetes.

The second taxa had a feature ID of 12e322c0e532976a22ea5fca1d4d9845. This taxa that was differentially expressed was in the kingdom bacteria, the phylum cyanobacteria, the class Synechococcophycideae, the order Pseudanabaenales, the family Pseudanabaenaceae, and the genus Pseudanabaena.

The third taxa had a feature ID of 204a9ca98cd8cb0b409a3b6ab13f1601. This taxa that was differentially expressed was in the kingdom bacteria and the phylum Bacteroidetes.

To repeat this analysis for life stage, you would use the code:

```
qiime composition ancombc \
    --i-table filtered-table.qza \
    --m-metadata-file metadata.txt \
```

```
--p-formula 'life_stage' \  
--o-differentials ancombc-stage.qza
```

and

```
qiime composition da-barplot \  
--i-data ancombc-stage.qza \  
--p-significance-threshold 0.05 \  
--o-visualization da-barplot-stage.qzv
```

Consequently, there were no significantly differentially expressed sites with the life stages analysis.