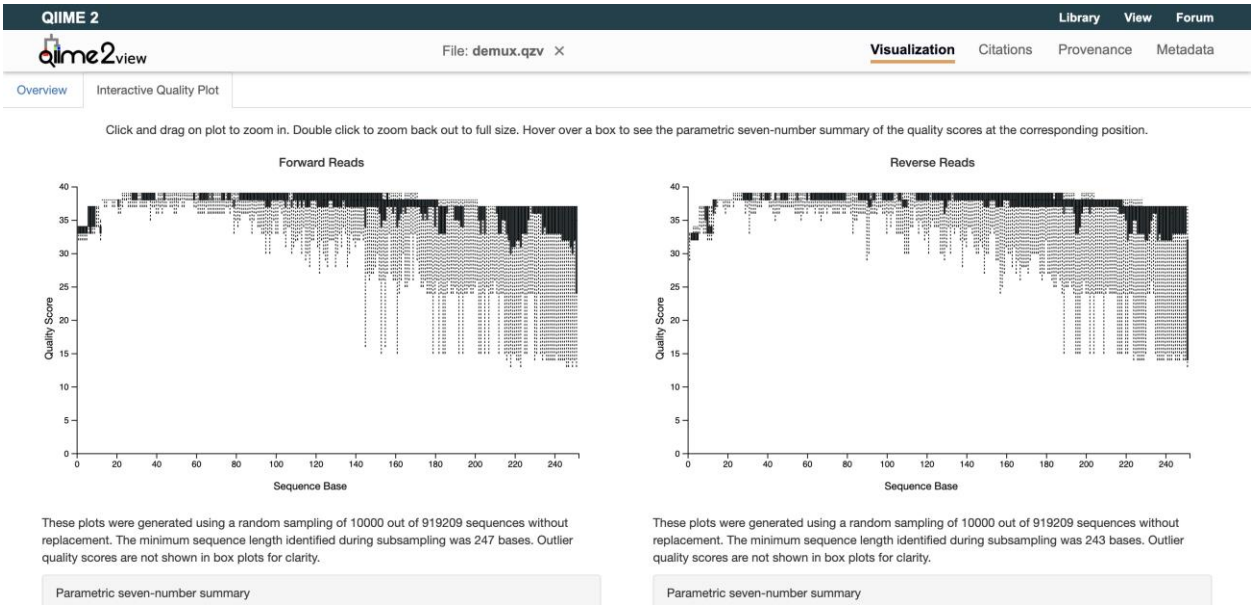
1. **Screenshot of demux.qzv**



2. **For each data set, you will need to make a judgement call. For this data set, any samples with fewer than 10,000 reads will probably need to be cut. How many samples is that?**

In this data set none of the samples were fewer than 10,000 reads.

3. **What do you see if you filter by taxon? What is this file showing you?**

When the file is filtered by taxon, I see the list is categorized by alphabetical order of the taxon group. This places the archaea species first than bacteria. This shows us the taxonomy of multiple species being compared in the file. The confidence is also provided which is the “correctness” of the data.

4. **Sometimes the file might have reads that match things other than Bacteria. Do you see that in your file? These are samples from frogs...what else, besides bacteria, could theoretically be amplified if you're amplifying 16S rRNA?**

In my file there are other matches besides Bacteria including the archaea kingdom. The mitochondria and chloroplast could theoretically be amplified when amplifying the 16S rRNA. This is because they contain their own 16s rRNA.

5. **When you visualize the taxa-bar-plots.qzv file, what do you notice? How does changing the different levels change the visualization? Why do you think this is? When you sort the samples by life stage or site of collection, do you notice any trend**

When visualizing the taxa-bar-plot I noticed the plot was all green (bacteria) when at level 1 taxonomy. When the levels are changed there is an increase in different colors. These colors represent different taxonomic levels such as Kingdom, Phylum, Class, Order, Family, Genus, and Species. I think there are different levels to represent the different taxonomies of the species being compared. By having the different taxonomic levels, the similarities and differences between the species can be compared. When it's sorted by life stage the order of the samples changes on the bar graph. The samples are now listed from adult to juvenile to tadpole.

6. **The command above runs both alpha and beta diversity metrics. There are a number of each that are performed. We will take a look at 2 for alpha, observed features, and Shannon diversity. In lecture, we talked about Observed and Shannon. In your own words, summarize briefly what each of these metrics is doing and how they are different.**

Observed features for alpha diversity summarize the structure of an ecological community by measuring its richness and evenness. Shannon diversity measures a species diversity by considering its richness and evenness. The Observed features is counting the richness and evenness for a community to determine diversity while the Shannon diversity compares the richness and evenness to each other to determine teh diversity of the system.

7. **Take a screenshot of your Observed Features for life stage and Shannon for site of collection. Were any comparisons significant for any metric? If so, which ones.**

The p-value for the Shannon site was approximately 0.386 which is greater than 0.05. This means the difference was not significant. The p-value for the observed features life stage was 0.09 whch is greater than 0.05. This means the difference was not significant.

8. **We can now look at beta diversity. How do alpha and beta diversity differ in what they are trying to tell you?**

Alpha diversity focuses on a single community/sample while beta diversity focuses on the differences between multiple communities/samples. Therefore, alpha diversity tells you about the diversity within a single community while beta diversity can tell you how communities differ from one another.

9. **Here are the commands for visualizing the Bray Curtis data. How will you change these to look at the Weighted Unifrac?**

You would change the commands to look for weighted unifrac by renaming each part of the command that has “bray curtis” to be “weighted unifrac.” Shown below.

10. **Do any life stages appear to have significantly different community composition based on either metric? Please include a screenshot of the table for both Bray Curtis and Weighted Unifrac. For the site comparison, we can just look at the p value. Do the sites differ in community composition?**

This is the table for the Bray Curtis life stage.

Pairwise permanova results

Download CSV

		Sample size	Permutations	pseudo-F	p-value	q-value
Group 1	Group 2					
adult	juvenile	11	999	0.998771	1.000	1.000
	tadpole	12	999	1.004292	0.466	0.699
juvenile	tadpole	11	999	1.004046	0.459	0.699

https://view.qiime2.org/_/butuneot644/04bcd723-05f2-475b-8b8f-c29fb12d6118/data/tadpole-boxplots.pdf

Pairwise permanova results

Download CSV

		Sample size	Permutations	pseudo-F	p-value	q-value
Group 1	Group 2					
adult	juvenile	11	999	1.462488	0.183	0.1830
	tadpole	12	999	2.033906	0.061	0.0915
juvenile	tadpole	11	999	4.568218	0.007	0.0210

https://view.qiime2.org/_/butuneot644/6d5eaf29-482e-4409-8ff0-d1ccec8b4d66/data/tadpole-boxplots.pdf

This is the table for the weighted unifrac life stage.

For the life stage data, the Bray Curtis table showed there was no significant difference between the different communities. On the table for the weighted unifrac there was a significant difference for the juvenile community with the tadpole p value being 0.007.

QIIME 2 Library View Forum

qiime2view File: weighted-unifrac-site-significance.qzv × Visualization Citations Provenance Metadata

Overview

	PERMANOVA results
method name	PERMANOVA
test statistic name	pseudo-F
sample size	17
number of groups	2
test statistic	1.281542
p-value	0.214
number of permutations	999

QIIME 2 Library View Forum

qiime2view File: bray-curtis-site-significance.qzv × Visualization Citations Provenance Metadata

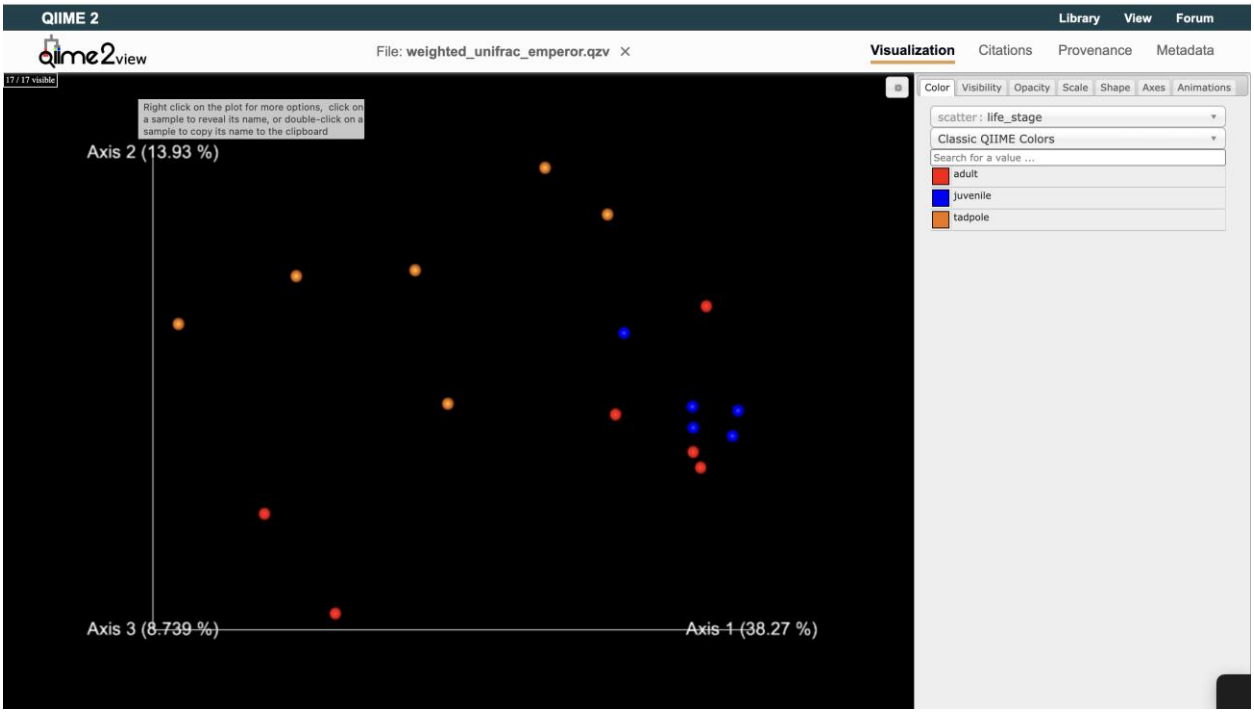
Overview

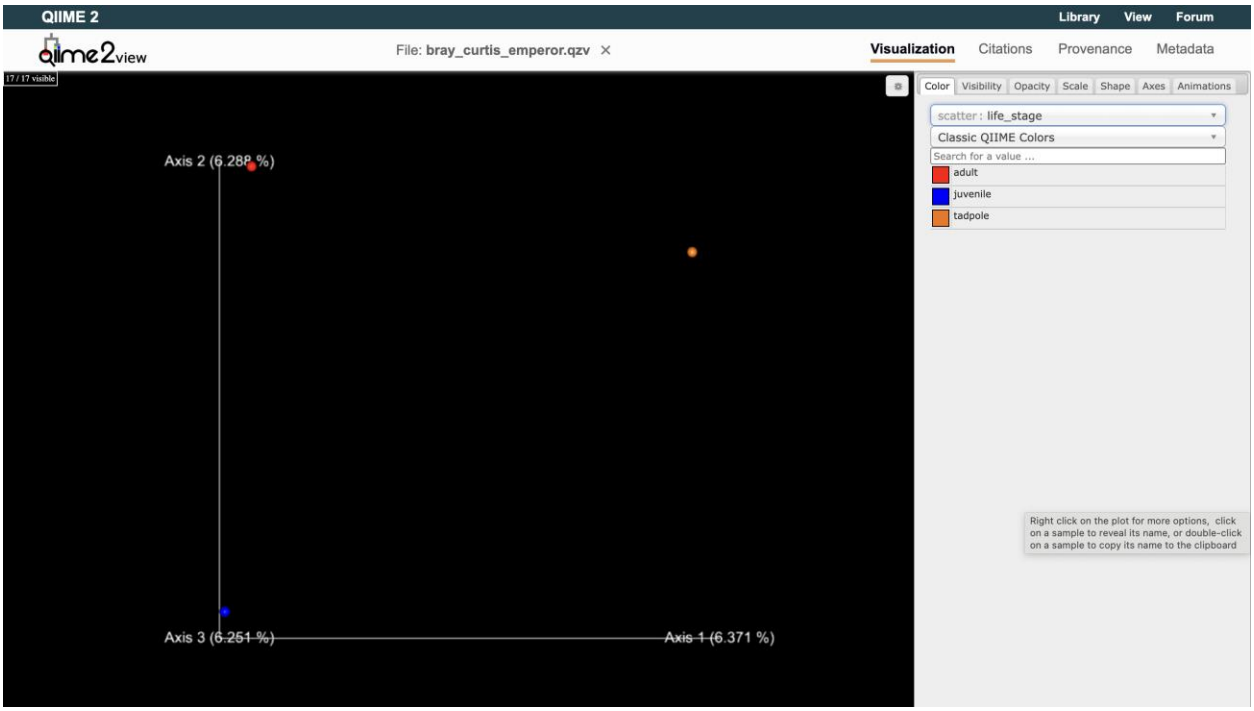
	PERMANOVA results
method name	PERMANOVA
test statistic name	pseudo-F
sample size	17
number of groups	2
test statistic	0.998484
p-value	0.517
number of permutations	999

For the site comparison, the Bray Curtis table and Weight Unifrac table showed no significant difference with both values being greater than 0.05.

11. **Once you have determined if there are any differences, you can view a plot of a Principle Components Analysis. This is basically a way to try and graphically represent community differences. You should view the Bray Curtis and Weighted Unifrac Emperor .qzv files. You can color code the individual points by site of collection or life stage. Do the points cluster like you would expect based on significance? For example, if you saw differences in life stage, do the various life stages seems to be close to one another (e.g., tadpole close to tadpole, but father from juvenile or adult). Include a screenshot of each Emperor plot (Bray Curtis and Weighted Unifrac) with the life stages color coded.**

The points do cluster as expected because the juvenile groups have a significant difference with the tadpole being farther from the other categories. The various life stages seem to be close and far from one another. The adult and juvenile stages are closer to each other while the tadpole stage is farther. The tadpoles are close to the tadpoles, the juveniles are close to the juveniles, and the adults are farther from other adults.





12. **If you find any that are differentially expressed, you can figure out what taxa they belong to by searching the taxonomy.qzv file for the alphanumeric code. Were there any differentially expressed taxa in the different sites? If so, provide at most 3 of them. You should repeat this analysis for the life stage as well.**

They were not differentially expressed, and the results were negative. This means the two communities being compared show no significant difference between them. The types of species in the communities are relatively the same.