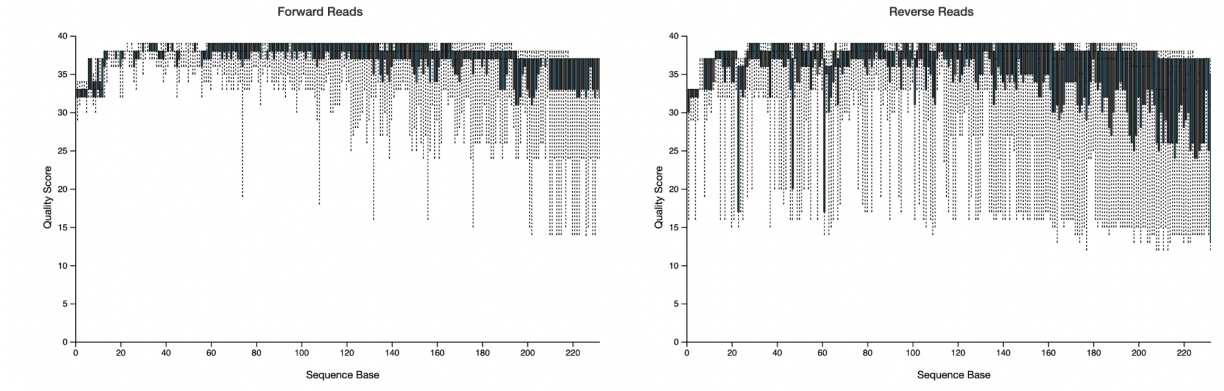
1. What columns do you see that you think might be useful for later when doing alpha and beta diversity metrics?

Sampleid is important to map metadata for feature/sequence tables. Population tests for diversity differences. Sex is important to discover sex based differences.

2. 

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

For the forward reads I chose 200 and 5 and for the reverse reads I chose 200 and 0. I chose these reads to eliminate the reads that were lower than 25.

3.

Table summary

Summary Statistic	Value
Number of samples	24
Number of features	2,210
Total frequency	311,390

Include a screenshot of the table summary from visualizing your table and a screenshot of the sequence length statistics from the rep-seqs file. Remember, we may eventually want to cut any samples with less than 10,000 reads. Do you see any in the interactive sample detail that might need to be cut? If so, which ones?

qiime2view File: table.qzv × Visualization Citations Provenance Metadata

Sample ID	Frequency
361_S168_L001	62,364
220_S155_L001	39,919
165_S230_L001	29,849
176_S154_L001	28,151
236_S241_L001	27,241
281_S130_L001	20,330
72_S206_L001	20,090
50_S144_L001	13,324
282_S217_L001	12,152
57_S153_L001	8,546
208_S177_L001	7,628
331_S131_L001	7,459
368_S129_L001	6,481
94_S278_L001	6,359
90_S107_L001	5,858
210_S336_L001	4,837
260_S178_L001	2,703
306_S120_L001	2,527
133_S265_L001	1,634
122_S207_L001	1,172
332_S105_L001	912
119_S106_L001	756
252_S179_L001	611
41_S254_L001	487

All of the values that are less than 10,000 are getting cut, over half of the bottom portion.

Sequence Length Statistics

Download sequence-length statistics as a TSV

Sequence Count	Min Length	Max Length	Mean Length	Range	Standard Deviation
2210	200	360	245.93	160	10.63

4. Once you have generated your taxonomy visualization, sort it by confidence. What are your top hits? What about if you sort by taxon? What hits do you see?

   My top hits when I visualize the taxonomy are bacteria Bacteroidetes and bacteria Proteobacteria. If I sort it by taxon my top hits show all bacteria.

5. When you visualize level 3 of taxonomy, what level is this? Do you see any trends as you sort by various metadata categories?

   When you visualize level 3 of taxonomy you can see the class of the bacteria. You can see the relative frequency of the bacteria. I notice when I sort by other various metadata categories, the same bacteria remain with the highest relative frequency, while the other bacteria vary when you sort by other trends.

6. After visualizing your filtered-table.qzv, what cutoff value will you use for generating alpha and beta diversity? Why? Include a screenshot of the interactive sample view to help justify your reasoning.

The cutoff value I'll use for generating alpha and beta diversity is 1,514 I chose this value because the values tend to get super low after this cut off.

7. The first metric we will analyze is alpha diversity. In your own words, what is alpha diversity and what are the differences between the two types of alpha diversity we will analyze (Shannon and Observed features)?

Alpha diversity measures the variety of species within a sample to see how diverse that individual's environment is. Observed features focus on the richness by the number of unique species. Shannon diversity measures richness and evenness, taking in consideration the evenness of their population's distribution.

8. Since you are looking at two metrics of diversity for 3 metadata categories, it would be helpful to make a table of the significance values. Are any of your comparisons significant? For one of the metadata data columns, there are actually 4 options. Include a screenshot of the pairwise comparisons for Shannon and Observed. Are there any significant comparisons?

Metadata Category	Diversity Metric	q-value	Significant
Population	Observed Features	0.297386	No
	Shannon Entropy	0.618402	No
Sex	Observed Features	0.306625	No
	Shannon Entropy	0.286321	No

Flock	Observed Features	0.201243-0.916816	No, all pairs
	Shannon Entropy	0.273322	No, all pairs

Shannon (flock, sex, population)

Kruskal-Wallis (pairwise)

Download CSV

Group 1	Group 2	H	p-value	q-value
migratoryfemale (n=5)	migratorymale (n=2)	0.000000	1.000000	1.000000
	residentfemale (n=5)	0.098182	0.754023	0.904827
	residentmale (n=6)	1.200000	0.273322	0.904827
migratorymale (n=2)	residentfemale (n=5)	0.150000	0.698535	0.904827
	residentmale (n=6)	0.444444	0.504985	0.904827
residentfemale (n=5)	residentmale (n=6)	0.833333	0.361310	0.904827

Kruskal-Wallis (pairwise)

Download CSV

Group 1	Group 2	H	p-value	q-value
female (n=10)	male (n=8)	1.136842	0.286321	0.286321

Kruskal-Wallis (pairwise)

Download CSV

Group 1	Group 2	H	p-value	q-value
migratory (n=7)	resident (n=11)	0.24812	0.618402	0.618402

Observed features (flock, sex, population)

Kruskal-Wallis (pairwise)

Download CSV

Group 1	Group 2	H	p-value	q-value
migratoryfemale (n=5)	migratorymale (n=2)	0.038182	0.845078	0.916815
	residentfemale (n=5)	0.010909	0.916815	0.916815
	residentmale (n=6)	1.633333	0.201243	0.634621
migratorymale (n=2)	residentfemale (n=5)	0.150000	0.698535	0.916815
	residentmale (n=6)	1.000000	0.317311	0.634621
residentfemale (n=5)	residentmale (n=6)	1.200000	0.273322	0.634621

Kruskal-Wallis (pairwise)

Download CSV

Group 1	Group 2	H	p-value	q-value
female (n=10)	male (n=8)	1.045158	0.306625	0.306625

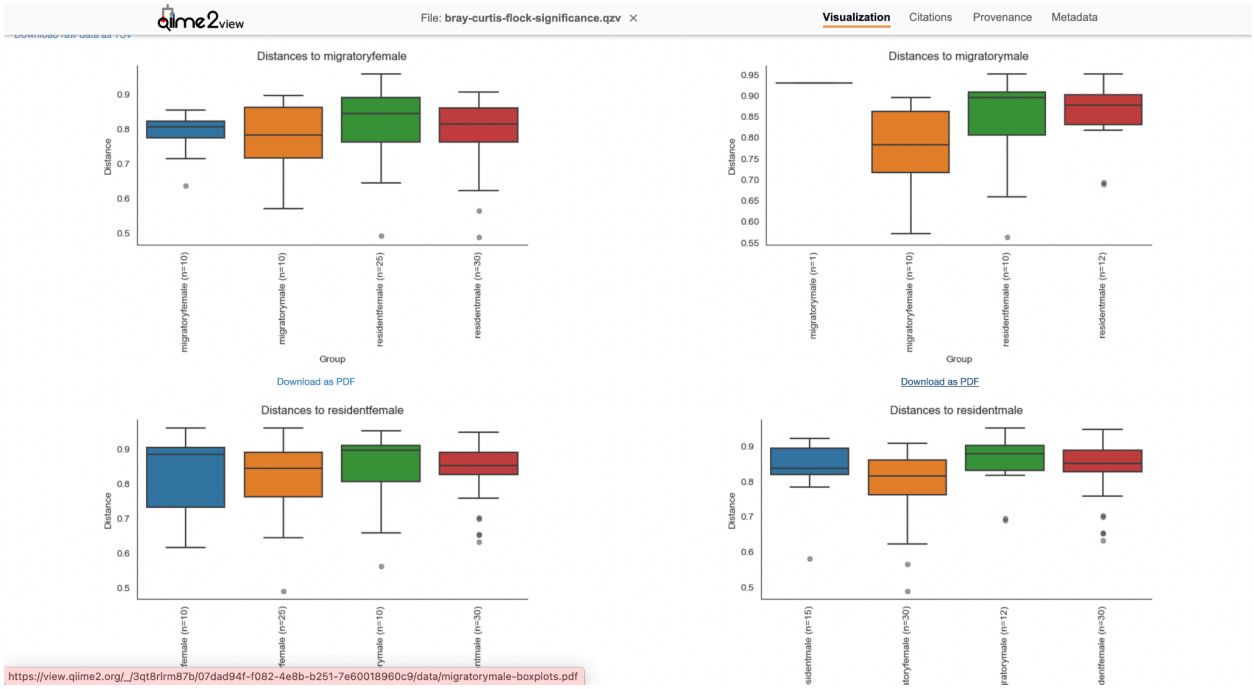
Kruskal-Wallis (pairwise)

Download CSV

		H	p-value	q-value
Group 1	Group 2			
migratory (n=7)	resident (n=11)	1.085878	0.297386	0.297386

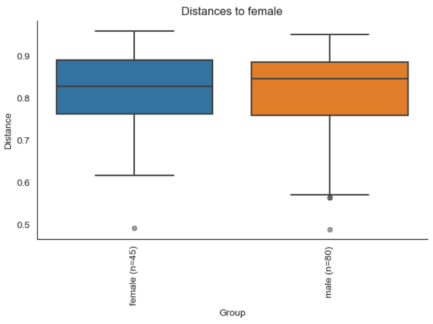
10) For beta diversity, you will need to create visualizations for Bray Curtis dissimilarity and Weighted Unifrac distance. This will require you to modify the `beta-group-significance` code. You should have one visualization for each metadata column you are interested in. Include a screenshot of each visualization. Is there any significance? Regardless of significance, how can you interpret these results (hint: what is beta diversity looking at?) How are Bray Curtis dissimilarity and Weighted Unifrac distance different?

Bray Curtis (flock, sex, population)

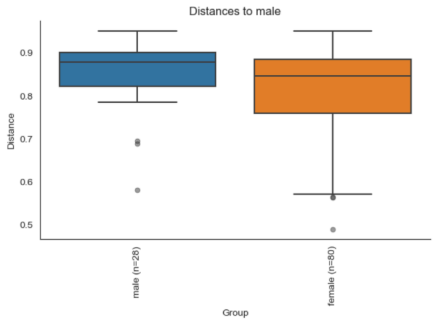


Group significance plots

Download raw data as TSV



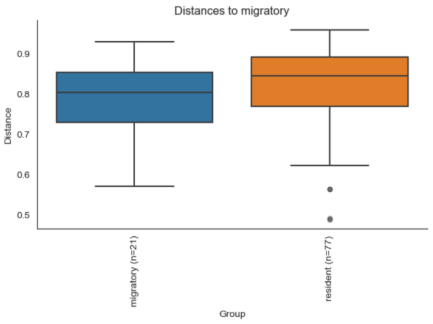
Download as PDF



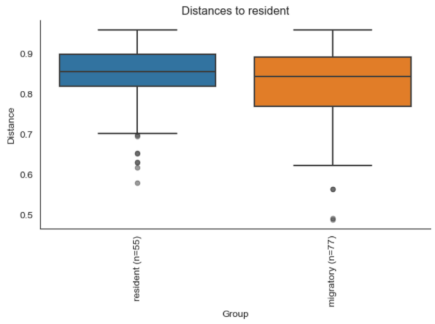
Download as PDF

Group significance plots

Download raw data as TSV



Download as PDF

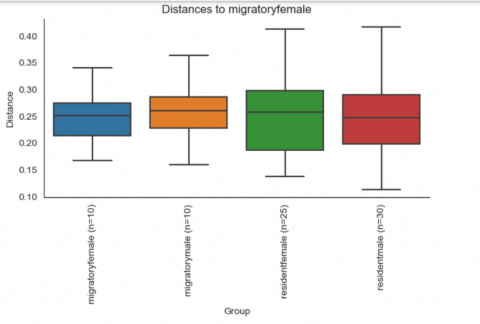


Download as PDF

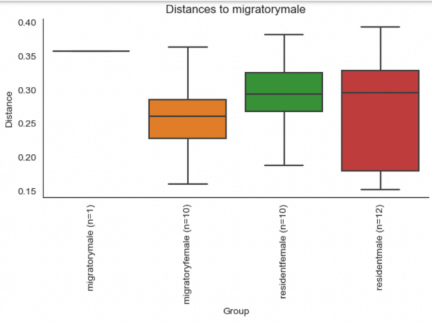
Weighted Unifrac distance (flock, sex, population)

QIIME 2 Library View Forum

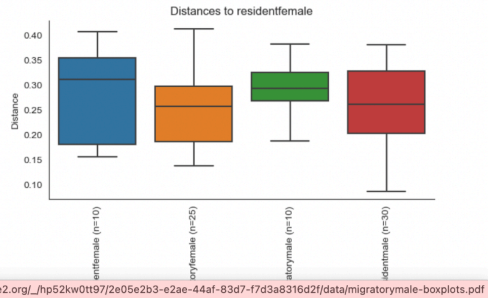
qiime2view File: weighted-unifrac-flock-significance.qzv × Visualization Citations Provenance Metadata

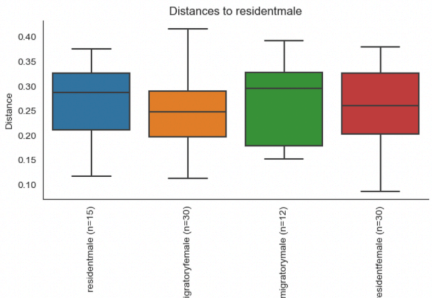


Download as PDF



Download as PDF

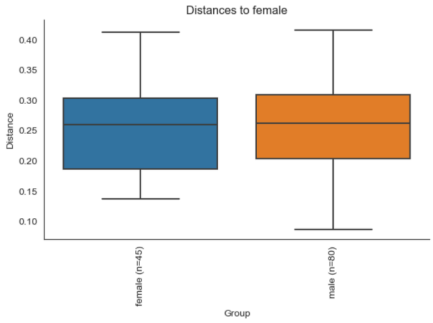




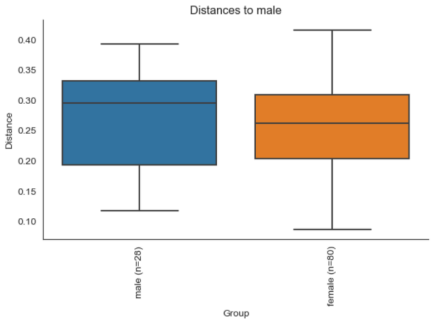
https://view.qiime2.org/_/hp52kw0tt97/2e05e2b3-e2ae-44af-83d7-f7d3a8316d2f/data/migratorymale-boxplots.pdf

Group significance plots

Download raw data as TSV



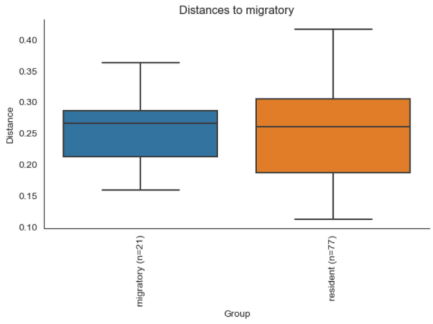
Download as PDF



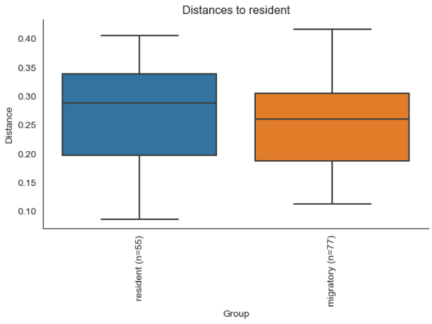
Download as PDF

Group significance plots

Download raw data as TSV



Download as PDF

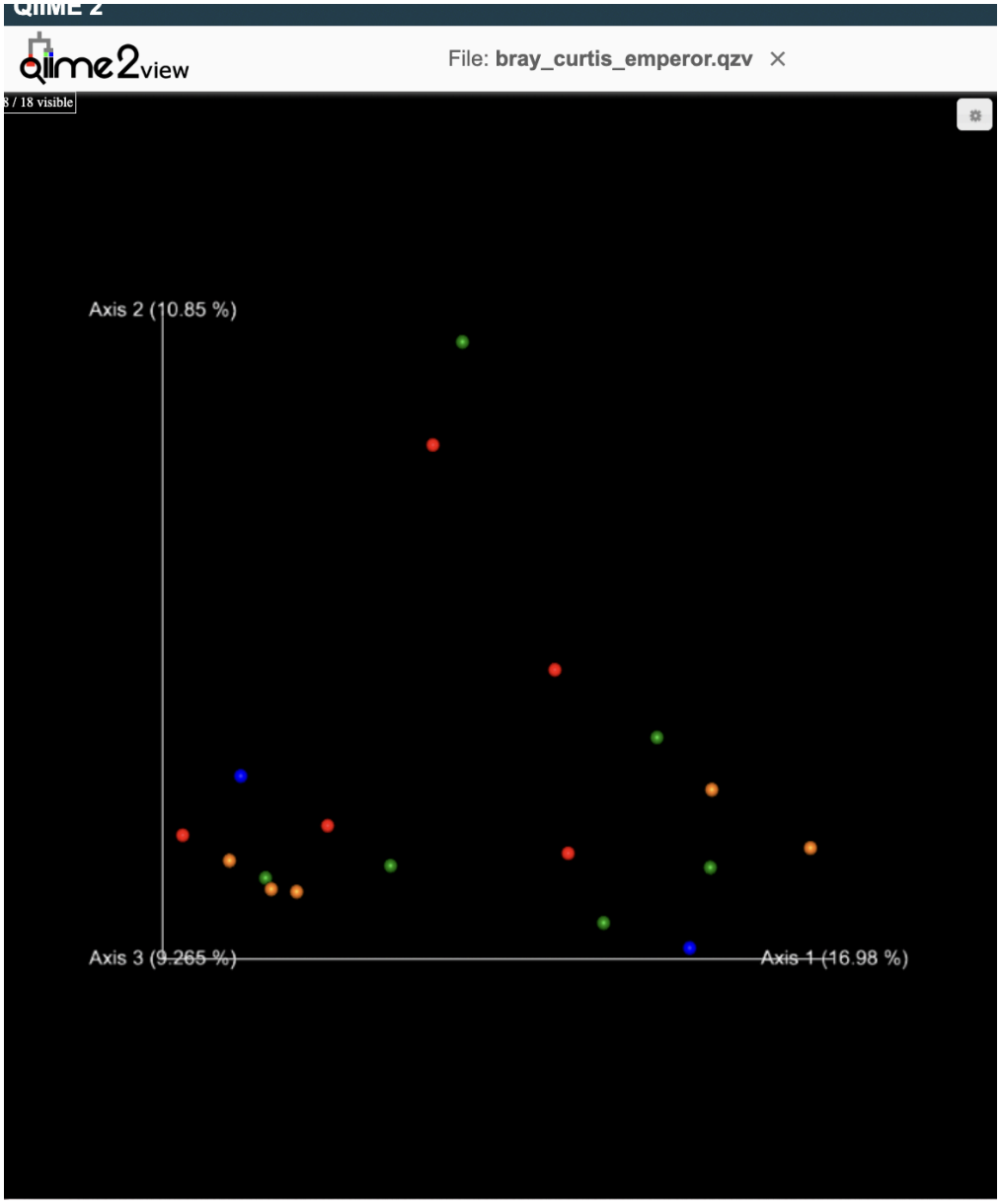


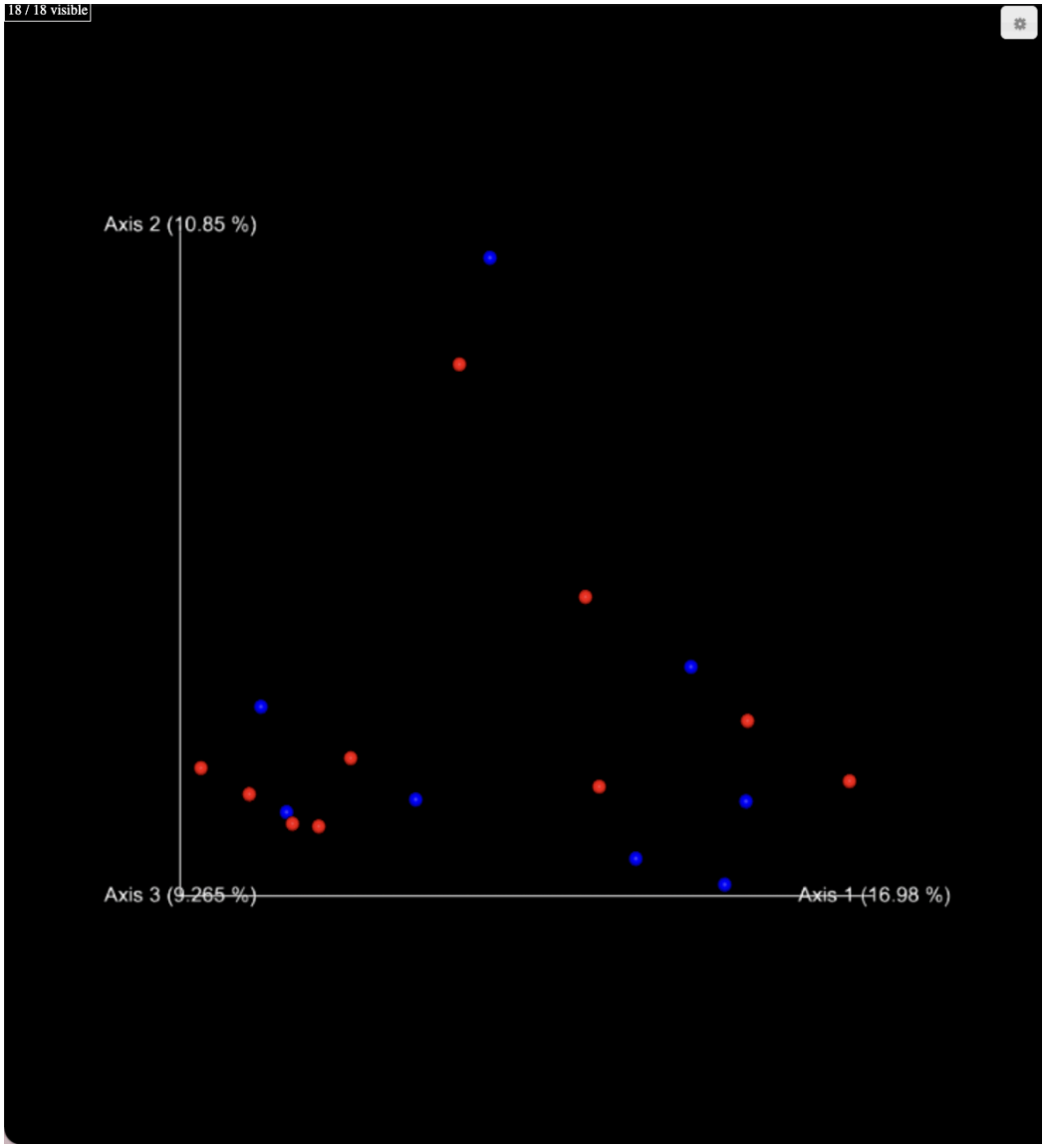
Download as PDF

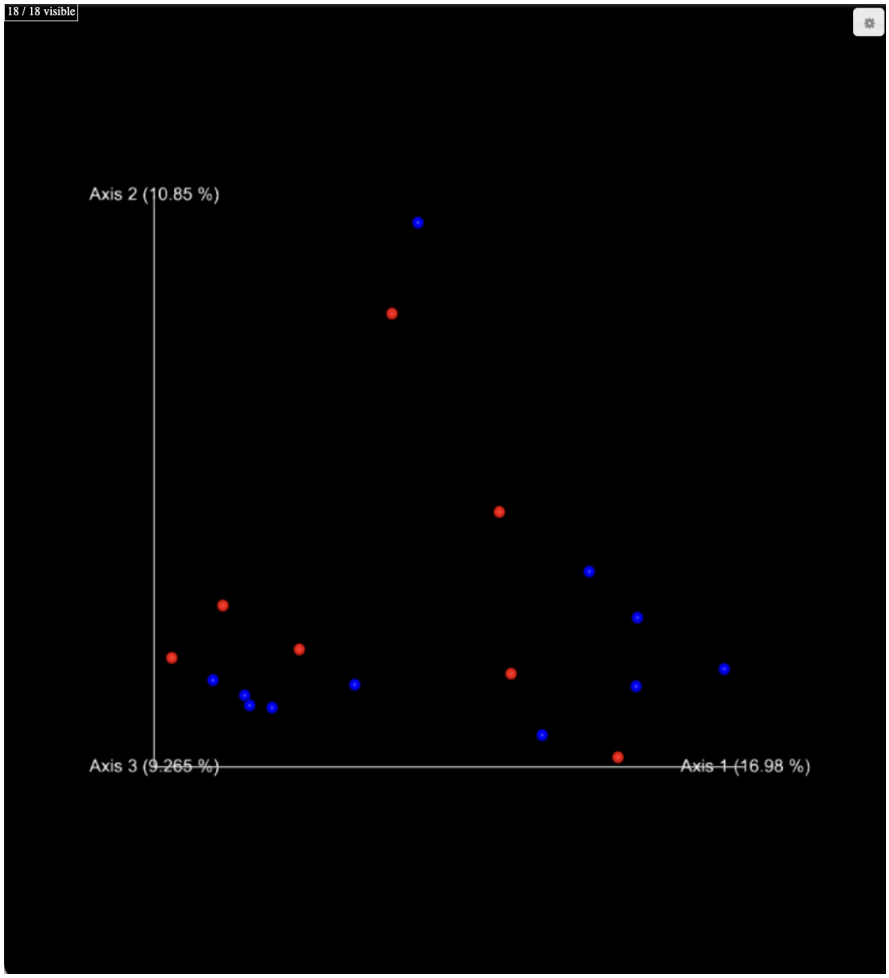
Bray Curtis measures differences in species abundance while Weighted Unifrac distance measures the difference in their abundance and how closely related the species are.

11) The `core-metrics-phylogeny` command generates a file called `bray-curtis-emporer.qzv.` Include 3 screenshots total (where the points are colored based on the metadata metrics). How do these results help you make sense of the results you got from question 10?

Flock, sex, and population


QIIME 2
qiime2view
File: bray_curtis_emperor.qzv
8 / 18 visible
Axis 2 (10.85 %)
Axis 3 (9.265 %)
Axis 1 (16.98 %)

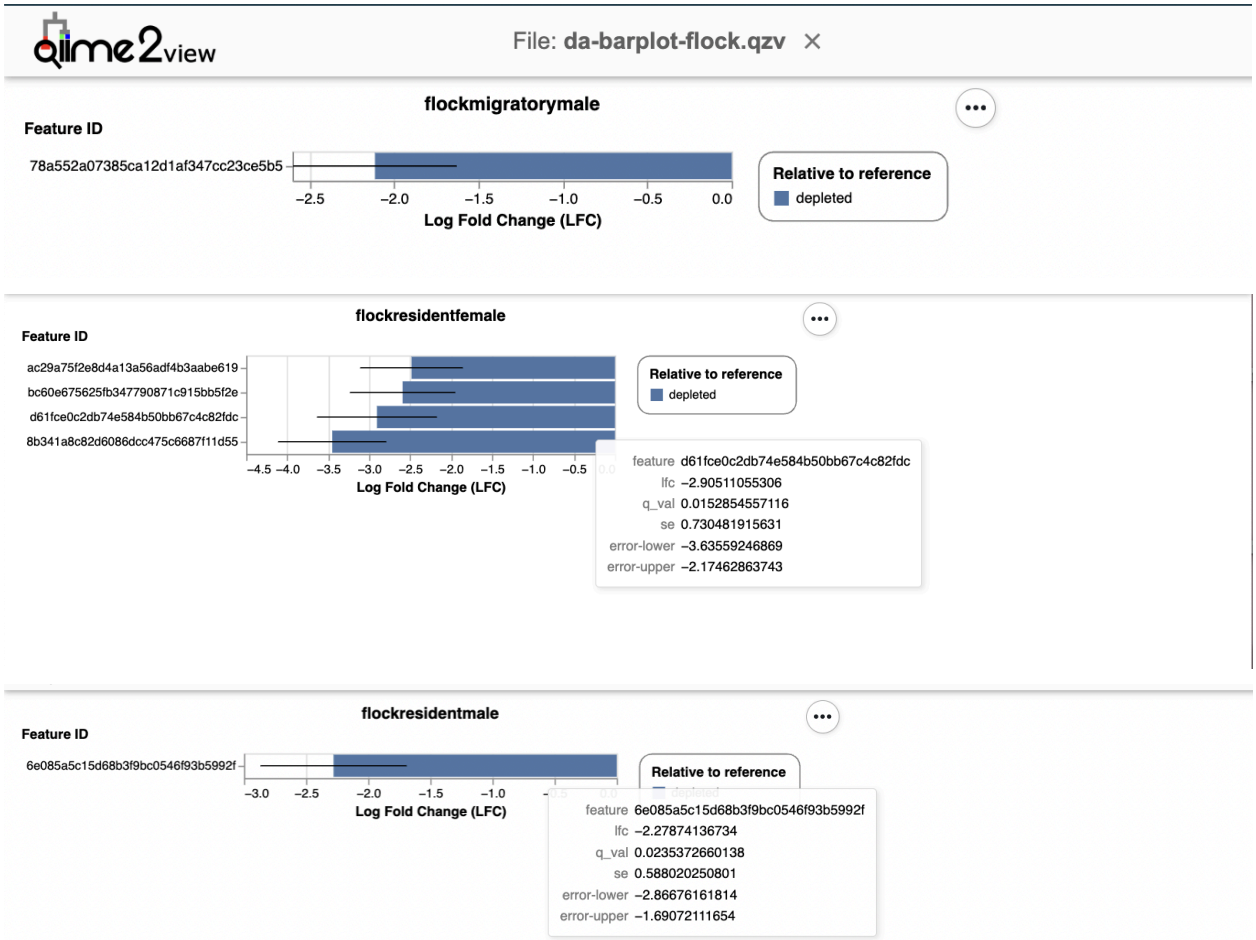

18 / 18 visible
Axis 2 (10.85 %)
Axis 3 (9.265 %)
Axis 1 (16.98 %)



This results help make sense of the results from question 10 because they show no clear separation between groups. Both Bray Curtis and Weighted Unifrac showed no significant differences between the groups indicating that the communities are similar across metadata categories.

12) Using ANCOMBC, do you find any specific taxa that are differentially expressed with your three metadata categories? Please include a screenshot of any differentially expressed taxa and identify the species using the taxonomy.qzv file.

Flock



Feature ID	Taxon
78a552a07385ca12d1af347cc23ce5b5	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales
ac29a75f2e8d4a13a56adf4b3aabe619	k__Bacteria; p__Proteobacteria; c__Betaproteobacteria; o__Burkholderiales; f__Comamonadaceae; g__Comamonas; s__
bc60e675625fb347790871c915bb5f2e	k__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__
d61fce0c2db74e584b50bb67c4c82fdc	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria
8b341a8c82d6086dcc475c6687f11d55	k__Bacteria; p__Bacteroidetes;

	c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__
6e085a5c15d68b3f9bc0546f93b5992f	**k__Bacteria; p__Acidobacteria; c__Acidobacteriia; o__Acidobacteriales; f__Acidobacteriaceae; g__; s__**

There are no species but a list of taxonomies above.