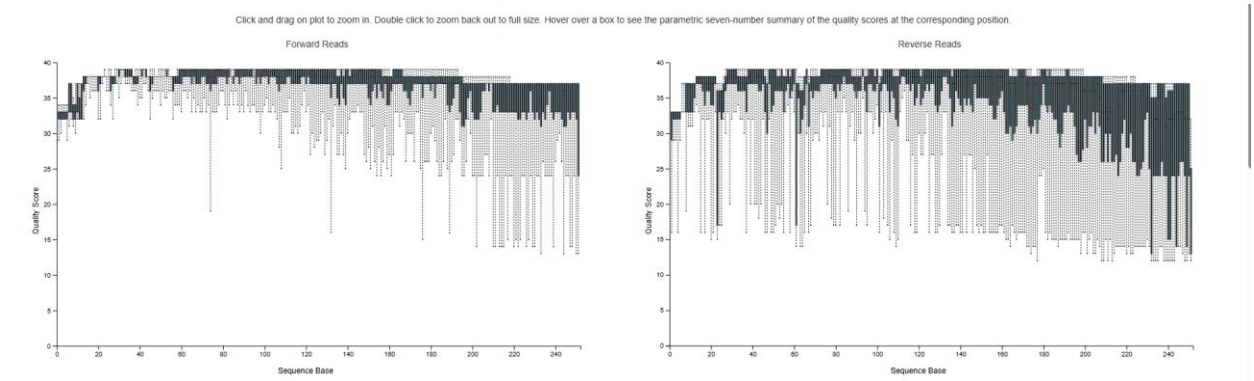
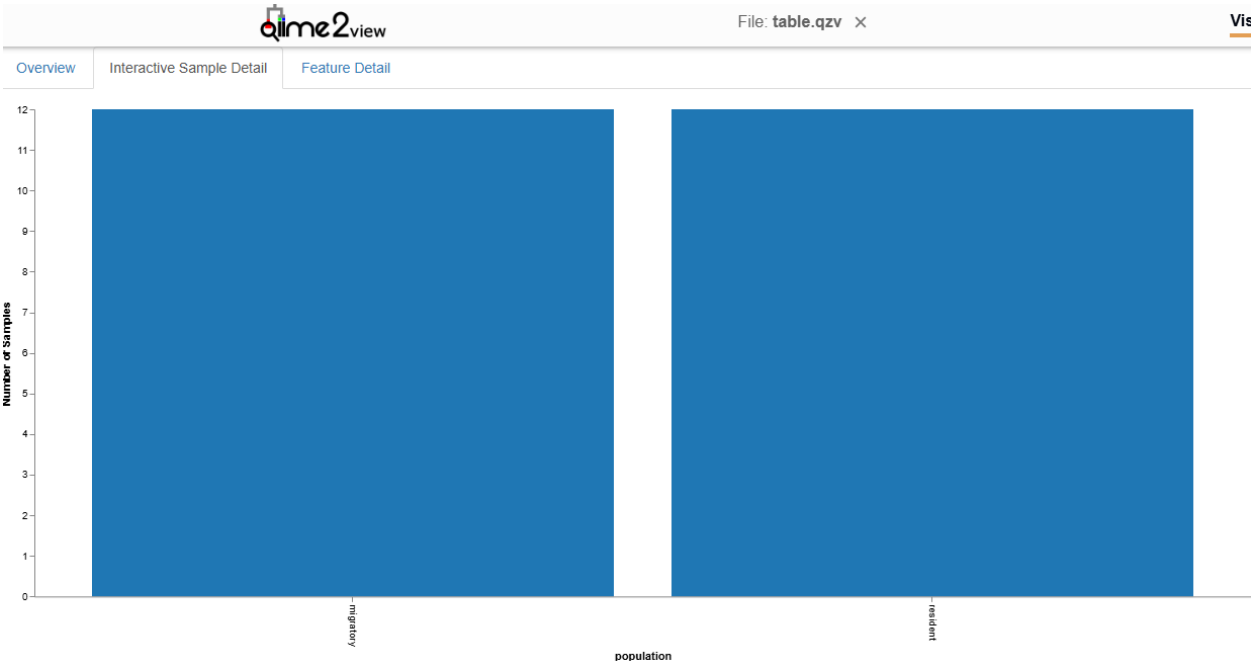
1. The 3 group samples that we can use for alpha, and beta would be Population, sex, and flock.
2. I chose the number 0, 190, and 23,230. I did this because I felt the beginning of the reads are important and useful, so I kept the first at 0, however I made a typo, and entered 23 for the 2nd one when I meant to keep it at 0. I then trimmed the right hand more because these are the points I felt they peaked and had enough information. So 190 for the forward and 230 for the reverse.



3. About 15 of the reads will need to be cut because they are below 10,000 reads.



Sequence Length Statistics

Download sequence-length statistics as a TSV

Sequence Count	Min Length	Max Length	Mean Length	Range	Standard Deviation
2122	230	342	230.56	112	6.29

4. My top hits at ~0.99 confidence is “k__Bacteria; p__Firmicutes; c__Bacilli; o__Bacillales; f__Paenibacillaceae; g__Saccharibacillus; s__kuerlensis” When I sort by taxon the hits, I see are all bacteria for the first 7 pages.
5. At level 3 we are at the class level of the taxonomy hierarchy. There is more diversity on the left of the graph with classes like Actinobacteria, Bacilli, and flavobacteria. As you move to the right classes like Betaproteobacteria and Alphaproteobacteria increase in abundance. Proteobacteria are the most dominant samples.
6. The cutoff value I have determined will be 903, because it includes as many samples, super low reads could skew the results, so I chose 918, since the ones below it are too small.

qiime2view File: **filtered-table.qzv** ×

Sample ID	Frequency
361_S168_L001	52,840
220_S155_L001	31,711
236_S241_L001	22,552
165_S230_L001	22,023
176_S154_L001	21,962
72_S206_L001	17,123
281_S130_L001	12,206
282_S217_L001	7,276
50_S144_L001	6,261
208_S177_L001	6,101
57_S153_L001	5,850
94_S278_L001	5,405
331_S131_L001	5,011
368_S129_L001	4,038
210_S336_L001	3,576
90_S107_L001	2,784
260_S178_L001	2,185
306_S120_L001	1,302
122_S207_L001	918
133_S265_L001	903
332_S105_L001	496
41_S254_L001	377
119_S106_L001	371
252_S179_L001	347

7. Alpha diversity is the measure of diversity within a single sample, this tells us how complex the microbial community is in that specific environment. Observed deals with the number of unique features in a sample, so how many different types are present. Meanwhile, Shannon measures how many types and how they are distributed. Basically, seeing if its balances, so a higher Shannon would mean its more diverse, which is a balanced community.
8.

Table summarization:

Metadata category	Diversity metric	q-value	Signifigance
Sex	Shannon	0.21	No
	Observed	0.34	No
Flock	Shannon	0.88 0.546	No
	Observed	0.4 0.7 1.0	No
Population	Shannon	0.305	No
	Observed	0.11	No

Shannon (sex,flock, population)

Kruskal-Wallis (pairwise)

Download CSV

		H	p-value	q-value
Group 1	Group 2			
female (n=11)	male (n=9)	1.571429	0.21	0.21

Kruskal-Wallis (pairwise)

Download CSV

		H	p-value	q-value
Group 1	Group 2			
migratoryfemale (n=6)	migratorymale (n=3)	0.266667	0.605577	0.881497
	residentfemale (n=5)	0.033333	0.855132	0.881497
	residentmale (n=6)	1.641026	0.200185	0.546643
migratorymale (n=3)	residentfemale (n=5)	0.022222	0.881497	0.881497
	residentmale (n=6)	1.666667	0.196706	0.546643
residentfemale (n=5)	residentmale (n=6)	1.200000	0.273322	0.546643

Kruskal-Wallis (pairwise)

Download CSV

		H	p-value	q-value
Group 1	Group 2			
migratory (n=9)	resident (n=11)	1.051948	0.305059	0.305059

Observed (sex, flock, population)

Kruskal-Wallis (pairwise)

Download CSV

		H	p-value	q-value
Group 1	Group 2			
female (n=11)	male (n=9)	0.902555	0.342098	0.342098

Kruskal-Wallis (pairwise)

Download CSV

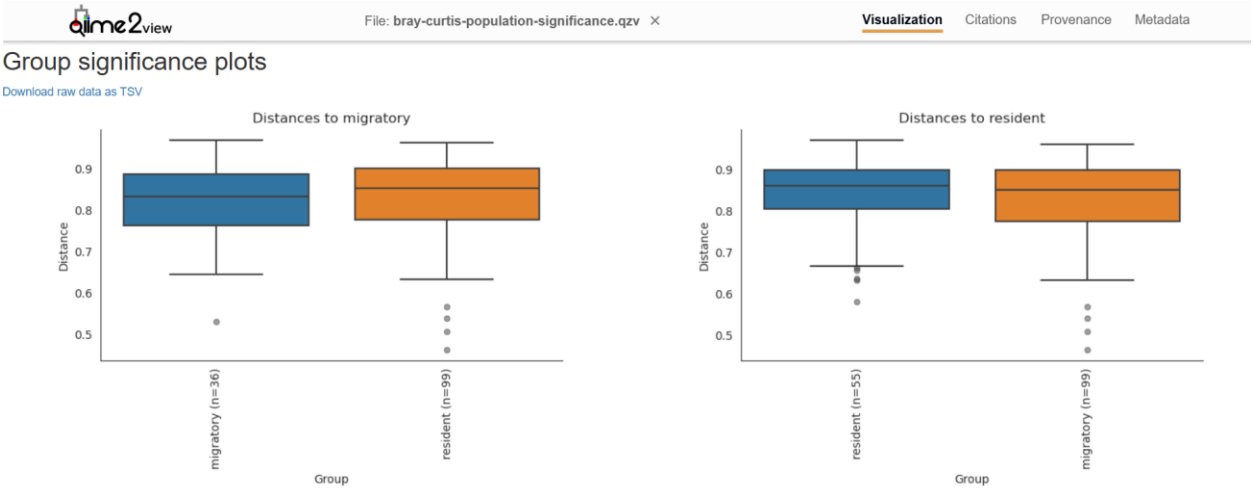
		H	p-value	q-value
Group 1	Group 2			
migratoryfemale (n=6)	migratorymale (n=3)	0.000000	1.000000	1.000000
	residentfemale (n=5)	0.209285	0.647329	0.776794
	residentmale (n=6)	2.076923	0.149541	0.402485
migratorymale (n=3)	residentfemale (n=5)	1.088889	0.296718	0.445076
	residentmale (n=6)	2.400000	0.121335	0.402485
residentfemale (n=5)	residentmale (n=6)	1.633333	0.201243	0.402485

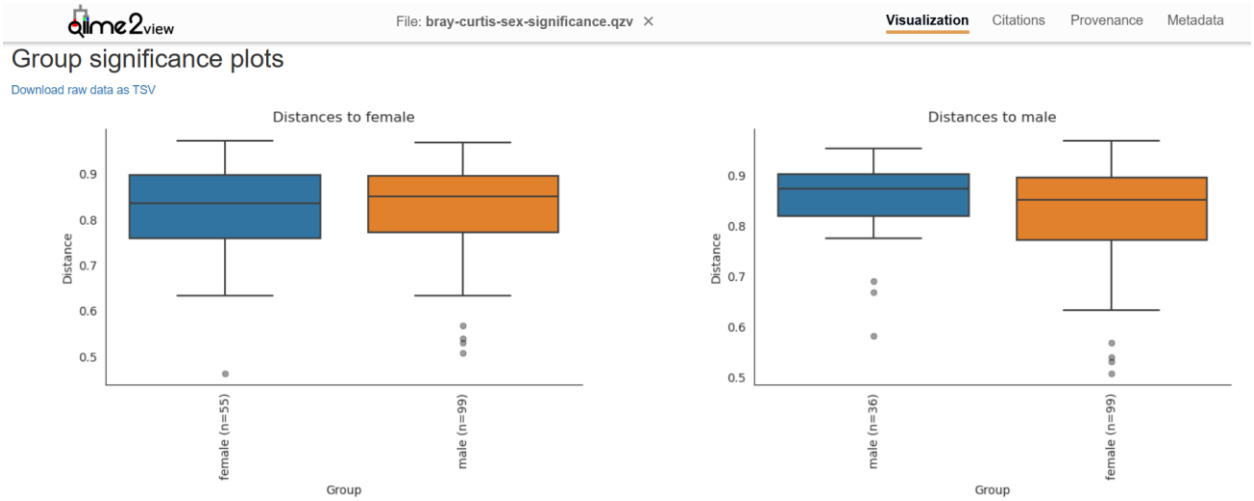
Kruskal-Wallis (pairwise)

Download CSV

		H	p-value	q-value
Group 1	Group 2			
migratory (n=9)	resident (n=11)	2.54737	0.110478	0.110478

9. **bray curtis: flock,population,sex**





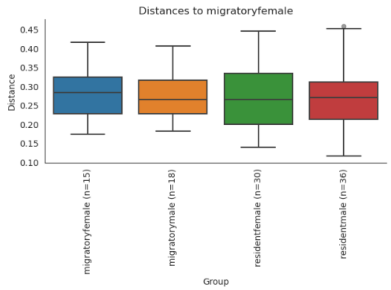
Weighted Unifrac: flock, population, sex

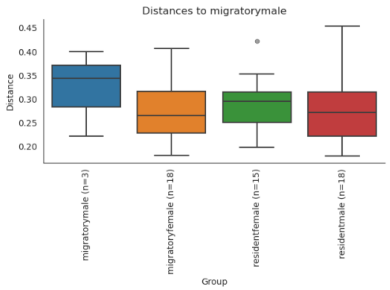
File: weighted-unifrac-flock-significance.qzv ×

Visualization Citations Provenance Metadata

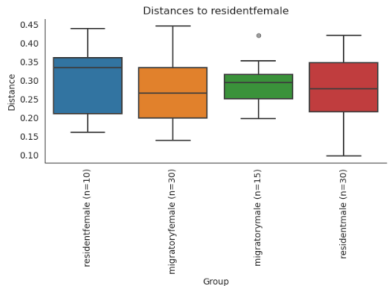
Group significance plots

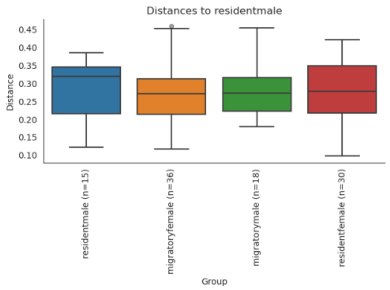
Download raw data as TSV





Download as PDF





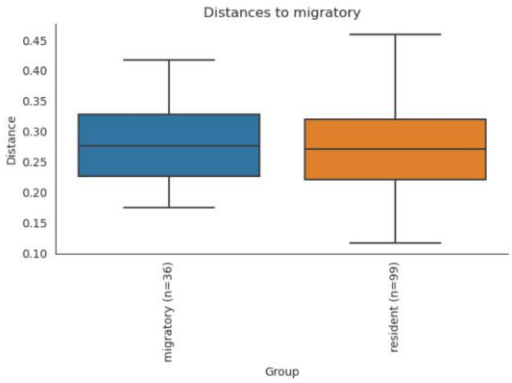
Download as PDF

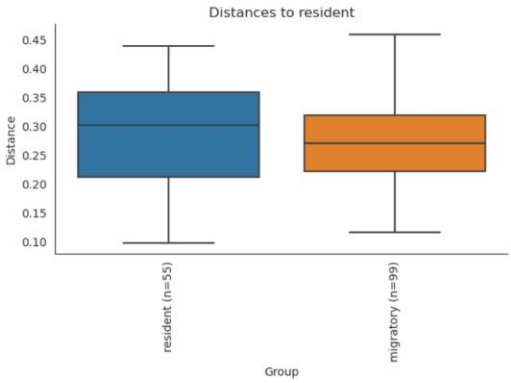
File: weighted-unifrac-population-significance.qzv ×

Visualization Citations Provenance Metadata

Group significance plots

Download raw data as TSV



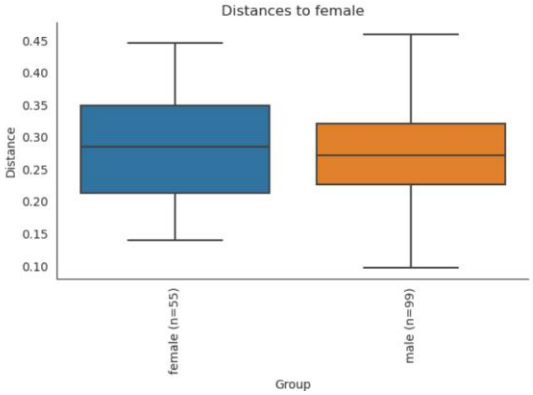


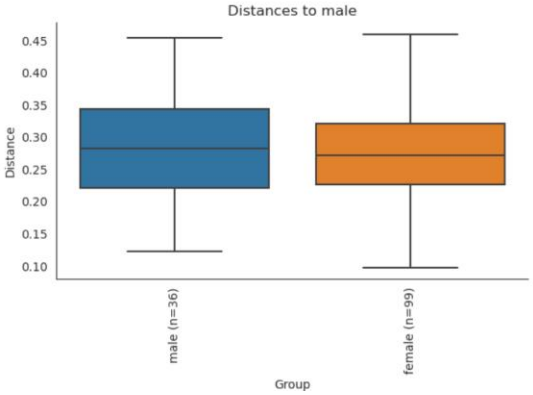
File: weighted-unifrac-sex-significance.qzv ×

Visualization Citations Provenance Metadata

Group significance plots

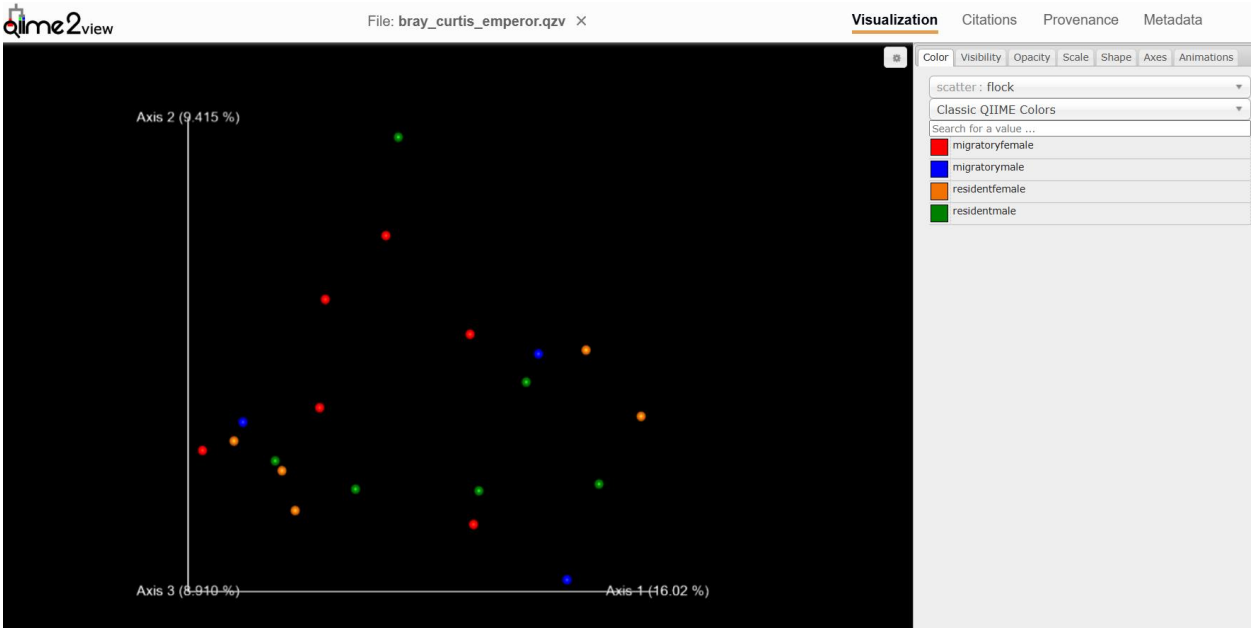
Download raw data as TSV

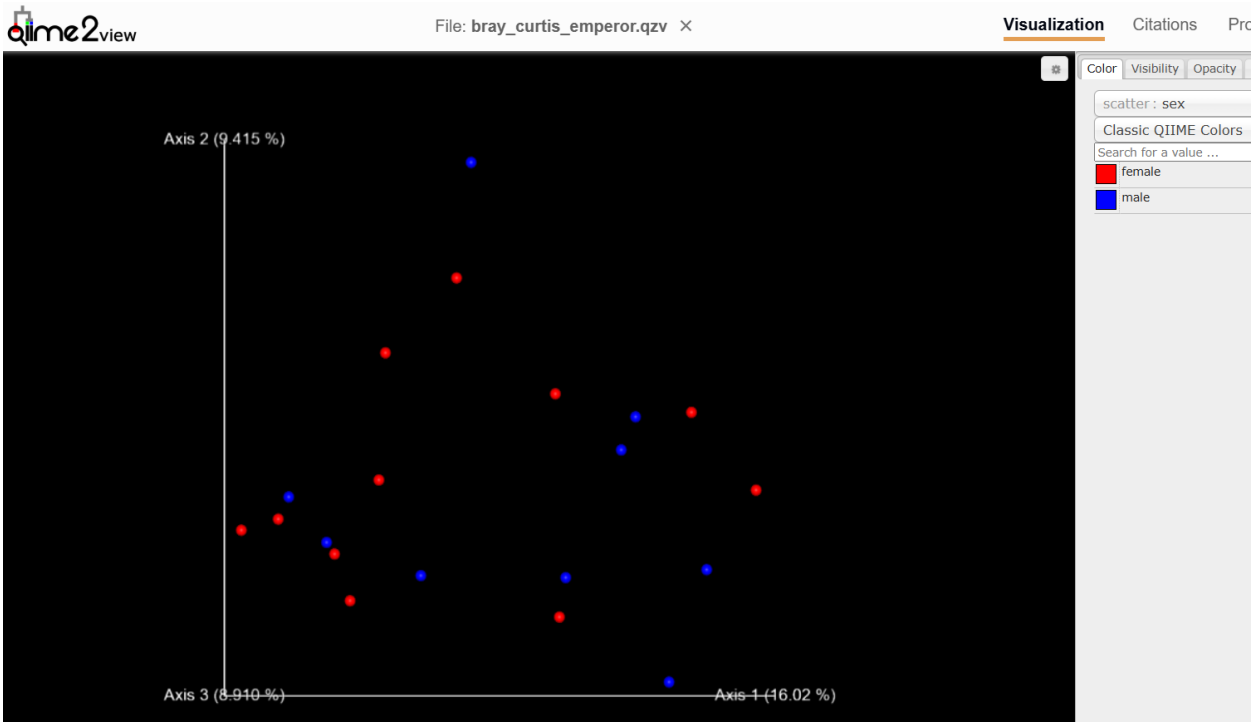


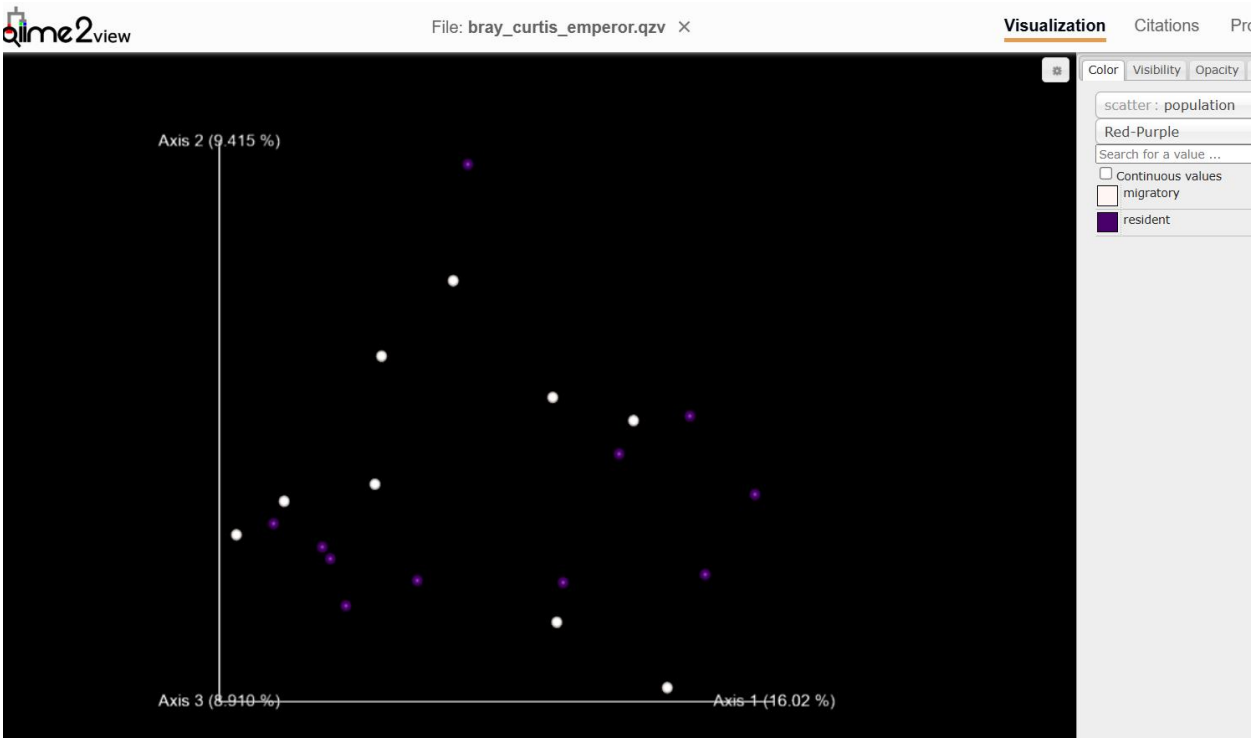


There were no significant values for any of the metadata groups, from either bray curtis or weighted unifrac. Beta diversity measures how different microbial communities are between samples. Bray curtis is based on only species abundance, so it ignores evolutionary and phylogenetic relationships, so it sees if they are identical or completely different. Meanwhile weighted unifrac will have a focus on phylogenetic relationships, since it is weighted, it will consider both presence, absence and abundance.

10.

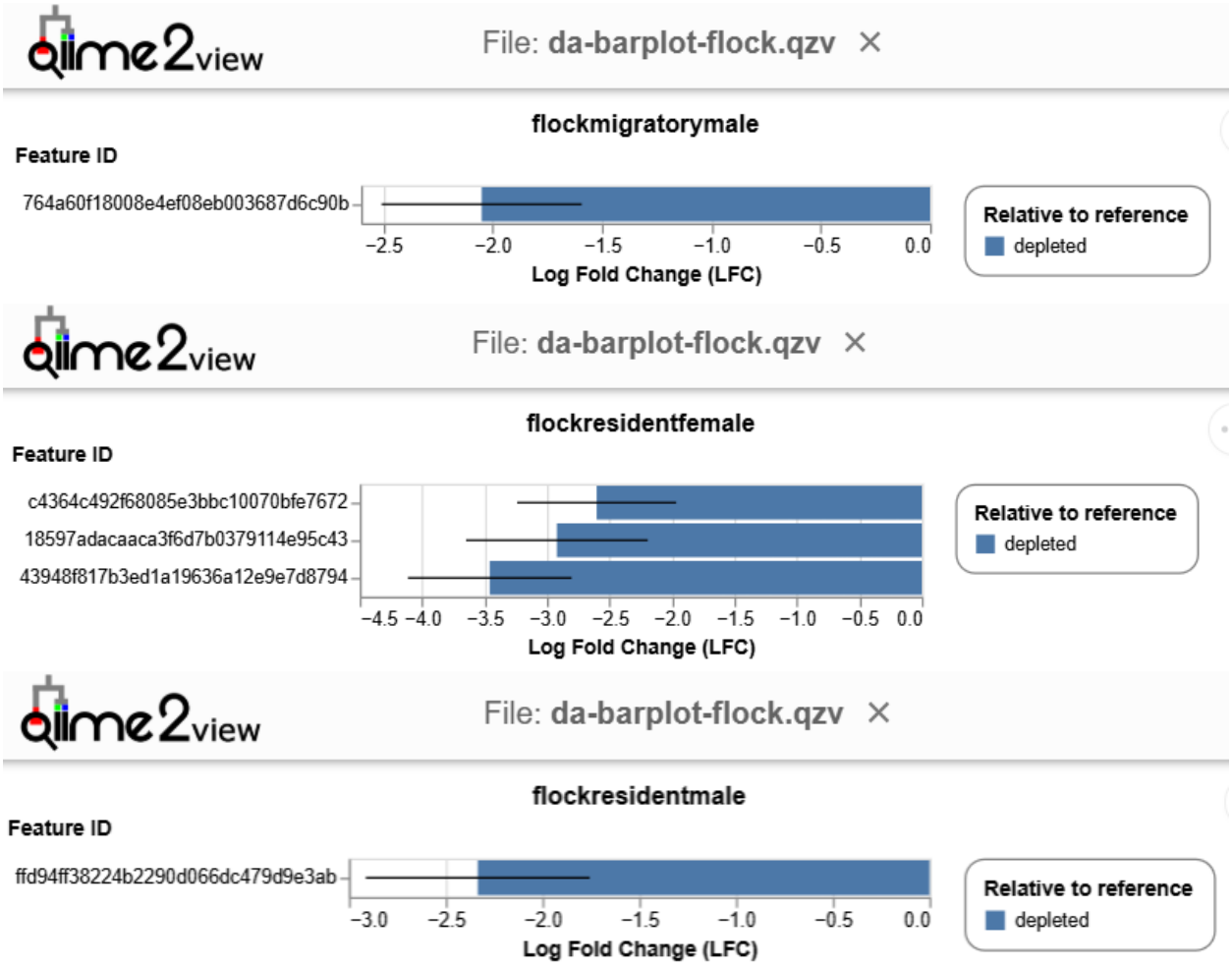






I see overlap between groups, which may be why I am getting no significant p-values. I see a trend of more migratory females on the left hand side of these graphs.

11.



Flock migratory female

Feature ID #q2:types	Taxon categorical
764a60f18008e4ef08eb003687d6c90b	k__Bacteria; p__Actinobacteria; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__Arcanobacterium; s__

Flock resident female

c4364c492f68085e3bbc10070bfe7672	k__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__	0.8902393584027176

Feature ID #q2:types	Taxon categorical	Confidence categorical
18597adacaaca3f6d7b0379114e95c43	k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria	0.8223747931836406

Feature ID #q2:types	Taxon categorical	Confidence categorical
43948f817b3ed1a19636a12e9e7d8794	k__Bacteria; p__Bacteroidetes; c__Flavobacteriia; o__Flavobacteriales; f__Flavobacteriaceae; g__Flavobacterium; s__	0.8888279367139189

Flock resident male:

Feature ID #q2:types	Taxon categorical	Confidence categorical
ffd94ff38224b2290d066dc479d9e3ab	k__Bacteria; p__Acidobacteria; c__Acidobacteriia; o__Acidobacteriales; f__Acidobacteriaceae; g__; s__	0.9421323999726458

Click a link to see the differential abundance bar plot for the specified category:

- Couldn't generate plot for populationresident: No features remaining after applying filters.

Notes on interpreting plots with taxonomic feature identifiers:

- If taxonomic labels are used to identify features, the feature labels (y-axis labels) in each plot represent the most specific named taxonomic level associated with that feature.
- Hover over the bars in plots to see the full taxonomic label of each feature identifier and information about its differential abundance relative to the reference.
- Feature identifiers (y-axis labels) that are followed by an asterisk (*) represent instances of a duplicated taxonomic name at the level displayed in the feature identifier. The number preceding the feature identifiers in these cases is used only for unique identification in the current figure. It is not taxonomically meaningful, and it won't be consistent across visualizations.

Click a link to see the differential abundance bar plot for the specified category:

- Couldn't generate plot for sexmale: No features remaining after applying filters.

Notes on interpreting plots with taxonomic feature identifiers:

- If taxonomic labels are used to identify features, the feature labels (y-axis labels) in each plot represent the most specific named taxonomic level associated with that feature.
- Hover over the bars in plots to see the full taxonomic label of each feature identifier and information about its differential abundance relative to the reference.
- Feature identifiers (y-axis labels) that are followed by an asterisk (*) represent instances of a duplicated taxonomic name at the level displayed in the feature identifier. The number preceding the feature identifiers in these cases is used only for unique identification in the current figure. It is not taxonomically meaningful, and it won't be consistent across visualizations.