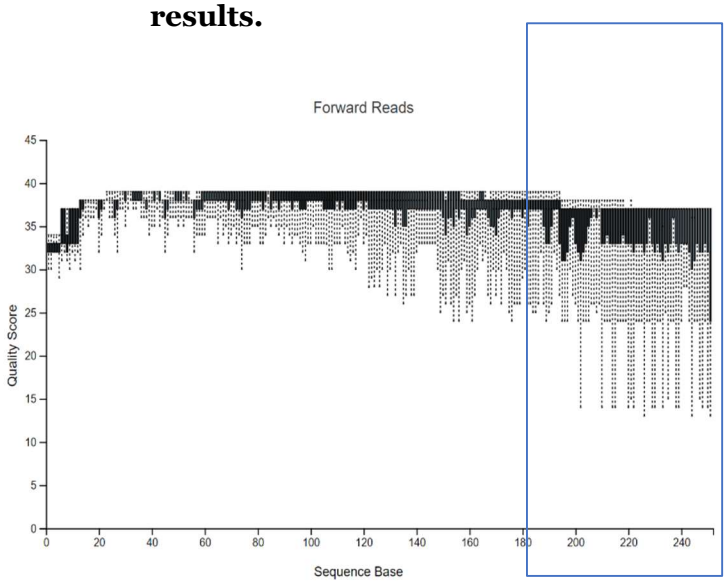
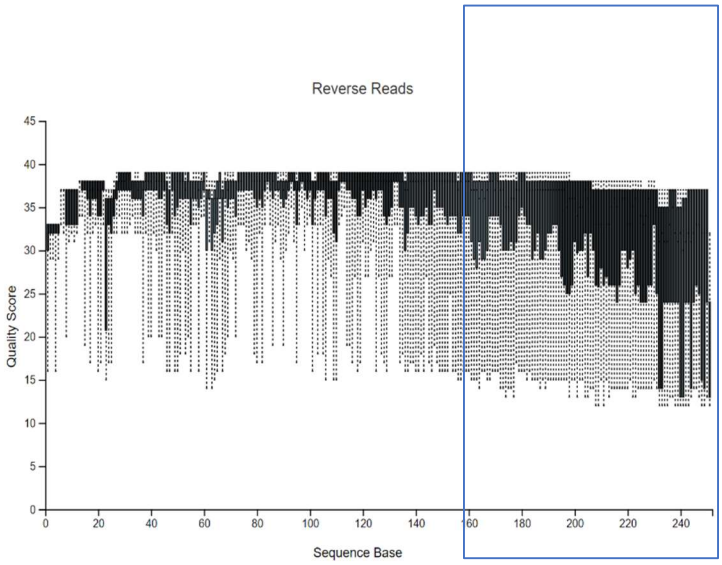
1) Include a screenshot of your interactive quality plot. Based on this plot, what values would you choose for --p-trunc-len and --p-trim-left for both the forward and reverse reads? Why have you chosen those numbers?

- **Reverse reads 0 for trim left and 160 trunc len**
- **Reverse reads 3' good quality until 160 bp**
- **Forward reads 0 for trim left and 190 for trunc len**

Based on the interactive quality plots for both forward and reverse reads, the chosen settings for data trimming and truncation are well-aligned with the observed quality scores. For forward reads, a trimming value of --p-trim-left 0 and a truncation length of --p-trunc-len 190 are selected to use the consistently high-quality scores above 30 across the initial 190 bases. Similarly, for reverse reads, the settings of --p-trim-left 0 and --p-trunc-len 160 are chosen because the quality stays high up to 190 bases, beyond which there is a noticeable decline. These parameters ensure that the sequence data used in subsequent analyses are of high quality, reducing the potential for errors and improving the reliability of the results.





2) How would you modify the code above to truncate and trim in your desired way?

Edit the code to reflex parameters showing best quality regions of the reads. The command is configured to trim at position 0 for both forward and reverse reads and to truncate at 190 bases for forward reads and 160 bases for reverse reads. This setup ensures that only the high-quality portions of the reads are kept, enhancing the reliability and accuracy of the subsequent microbiome analysis.

```
--p-trim-left-f 0 \
--p-trunc-len-f 190 \
--p-trim-left-r 0 \
--p-trunc-len-r 160 \
```

3) In the tutorial, you had to mv the files to rename them to just rep-seqs.qza, table.qza, and stats.qza. How could you modify the above code to skip that step? How do you need to modify qiime metadata tabulate in order to account for the renamed files being generated?

Specify output filenames such as rep-seqs.qza, table.qza, and stats.qza in the qiime dada2 denoise-paired command. Then, ensure the qiime metadata tabulate command references these filenames accurately, thus streamlining the process and reducing potential errors; essentially deleting DADA2 from the code will make it successful.

--o-representative-sequences rep-seqs-dada2.qza \

--o-table table-dada2.qza \

--o-denoising-stats stats-dada2.qza

**--o-representative-sequences rep-seqs.qza **

**--o-table table.qza **

--o-denoising-stats stats.qza

4) Your metadata file has a different name than that in the tutorial. How do you adjust your code in order to use the metadata file you have been given?

Make it a txt file remove "sample" from tutorial code, also remove "sample" from the metadata default. To accommodate a metadata file with a different name and format in your QIIME2 analysis, first convert the file from Excel to a Unicode text file (.txt), then update your command line references by removing "sample" and specifying the new file name, metadata.txt.

sampleid	population	sex	timepoint	bird	flock	
#q2:types	categorical	categorical	categorical	categorical	categorical	categorical
4_S157_L001	resident	female	B	BIRD22	residentfemale	
13_S95_L001	resident	male	B	BIRD19	residentmale	
18_S71_L001	resident	male	B	BIRD16	residentmale	
61_S109_L001	migratory	female	B	BIRD03	migratoryfemale	
78_S46_L001	migratory	male	B	BIRD12	migratorymale	
84_S61_L001	migratory	female	B	BIRD06	migratoryfemale	
100_S359_L001	migratory	female	B	BIRD01	migratoryfemale	
104_S93_L001	migratory	male	B	BIRD09	migratorymale	
106_S98_L001	resident	female	B	BIRD23	residentfemale	
125_S13_L001	migratory	female	B	BIRD02	migratoryfemale	
128_S36_L001	resident	male	B	BIRD18	residentmale	
163_S60_L001	migratory	female	B	BIRD04	migratoryfemale	
168_S37_L001	migratory	female	B	BIRD05	migratoryfemale	
174_S146_L001	resident	female	B	BIRD24	residentfemale	
189_S23_L001	resident	male	B	BIRD17	residentmale	
212_S94_L001	resident	male	B	BIRD15	residentmale	
245_S122_L001	resident	female	B	BIRD26	residentfemale	
254_S69_L001	migratory	male	B	BIRD08	migratorymale	
265_S133_L001	resident	female	B	BIRD21	residentfemale	
307_S70_L001	migratory	male	B	BIRD11	migratorymale	
309_S47_L001	resident	male	B	BIRD20	residentmale	
364_S22_L001	migratory	male	B	BIRD10	migratorymale	
366_S45_L001	migratory	male	B	BIRD07	migratorymale	
385_S170_L001	resident	female	B	BIRD25	residentfemale	

5) Include a screenshot of the table summary from visualizing your table and a screenshot of the sequence length statistics from the rep-seqs file.

The first graph is individual birds, the second graph represents the population.

Rep seq file

Sequence Length Statistics

Download sequence-length statistics as a TSV

Sequence Count	Min Length	Max Length	Mean Length	Range	Standard Deviation
5521	190	337	250.89	147	9.2

Seven-Number Summary of Sequence Lengths

Download seven-number summary as a TSV

Percentile:	2%	9%	25%	50%	75%	91%	98%
Length* (nts):	221	251	253	253	253	253	254

*Values rounded down to nearest whole number.

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.

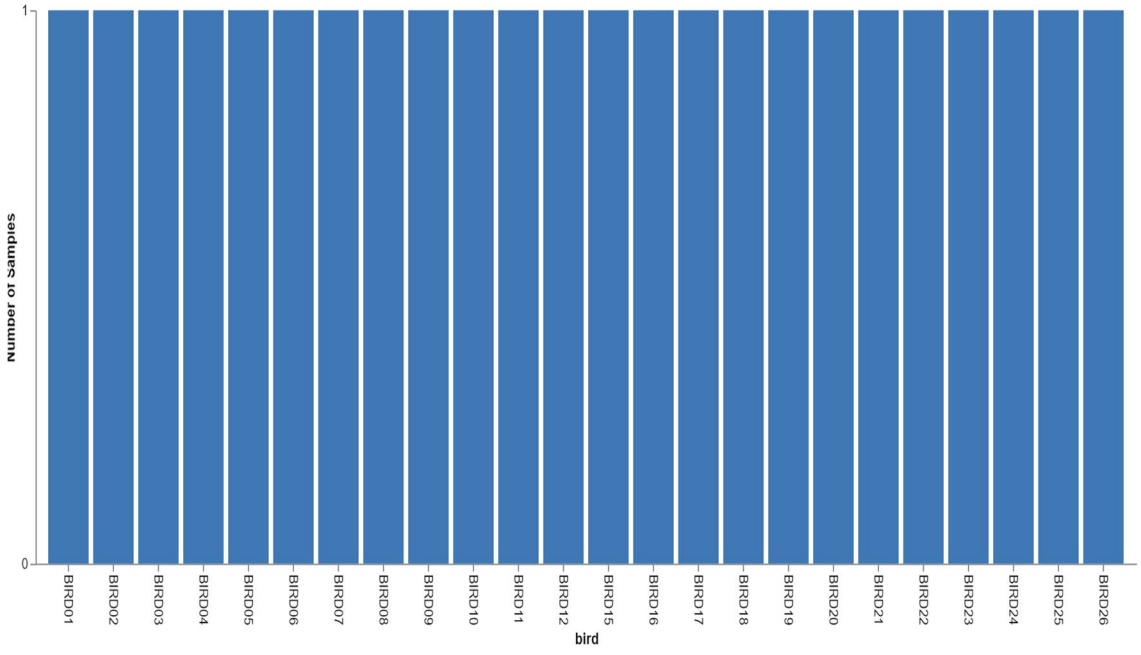
Download your sequences as a raw FASTA file

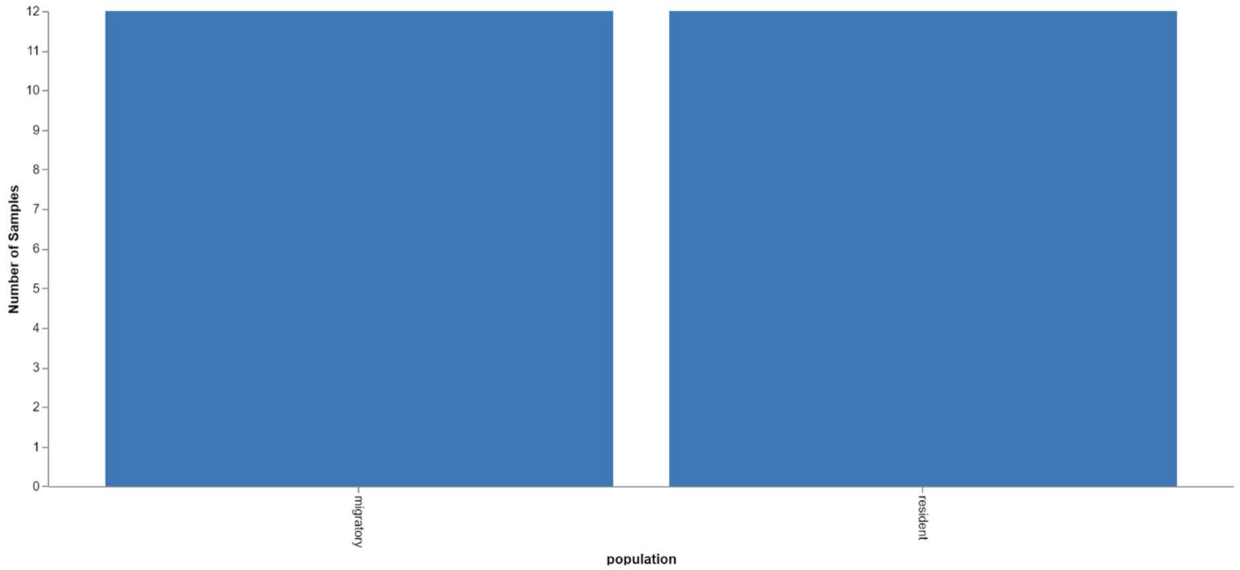
Click on a Column header to sort the table.

Feature ID	Sequence
e76eaf48ba6fbb414f386b0c384c6f67	TACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCGCGTAGGCGGACAGTTTAGTCAGAGGTGAAAGCCCAGGGCTCAACCTTGGAATAGCCTTTGAAACTGACTGCTTGAGTACGGGAGAGGTGTGTGGAACTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGAACACCGGTGGCGAAGGCGACACACTGGCCCGTTACTGACGCTGA
486c89fe8aaddf66fb9f27a49b5eda93	TACGTATGGTGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGCAGGCGGTACGGCAAGTCTGATGTGAAAGCCCGGGGCTCAACCCCGGTACTGCATTGGAAACTGTCGAACTGGAGTGTCGGAGGGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACGATTACTGACGCTGA
d2af52fc6f48ab4d05c96dad9527d8a2	TACGTAGGGTGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAACTGGAGGCTCAACCTCCAGCCTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGATTGGAATTCCTGGTGTAGCGGTGGAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGATCTCTGGGCCGTAACTGACGCTGA
44c21e29adae97a53247abbd73978395	TACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTAGTTAAGTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATGGCGATGGATACTGGCTGGCTAGAGTGTGTCAGAGGATGGTGGAATTTCCGGTGTAGCGGTGAAATGCGTAGAGATCGGAAGGAACATCAGTGGCGAAGGCGGCCATCTGGGACAACACTGACGCTGA
5005073a496b374b2352dea23782082c	TACGTATGTCCCAAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGACGGTTATTTAAGTCTGAAGTGAAAGCCCTCAGCTCAACTGAGGAATTGCTTTGGAAACTGGATGACTTGAGTGCAGTAGAGGAAAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACGTTGA
2c87d01f359e7d77cbdf931e0868986a	TACGGAGGGTCCAAGCGTTATCCGGAATCATTGGGTTTAAAGGGTCCGTAGGCGGTTTAATAAGTCAGTGGTGAAAGCCCATCGCTCAACGATGGAACGGCCATTGATACTGTTAAACTTGAATTATTAGGAAGTAACTAGAATATGTAGTGTAGCGGTGAAATGCTTAGAGATTACATGGAATACCAATTGCGAAGGCAGGTTACTACTAATATATTGACGCTGA
c67aa044b45a996c17e3b83bbbc3c075	TACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATACTTAAGTCAGGGGTGAAATCCCGAGGCTCAACCTCGGAACTGCCTTTGATACTGGGTATCTTGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACGCTGA
f7447c8e079023f4f2579d8580575dd3	TACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGTGCAGGCGGCTCGATAAGTCTGATGTGAAAGCCTTCGGCTCAACCGGAGAATTGCATCAGAAACTGTCGAGCTTGAGTACAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAACACCGGTGGCGAAGGCGGCTCTCTGGTCTGTTACTGACGCTGA

Table summary

Metric:	Sample
Number of samples	24
Number of features	5,521
Total frequency	614,415





Feature ID	Sequence
e76eaf48ba6fbb414f386b0c384c6f67	TACGAAGGGGGCTAGCGTTGCTCGGAATTACTGGGCGTAAAGGGCGCGTAGGCGGACAGTTTAGTCAGAGGTGAAAGCCCAGGGCTCAACCTTGGAATAGCCTTTGAAACTGACTGTCTTGAGTACGGGAGAGGTGTGTGGAACTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGAACACCGGTGGCGAAGGCGACACACTGGCCCGTTACTGACG
486c89fe8aaddf66fb9f27a49b5eda93	TACGTATGGTGCAAGCGTTATCCGGATTTACTGGGTGTAAAGGGAGCGCAGGCGGTACGGCAAGTCTGATGTGAAAGCCCGGGGCTCAACCCCGGTACTGCATTGGAAACTGTCGAACTGGAGTGTCGGAGGGGTAAGTGGAATTCCTAGTGTAGCGGTGAAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGGCTTACTGGACGATTACTGACG
d2af52fc6f48ab4d05c96dad9527d8a2	TACGTAGGGTGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGAGCTCGTAGGCGGTTTGTCGCGTCTGCTGTGAAAACTGGAGGCTCAACCTCCAGCCTGCAGTGGGTACGGGCAGACTAGAGTGCGGTAGGGGAGATTGGAATTCCTGGTGTAGCGGTGGAAATGCGCAGATATCAGGAGGAACACCGATGGCGAAGGCAGATCTCTGGGCCGTAACTGACG
44c21e29adae97a53247abbd73978395	TACGAAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGGGTGCGTAGGCGGTTAGTTAAGTCTGTCGTGAAATCCCCGGGCTCAACCTGGGAATGGCGATGGATACTGGCTGGCTAGAGTGTGTCAGAGGATGGTGGAATTTCCGGTGTAGCGGTGAAATGCGTAGAGATCGGAAGGAACATCAGTGGCGAAGGCGGCCATCTGGGACAACACTGACG
5005073a496b374b2352dea23782082c	TACGTATGTCCCAAGCGTTATCCGGATTTATTGGGCGTAAAGCGAGCGCAGACGGTTATTTAAGTCTGAAGTGAAAGCCCTCAGCTCAACTGAGGAATTGCTTTGGAAACTGGATGACTTGAGTGCAGTAGAGGAAAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAACACCAGTGGCGAAGGCGGCTTTCTGGACTGTAACTGACG
2c87d01f359e7d77cbdf931e0868986a	TACGGAGGGTCCAAGCGTTATCCGGAATCATTGGGTTTAAAGGGTCCGTAGGCGGTTTAATAAGTCAGTGGTGAAAGCCCATCGCTCAACGATGGAACGGCCATTGATACTGTTAAACTTGAATTATTAGGAAGTAACTAGAATATGTAGTGTAGCGGTGAAATGCTTAGAGATTACATGGAATACCAATTGCGAAGGCAGGTTACTACTAATATATTGACG
c67aa044b45a996c17e3b83bbbc3c075	TACGAAGGGGGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGATACTTAAGTCAGGGGTGAAATCCCGAGGCTCAACCTCGGAACTGCCTTTGATACTGGGTATCTTGAGTCCGGAAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAGGAACACCAGTGGCGAAGGCGGCTCACTGGTCCGGTACTGACG
f7447c8e0790234f2579d8580575dd3	TACGTAGGTGGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGAGTGCAGGCGGCTCGATAAGTCTGATGTGAAAGCCTTCGGCTCAACCGGAGAATTGCATCAGAAACTGTCGAGCTTGAGTACAGAAGAGGAGAGTGGAACTCCATGTGTAGCGGTGAAATGCGTAGATATATGGAAGAACACCGGTGGCGAAGGCGGCTCTCTGGTCTGTTACTGACG
60ecd2559d8e8cdd1226aab6b52bdffd	TACGAGGGGGGCAAGCGTTGTTCGGAATTATTGGGCGTAAAGGGTGCGTAGGCGGTTTGATAAGTCTTGTGTGAAATCTATGGGCTCAACCCATAGTCTGCATAGGAAACTGTCGGCTTGAGAGTGGGAGAGGTGAGTGGAATTTCCGGTGTAGCGGTGAAATGCGTAGATATCGGAAGGAACACCTGTGGCGAAAGCGGCTCACTGGACCATTTCTGACG
806da84a4d3359c41378330fbb8a2cf63	TACAGGGGGGGCAAGCGTTGTTCGGAATTACTGGGCGTAAAGGGTTCGTAGGTGGCCACTTAAGTCAGACGTGAAATCCCTCAGCTTAACTGGGGAACTGCGTCTGATACTGAGTGGCTTGAGTGCAGGAGAGGAATGCGGAATTCCAGGTGTAGCGGTGAAATGCGTAGATATCTGGAGGAACACCGGTGGCGAAGGCGGCATTCTGGACTGTAACTGACA
6ab5af96116253dfc6e460fa7cc61e65	TACGTAGGACCCGAGCGTTGTCCGGATTTACTGGGTATAAAGGGTGCGTAGGCGGTCTTGTGCGTCAGAGGTGAAATCTCCGGGCTTAACCCGGATGGTGCCTTTGATACGGCAGGACTTGAGTGCGGGAGAGGGTGATGGAATTCCTGGTGTAGCGGTGAAATGCGTAGATATCAGGAGGAACACCGGTGGCGAAGGCGGTCACCTGGCCCGTAACTGACG
f5a79df0019ecf61cb142b8562eb3a08	TACGTAGAAGACTAGTGTTATTCATCTTTAGTAGGTTTAAAGGGTACCTAGACGGTAAAATTAGTCCAAAACTGGATACAGTTTTACTAGAGTTATATATGAGAAGGTGAGTATTTAAGGAGTAGAGTTGAAATTCTTTGATACCTTAACGACTGGTAACGGCGAAAGCAACCTTTTATCTAATAACTGACGTTGAGGGACGAAGGCTTGGGTAGCGAACAG
f512ea8083e902e0a1e3e8a5bbca4f3b	TACGGAGGGTGCAAGCGTTGTCCGGATTTATTGGGTTTAAAGGGTGCGTAGGTGGCCTTCTAAGTCTGGTTTGAAAGTAGGTCGCTTAACGATCTAATGTGGCTGGAAACTGGGGGGCTTGAATGATTTGGCGGTAGCCGGAACGGGTCATGTAGCGGTGAAATGCATAGATATGACCCAGAACACCGATTGCGAAGGCAGGCTACTACGAGTTGATTGACA
0cc951954a255c148a5be384e2d5fe9c	TACAGAGGGTGCAAGCGTTAATCGGAATTACTGGGCGTAAAGCGAGCGTAGGTGGCTTGATAAGTCAGATGTGAAATCCCCGGGCTTAACCTGGGAACTGCATCTGATACTGTCAGGCTAGAGTAGGTGAGAGGAAGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGATGGCGAAGGCAGCCTTCTGGCATCATACTGACA

6) Jump down to taxonomy. Once you have generated your taxonomy visualization, sort it by confidence. What are your top hits?

Confidence table

284687df97108824c1f6a3d57668cba0	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	1.0000000000000007
f52cc1224cac807cd42d9a901ed53c9b	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	1.0000000000000049
a52970705dfd5642085748f4646b3b18	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	1.0000000000000013
1b87abf52962fba31cf34d77729d1b4c	k__Bacteria; p__Planctomycetes; c__Planctomycetia; o__Gemmatales; f__Isosphaeraceae	1.0000000000000002
025f8d52bf0fe507d3355b96f306b99b	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__Runella; s__	1.0
1524b090dbd05a604d54ed8c2ced1acb	k__Bacteria; p__Fusobacteria; c__Fusobacteriia; o__Fusobacteriales; f__Leptotrichiaceae; g__Leptotrichia; s__	1.0
1e781834af10358b3ddaaecde4c15959	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Anaerococcus; s__	1.0
245b2aa3f0304ce3a64102493b4261a1	k__Bacteria; p__Bacteroidetes; c__Cytophagia; o__Cytophagales; f__Cytophagaceae; g__Emticicia; s__	1.0
26ce2cd3e526fbd0a26604d642886a16	k__Bacteria; p__Proteobacteria; c__Deltaproteobacteria; o__MIZ46; f__; g__; s__	1.0
317c1c80d1c49b3857282ad38125cb7f	k__Bacteria; p__Verrucomicrobia; c__[Methylacidiphilae]; o__Methylacidiphilales; f__; g__; s__	1.0
41f3f39a0335a7392afa80327caf0f77	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Anaerococcus; s__	1.0
46e9544e9c5e7ed6d90a7b50a8a9c990	k__Bacteria; p__Firmicutes; c__Clostridia; o__Clostridiales; f__[Tissierellaceae]; g__Anaerococcus; s__	1.0
6db3b7eb4f438072f8fe720b9745ceb6	k__Bacteria; p__Armatimonadetes; c__SHA-37; o__; f__; g__; s__	1.0
74db75e62fcd4c68be66d7a5cb9b55c5	k__Bacteria; p__Verrucomicrobia; c__[Methylacidiphilae]; o__Methylacidiphilales; f__; g__; s__	1.0
7982f10e7208509a8591f1ee1d91ee3b	k__Bacteria; p__Verrucomicrobia; c__Verrucomicrobiae; o__Verrucomicrobiales; f__Verrucomicrobiaceae; g__Akkermansia; s__	1.0
809ce338577e7bcbed5ae3101fea51ff	k__Bacteria; p__Verrucomicrobia; c__[Methylacidiphilae]; o__Methylacidiphilales; f__; g__; s__	1.0
8fd253d51e7aca7bbc43bae8dcd26435	k__Bacteria; p__Fibrobacteres; c__Fibrobacteria; o__258ds10; f__; g__; s__	1.0
2bc5b2b80439ea1819f948224ea6be69	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999978
c0d5395792eadbf5f62e8ffb14fa0262	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999976
ce60a6ff9d1ebb20bc5c414825cd8833	k__Bacteria; p__Planctomycetes; c__Planctomycetia; o__Gemmatales; f__Isosphaeraceae	0.9999999999999956
99108a0151b211c8963bed3a74a06f15	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999953
9011737255ae3168845adf05114808ab	k__Bacteria; p__Proteobacteria; c__Alphaproteobacteria; o__Rickettsiales; f__mitochondria	0.9999999999999947

7) What do you think this code is doing? Why do you think this is a necessary or important step?

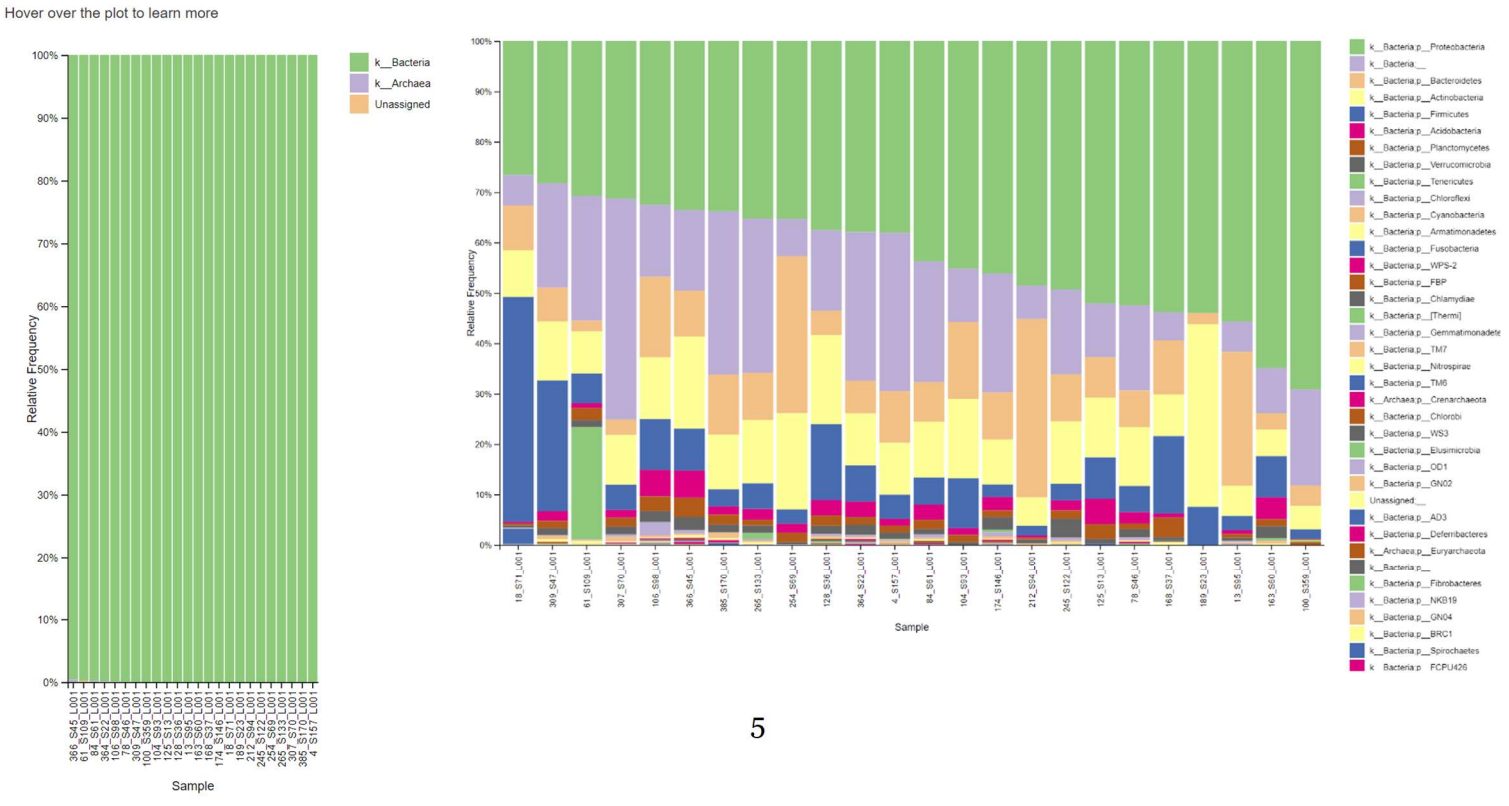
```
qiime taxa filter-table \
 --i-table table.qza \
 --i-taxonomy taxonomy.qza \
 --p-exclude mitochondria,chloroplast \
 --o-filtered-table table.qza
```

The code provided uses the QIIME2 qiime taxa filter-table command to filter out mitochondrial and chloroplast DNA from the analysis table, based on the taxonomy classifications. This step is crucial as it enhances the specificity and accuracy of microbial community analysis by focusing solely on bacterial (prokaryotic) sequences, thereby eliminating non-bacterial DNA which could skew the results and interpretations of the microbial ecosystem being studied.

8) Re-do your table visualization and re-do your taxonomy commands. Do you have any differences now in the hits with the highest confidence? Why or why not? Really think about what the code is doing.

After filtering out mitochondrial and chloroplast DNA and redoing the table visualization and taxonomy commands, there were no differences in the hits correlated to the highest confidence. This is because the filtering process only affects the abundance table and taxonomy classification, not the underlying sequence data in the representative sequences file. Thus, any high-confidence mitochondrial or chloroplast sequences still appear in results if not explicitly excluded from all data files used in the analysis. The end result did not remove mitochonial DNA from our rep seq file only from our taxonomy and table.

9) Looking at taxa bar plots, what are your top 2 phyla? Include a screenshot. What are the top 5 most abundant classes? Include a screenshot. (Color bars level 1 and level 2)



10) What is the difference between alpha and beta diversity? You will have to read outside resources to answer this question. Your response should be in your own words.

- **Alpha diversity refers to the diversity within a single ecological community or sample, measuring the variety and abundance of species present. It often includes metrics like species richness, evenness, and the Shannon diversity index, providing insights into the ecological complexity within a single habitat.**
- **Beta diversity, on the other hand, compares the diversity between different communities or ecosystems. It measures the difference in species composition across environmental gradients or geographic locations, highlighting how distinct the communities are from each other. Metrics like Bray-Curtis dissimilarity and Jaccard index are commonly used to quantify these differences, offering a view of species turnover or shared species between habitats.**

Together, alpha and beta diversity contribute to a comprehensive understanding of biodiversity across studies and populations.

11) Before you calculate your diversity metrics, you have to choose a sampling depth. What file previously generated will you use to help you determine what to choose? Defend your choice of sampling depth. How many samples do you retain and how many do you lose?

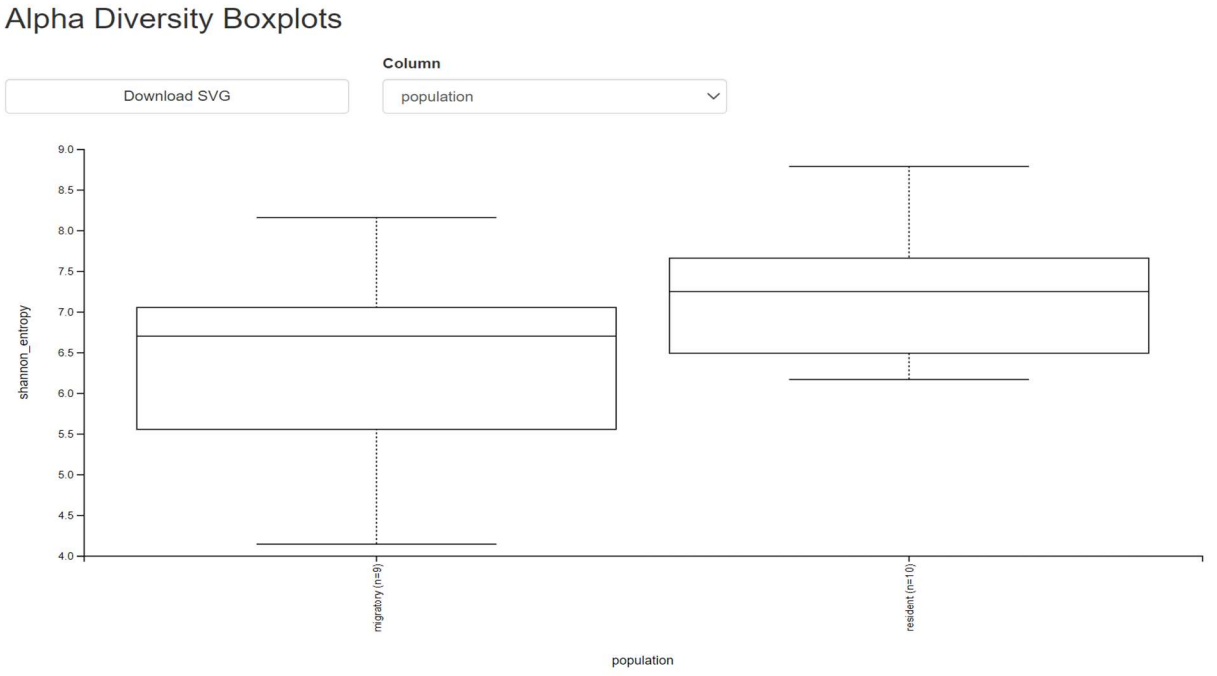
To calculate diversity metrics accurately, I will use a sampling depth of 10,000 based on the sequence data in the table.qzv file, which provides a summary of sequence counts per sample. This depth ensures high-quality data by excluding samples with fewer than 10,000 sequences, thus minimizing the impact of low-coverage data on the analysis. By setting this threshold, I kept 28 samples for in-depth analysis and exclude only 6, effectively balancing data quality with the number of samples analyzed.

Sample ID	Frequency
106_S98_L001	65,047
13_S95_L001	50,303
174_S146_L001	46,672
4_S157_L001	46,157
364_S22_L001	44,702
307_S70_L001	36,809
18_S71_L001	36,618
309_S47_L001	35,816
78_S46_L001	34,625
128_S36_L001	33,524
212_S94_L001	32,387
366_S45_L001	27,170
61_S109_L001	20,540
163_S60_L001	19,460
385_S170_L001	19,311
245_S122_L001	18,620
84_S61_L001	11,072
254_S69_L001	10,101
168_S37_L001	8,991
265_S133_L001	6,858
100_S359_L001	4,100
125_S13_L001	3,620
104_S93_L001	1,436
189_S23_L001	476

12) For alpha diversity, you need to create visualizations for Shannon diversity and Observed features. This will require you to modify the alpha-group-significance code. For which metadata values were graphs generated? Were any of those comparisons significant? How do you know whether they were or were not significant? Briefly describe what Shannon diversity and Observed features are measuring (less than 1 paragraph).

Based on the provided boxplot for Shannon diversity grouped by population, and the accompanying Kruskal-Wallis test results, it's clear that there are no significant differences in alpha diversity across different populations, as indicated by the p-value of 0.1615. This p-value, being greater than 0.05, suggests that any observed variations in Shannon diversity across populations are not statistically significant and could be due to chance.

Shannon diversity is a metric that evaluates both the richness (the total number of different species) and evenness (how evenly the individuals are distributed among those species) within a community. The Observed features metric, on the other hand, merely counts the number of distinct species present in a sample, without accounting for their abundance. Neither showed significant differences across the groups defined by sex, population, and flock, indicating uniformity in species richness and distribution patterns among these groups within the studied community.

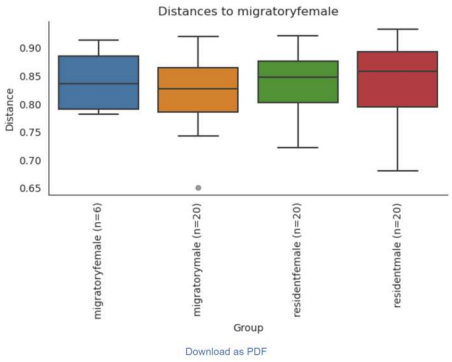


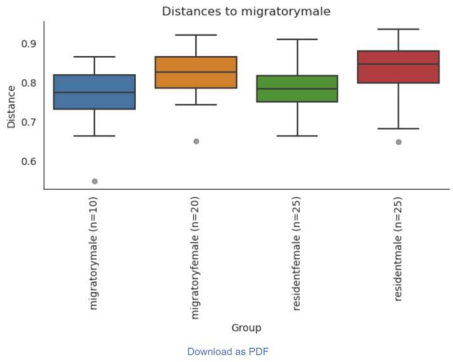
Kruskal-Wallis (all groups)

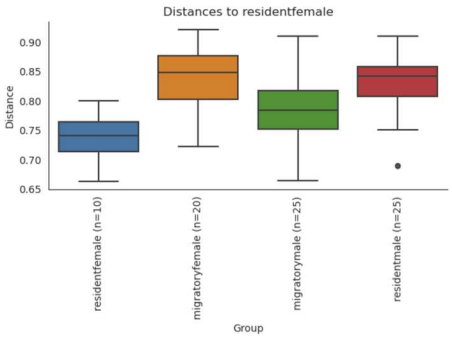
	Result
H	1.9266666666666694
p-value	0.1651235901283899

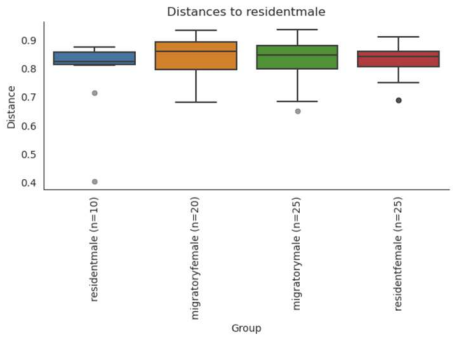
13) For beta diversity, you will need to create visualizations for Bray Curtis dissimilarity. This will require your to modify the beta-group-significance code. You should have one visualization for sex, one for population, and one for flock. Include a screenshot of each visualization. Is there any significance? Regardless of significance, how can you interpret these results (hint: what is beta diversity looking at?)

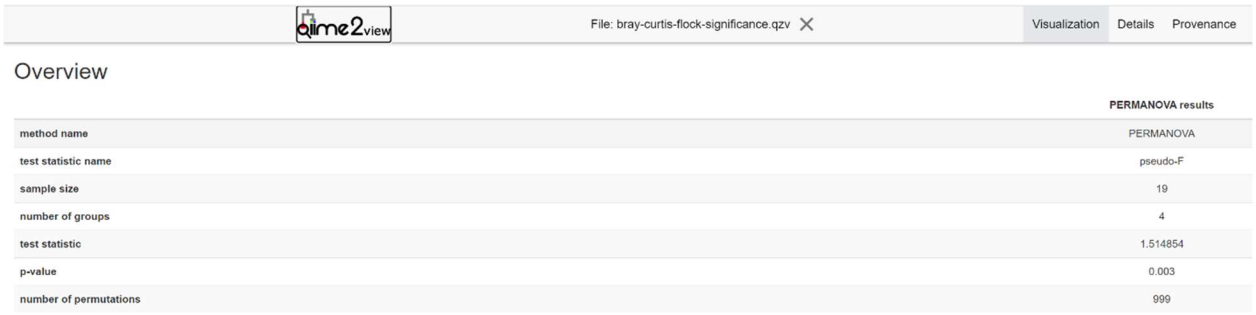
The Bray-Curtis dissimilarity visualizations show varied microbial community compositions across groups based on sex, population, and flock. The statistical analysis reveals that differences in microbial communities between sexes (p=0.167) and between migratory and resident populations (p=0.994) are not statistically significant, suggesting that these factors do not strongly influence microbial diversity. However, differences between flocks are statistically significant (p=0.003), indicating that the flock environment significantly affects microbial community structure





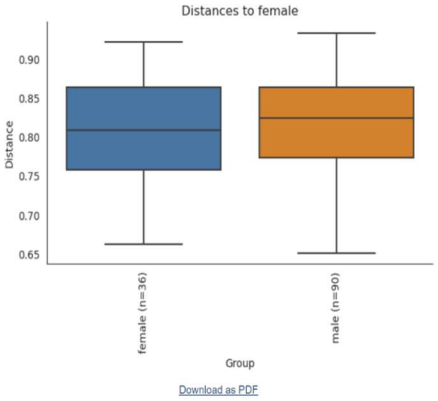




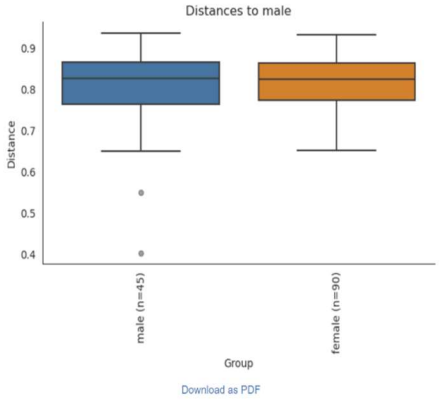


Group significance plots

Download raw data as TSV



Download as PDF



Download as PDF

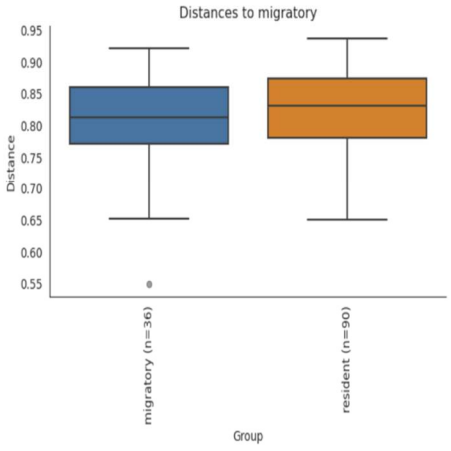
qiime2view File: bray-curtis-sex-significance.qzv Visualization Details Provenance

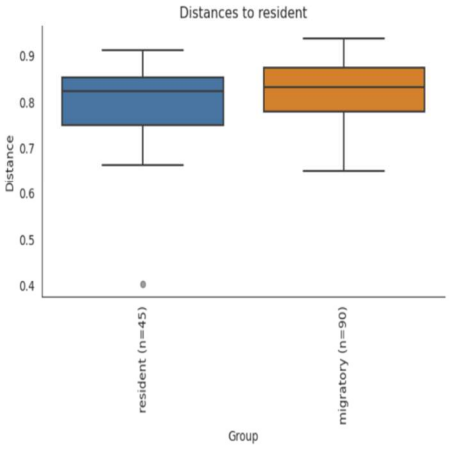
Overview

	PERMANOVA results
method name	PERMANOVA
test statistic name	pseudo-F
sample size	19
number of groups	2
test statistic	1.168064
p-value	0.167
number of permutations	999

Group significance plots

Download raw data as TSV





qiime2view File: bray-curtis-population-significance.qzv Visualization Details Provenance

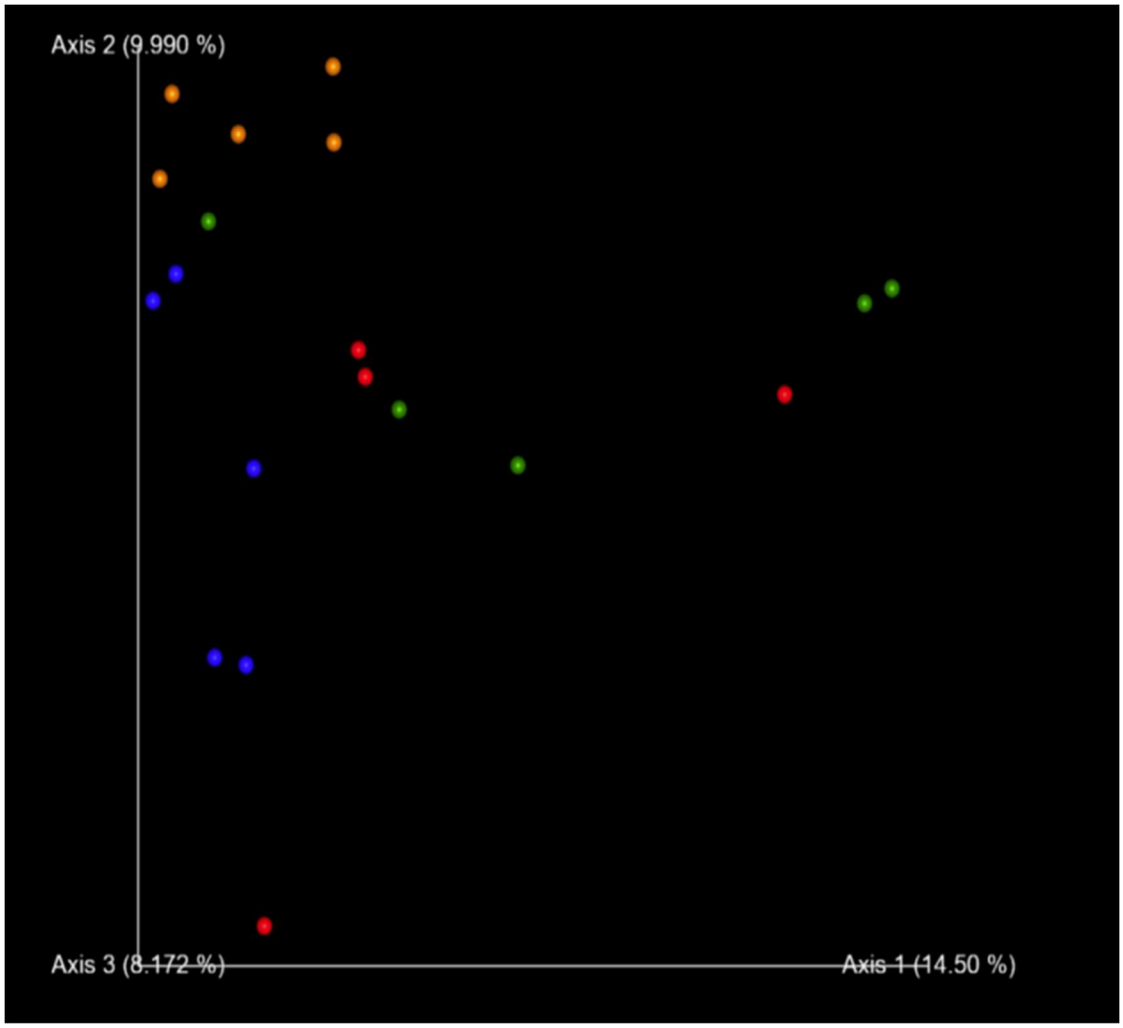
Overview

	PERMANOVA results
method name	PERMANOVA
test statistic name	pseudo-F
sample size	19
number of groups	2
test statistic	1.435619
p-value	0.034
number of permutations	999

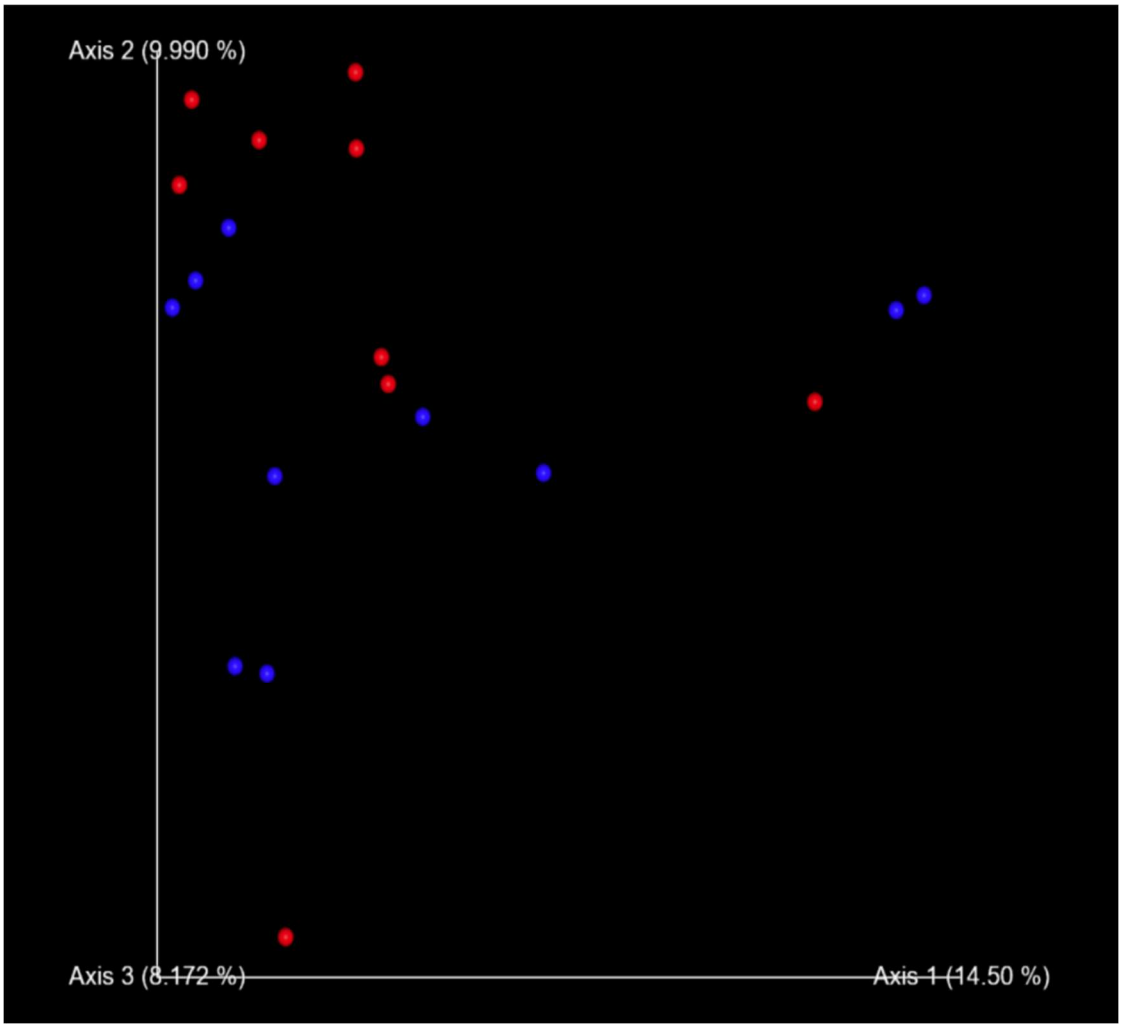
14) The core-metrics-phylogeny command generates a file called bray-curtis-emporer.qzv. Include 3 screenshots total (1 where the points are colored based on sex, one on population, one on flock). How do these results help you make sense of the results you got from question 13?

The Emperor PCoA plots from the bray-curtis-emporer.qzv file, colored by sex, population, and flock, visually reinforce the statistical outcomes from the PERMANOVA tests. The plots show no significant clustering by sex (p=0.167) or population (p=0.994), showing that these factors do not influence microbial community structures significantly. In contrast, the significant clustering by flock (p=0.003) visually shows that flock-specific environmental or social factors have a substantial impact on shaping microbial communities, confirming the statistical analysis results.

Population



Flock



Sex