MICB 475 Lab Notebook
JAAFV4723

Team 12

Jas Aujla
Anu Chauhan
Ally Howard
Fatima Nadeem
Vera Pu

TABLE OF CONTENTS

P002	JAAFV4723-039	PICRUSt2 Visualization with R	14-Nov-2023	039
P002	JAAFV4723-40	Diversity Metrics Analysis in R - Continued	15-Nov-2023	040
P002	JAAFV4723-41	PICRUSt2 MetaCyc Visualization with R	19-Nov-2023	041-043
P002	JAAFV4723-044	Diversity Metrics Analysis in R - Continued	20-Nov-2023	044-045
P002	JAAFV4723-046	Diversity Metrics Analysis in R - Continued	22-Nov-2023	046-047
P002	W12-TM	Week 12 Team Meeting	22-Nov-2023	048-049
P002	JAAFV4723-50	Core Microbiome R Processing	22-Nov-2023	050-051
P002	JAAFV4723-052	Indicator Species Analysis R Processing Continued	22-Nov-2023	052
P002	JAAFV4723-053	MetaCyc error bar editing	24-Nov-2023	053-054
P002	JAAFV4723--055	Indicator Species Analysis R Visualization	26-Nov-2023	055-056
P002	W13 - TM	Week 13 Team Meeting	29-Nov-2023	057-058
P002	JAAFV4723-059	Diversity Metrics Analysis in R - Continued	30-Nov-2023	059-60

Project Codes

P001 - Analysis of soil dataset to model QIIME 2 pipeline

P002 - Research project on soil dataset

START

W4-TM: Week 4 Team Meeting

Agenda

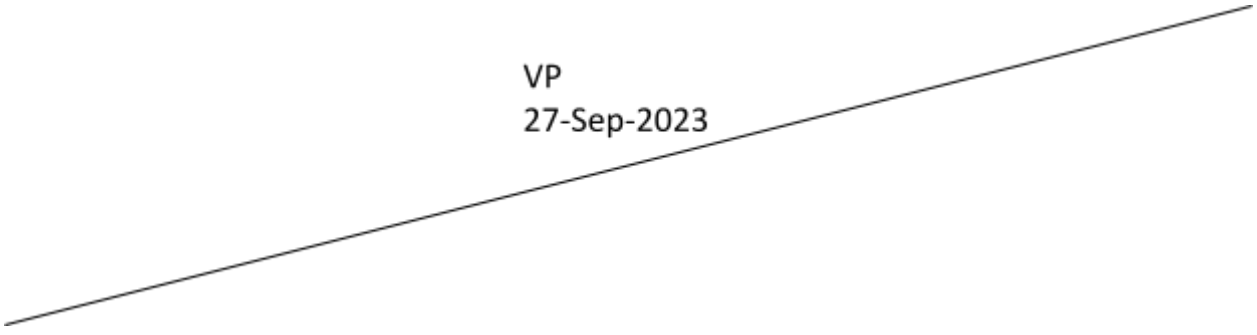
1. Introduction and coding experience
2. Exploration of research ideas and metadata sets
3. Selection of collaboration tool for lab notebook

Meeting Minutes

- Everyone has some level of coding knowledge from past coursework.
- Dataset Consideration: Zoo Metadata
    - Research focus: Evaluating animal health by comparing weights and analyzing microbiome data.
    - Approach: categorization of animals - create a structured list of categories for the captive and non-captive animals in the dataset.
- Collaboration Platform: GitHub
    - Assigned a designated member to take responsibility for recording meeting minutes.
    - Meeting agenda must be submitted to Avril by Tuesday midnight for review.
    - Lab notebook should be updated regularly throughout the project.

Action Items

- Prepare agenda for the next group meeting
- Read over the zoo metadata and associate paper
- Generate potential research questions and share
- Review proposal and lab notebook module
- Create GitHub account



Date: 27-Sep-2023 Signature/Initial: VP END

START

W5-TM: Week 5 Team Meeting

Agenda

1. Review previous meeting minutes
2. Share research questions and finalize the topic (possible questions below)
    a. Does country of captivity have an effect on the gut microbiome of captive animals
    b. Is body mass between wild and captive animals correlated with changes in gut microbiome
    c. How is the gut microbiome of mammalian animals affected by their habitat type (terrestrial or arboreal) and captivity status (captive or wild) → compare bacterial diversity across the four combinations
3. Notebook formatting

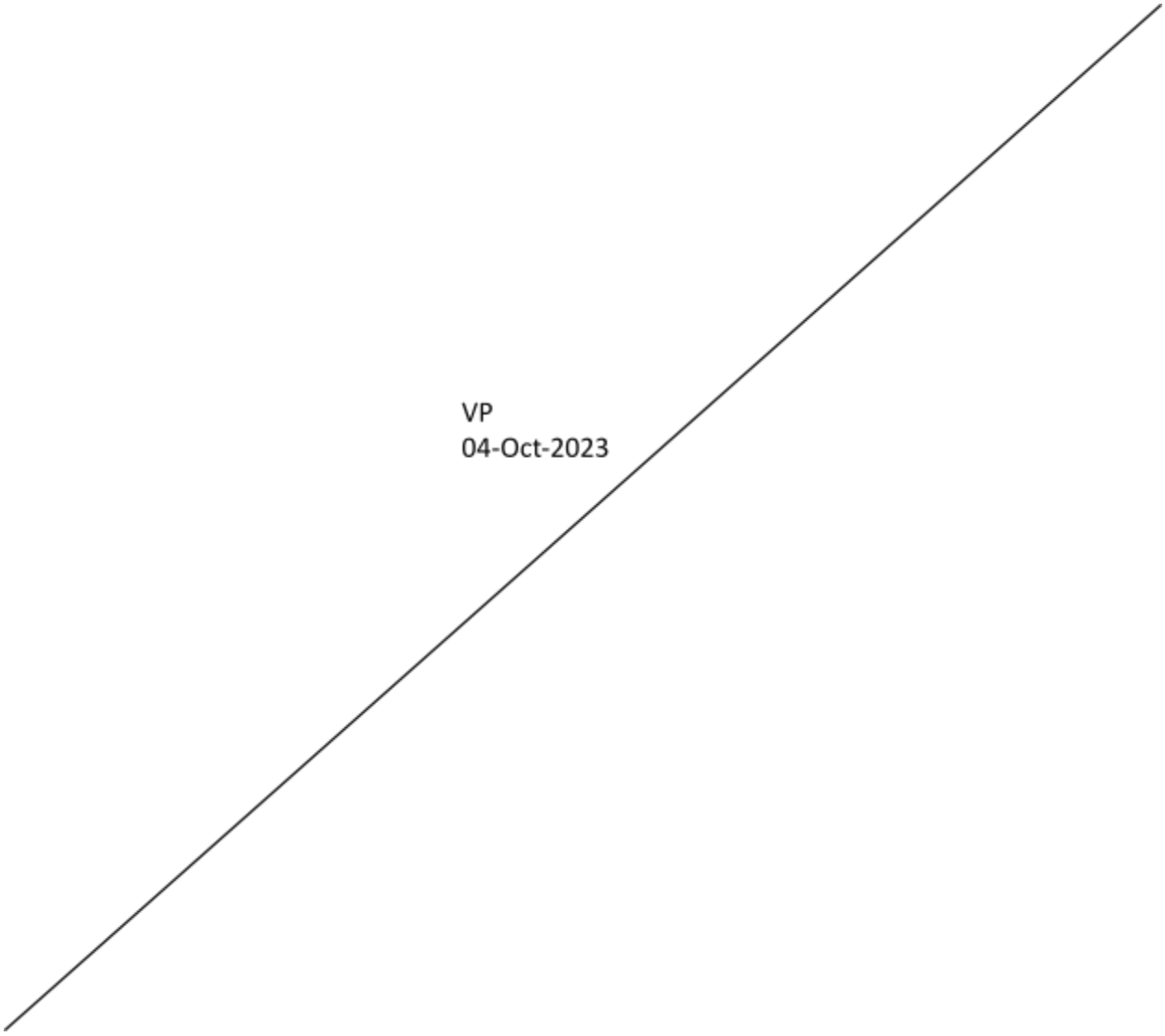
Meeting Minutes

- Zoo metadata is referenced wrongly (e.g., body mass, age of birth)
    - Reach out to the researcher for clarification → takes a long time
- Switching to new datasets
    - Anemia & infant feeding
        - Concerns: another group is already doing these two metadata, overlap between the two sets is limited
    - Tanzania
    - **Soil samples following logging & in the wetland**
        - Ideas: similar bacterial diversity and environmental composition? 'Tree cover' - filter specific tree types and investigate the soil types
        - **Possible focus: looking at differences between groups that have Douglas firs in BC but with different soil types, moisture content in the microbiome**
            - Try to get stuff from the same location
            - Look through old papers to see what has been done before
            - Beta-diversity plot to check if the latitude variable differs a lot
- Notebook format - digital is fine
- Lab proposal
    - 3 aims - alpha & beta diversity, differential abundance (statically differences in bacteria vs. community), indicator taxa (smaller amount of bacteria with strong differences)

- Check additional resources for other interesting analysis types (picrust, GO enrichment, Kegg terms)
- Can switch research direction after the proposal is done

Action Items

- Share one credential so we can all work on the same QIIME server
- review soil metadata and propose a list of new research questions and email to Avril



Date: 04-Oct-2023 Signature/Initial: VP END

START

W6-TM: Week 6 Team Meeting

Agenda

1. Discuss finalized project question (Do microbiomes of logged Douglas fir stands in BC differ based on soil type and moisture content?)
2. Discuss denoising step (and whose credentials to use for QIIME2 work)
3. Come up with 3 objectives for our proposal
4. Decide tasks for proposal

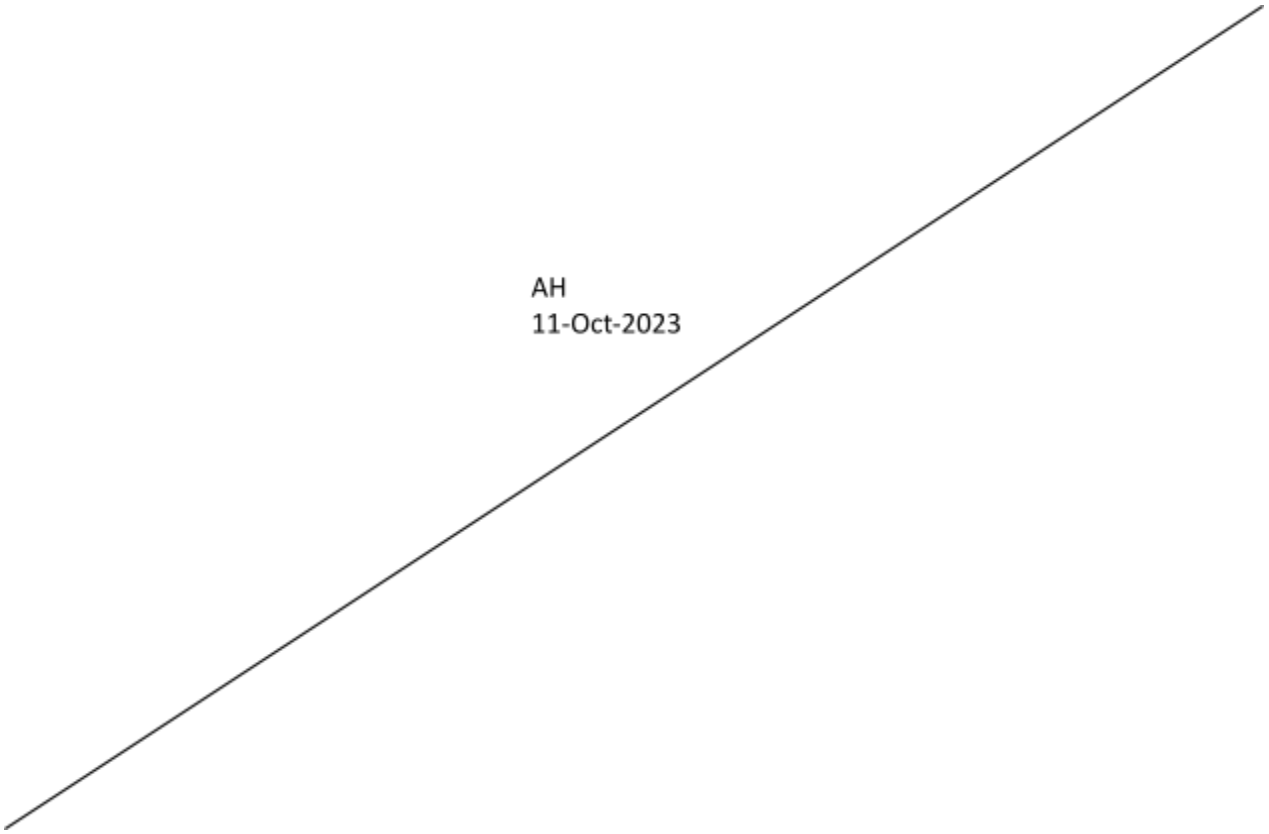
Meeting minutes

- Confirmed finalized project question (double check UJEMI papers to see if duplicates)
    - Checked if moisture content was a continuous variable (1-100, percentage)
- Objectives for proposaL
    - Look at alpha a beta diversity (check if one soil type is more diverse than another)
    - Differential abundance (looking for any stats difference between two values)
        - Not doing this one for our proposal
    - Indicator taxa (looking for one really abundant / prevalent group, smaller number of species
        - Might miss out on some differences that would be seen with differential abundance
        - Used more in ecology
        - Wouldn't typically use Picrust (DSeq instead) for this (but this might be a cool route to go down)
        - Use this one for objective 2 of our proposal
    - Picrust (gives Kegg data, pathway data, enzyme data all based on taxonomy data)
- Proposal
    - Approach chart (specify when using QIIME vs R) --> colour code with key
    - Be specific about what kind of diversity metrics you are evaluating
    - Be specific over all (mention indicator species package, stats)
        - Alpha diversity (boxplots)
            - Soil types --> use wilcoxin test (t-test)
            - Moisture --> spearman correlation (linear model, pearsons correlation coefficient)
            - Look over module 14 for this^^
        - Beta diversity --> permanova (module 15)

        - Indicator species and picrust automatically generate stats (write "use internal stats")
    - Write up picrust for indicator species (Avril will talk to Evelyn about this)
    - Background stuff doesn't talk about OUR project, just past work and gap
    - Research objective
        - Slightly more specific ("these variables are known to be related")
    - Experimental aims
    - Specific papers that are relevant to our project
    - Touches on how we'll do our project
    - Most narrow (aside from our approaches table)

Action Items

- Double check UJEMI papers
- Share QIIME credentials
- Decide tasks for proposal
- Start on proposal



Date: 11-Oct-2023 Signature/Initial: AH END

START

JAAFV4723-009: Initial data processing with QIIME 2

Purpose:
To generate features tables for downstream analysis, this experiment will import the soil dataset and perform demultiplexing, denoising and ASV determination, assign taxonomic information to the ASVs, and filter the dataset to remove mitochondria and chloroplast data, and create a filtered table based on tree cover for douglas fir.

Materials:
Laptop with Mac iOS

Methods:
See project2 … script for commands used in this experiment.

1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
    a. ssh root@10.19.139.169
    b. Biome2491
2. Create a directory for working with the current project's data:/data/project_2/
3. Import soil dataset to the using the manifest file to create a qza file with demultiplexed samples (output = demux.qza).
4. Create a visualization of the demultiplexed samples (output = demux.qzv).
5. Copy file to GitHub repository
6. View the demux.qzv file using the QIIME 2 View website (https://view.qiime2.org/) and determine an appropriate read truncation length for quality control after looking at the interactive quality plot.

Results and Observations:

Table 1. *Demultiplexed sequence count summary. Data was generated from the demux.qzv file using QIIME 2 View (https://view.qiime2.org/). The total number of samples was 697.*

Statistic	Number of forward reads
Minimum	4
Median	13286.0
Mean	13934.548063

Maximum	38473
Total	9712380

Figure 1. *Truncation length was selected to be 150 using the quality plot generated from demux.qzv. QIIME 2 View (https://view.qiime2.org/) noted the following: These plots were generated using a random sampling of 10000 out of 9712380 sequences without replacement. The minimum sequence length identified during subsampling was 30 bases. Outlier quality scores are not shown in box plots for clarity.*

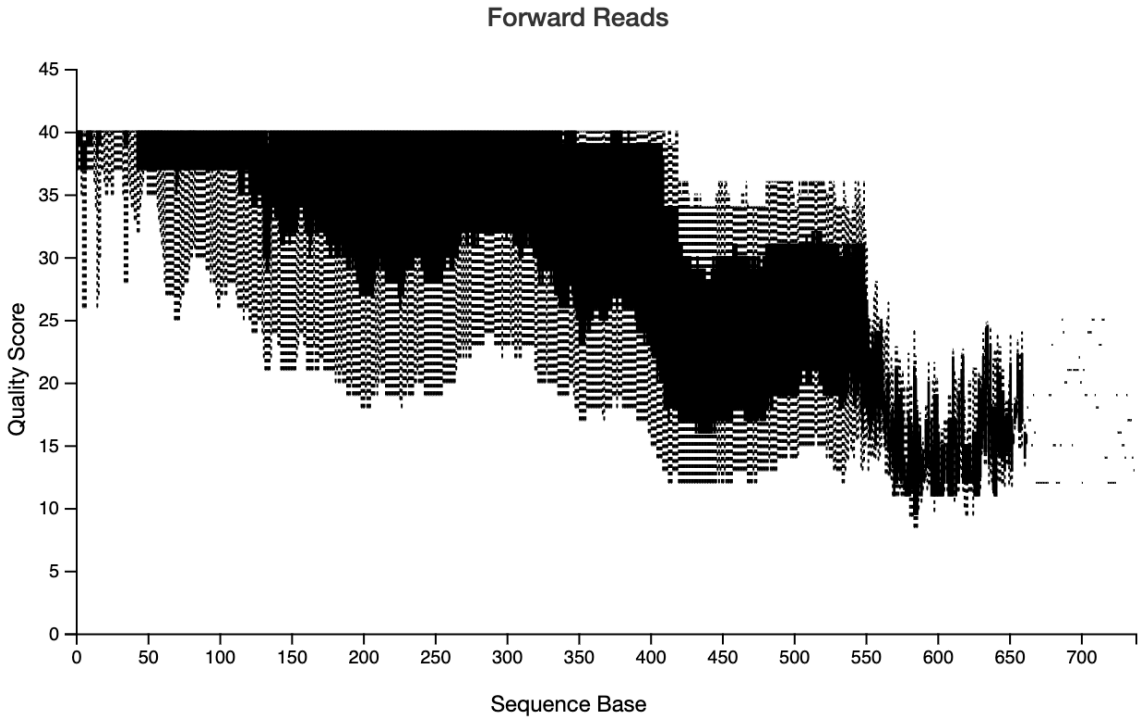


Table 2. *Summary of the demultiplexed sequence length showed that at least 98% of the remaining sequences have a length of greater than or equal to 533 nt. Statistics were generated by visualizing the demux.qzv file at the QIIME 2 View website (https://view.qiime2.org/). The total number of reads sampled randomly to generate this table is 10000 (out of 9712380).*

Percentile	**Sequence Length (nts)**
2%	46
9%	92

25%	340
50% (Median)	438
75%	488
91%	518
98%	533

Table 3. *Parametric seven-number summary of quality scores for bases at position 410. Data from the demux.qzv file was visualized using the QIIME 2 View website (https://view.qiime2.org/). Box plot in the table refers to the sequence base quality plot (Figure 1).*

Box Plot Feature	Percentile	Quality Score
(Not shown in box plot)	2nd	11
Lower Whisker	9th	13
Bottom of Box	25th	18
Middle of Box	50th (Median)	26
Top of Box	75th	34
Upper Whisker	91st	40
(Not shown in box plot)	98th	40

Discussion:

All data processing steps were carried out according to the protocol described in Methods.

The demultiplexed soil dataset included 697 samples and a total of 9712380 forward reads (no reverse reads as the dataset includes single-ended sequences) (Table 1).

After visualizing the demultiplexed soil dataset, we selected a sequence truncation length of 410 nt for the denoising step. This decision was based on the sequence length of the remaining reads and the sequence base qualities. A summary of the length of demultiplexed sequences showed that at least 50% of the reads were estimated to be 438 nt long and at least 98% of the

reads were longer than or equal to 46 nt (Table 2). Selecting a truncation length between 46 and 438 would result in retaining only 50% to 75% of the reads and may result in the loss of a number of samples. Trimming too much would result in the loss of base information which may be suboptimal when aligning reads to reference sequences during taxonomic classification. Based on this knowledge, we decided to trim sequences to 410 nt. This allows us to retain at least 98% of the reads and reduce the loss of samples (Table 2). Additionally, the truncation length of 410 nt ensures acceptable sequence base qualities. No more than 9% of the bases scored lower than 29 at position 150, and no more than 9% of bases scored lower than 27 at any base position between 1 and 149 (Table 3). This means that the chance of the occurrence of a base call error at each position in the 150 nt sequence is less than 1%, which is sufficient for the current study.

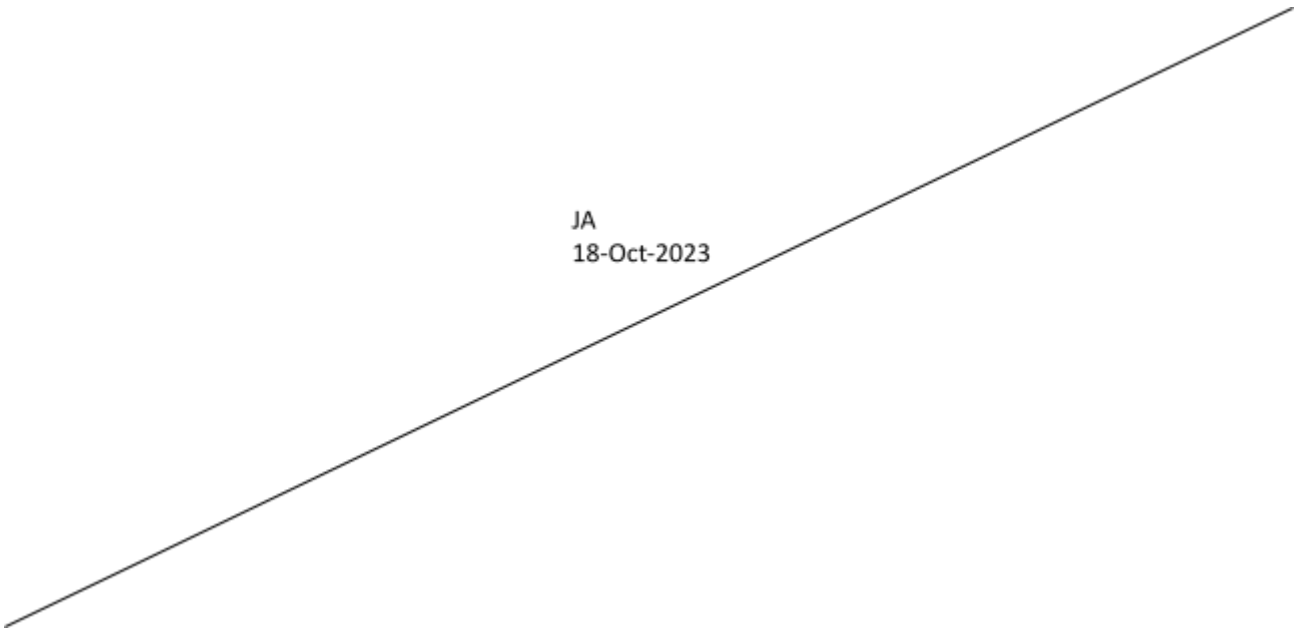
Conclusion:

This experiment completed the initial data processing steps of the project (steps 1-1, 2-1, and 3-1). The following files were generated and can be used for downstream analysis in QIIME 2 or R:

- demux.qza & demux.qzv

Future Directions:

The project will continue on at steps 1-2, 2-2, and 3-2.



Date: 18-Oct-2023 Signature/Initial: JA END

START

W7-TM: Week 7 Team Meeting

Agenda

1. Fatima + Jastina's questions for Avril
    a. When analyzing our data, do we run test looking at moisture then soil types and THEN both?
    b. Look through how we ran our QIIME2 code
    c. What types of alpha/beta diversity tests should we run?
2. Is our project too similar to this other UJEMI paper (please say no) and if so how can we adapt ours to be different?:
    a. Annual precipitation and soil moisture level strongly associate with the bacterial community structure in Interior Douglas-fir and Sub-Boreal Spruce ecozones in British Columbia
3. Can we use the above paper as a reference
4. Ok new project idea: look at pH and A and O horizons instead
5. Feasibility / training
6. Gitignore?

Meeting Minutes

- Addressed change to research question
    - UJEMI+ paper we found was very close to our initial research question - will now be used as background
    - Pivoting to analyze microbiome of Douglas Fir in BC based on their pH and horizon type
    - Not elevation because only two elevations for different sites
    - Not soil type because all Douglas Fir have same
    - Focus on Horizon more than pH
        - Because there is a previous UJEMI paper which discusses pH (as well as other variables for all trees in the dataset) must explicitly state the novelty
        - Could simply control for pH
        - Because not all Doulgas Fir have pH (some are listed as 0) we could chose to subset and compare pH for the Douglas Fir which DO have pH values
- Shared proposal with Avril
- Final project requires all graphs and plots to be done in R
    - Use QIIME2 for initial visualization

- No longer need to include Feasibility and Training section in proposal
- Aim to cover all rubric
    - Intro -
        - Talk about papers that take about the data
        - Overview of what's been done (prior UJEMI papers, original authors, background on metadata talked about)
        - Give background regarding pH and horizon
        - Do not draw any conclusions in intro
        - General info
        - What is the learning gap? what is the purpose of our paper?
        - Remain focused on project topic, must have clear connections
    - Experimental Aims -
        - For each tool you use, why is that particular tool important?
            - Eg: why bother with functional differences when you have taxonomic info?
            - Must address importance of what you are doing, what is the purpose
    - Gantt chart -
        - Update the example Gantt charts to be relevant to our project
        - Add in manuscript draft, final copy and presentation prep timelines as well
- Spell check!
- .gitignore = file that will remain on your computer which cannot be pushed or pulled
- Hidden files like .DS_score can be ignored, do not have to remove them from repository
- Primary filtering in QIIME but most work should be done in R (easier)
- Rarefaction should only be done on alpha and beta diversity
- Can cause issues if raw data is NOT used for all other analyses

- Primary focus of project will be regarding Horizon types
    - Determine if we want to also control for pH
- Group meetings to finish up proposal
    - Provide Avril with updates up until early Saturday morning to get feedback prior to submission

Action Items

- Finish proposal!

Date: 18-Oct-2023 Signature/Initial: AC END

START

JAAFV4723-015: Initial data processing with QIIME 2 Continued

Purpose:
Continue the initial data processing (denoising and ASV determination) from JAAFV4723-009.

Materials:
Laptop with Mac iOS

Methods:

1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
2. Start a new detached screen session (class) using the command: screen -S class
    a. detach by using [Ctrl + a] followed by [d]
    b. resume screen session using the command: screen -r class
    c. terminate session by typing exit when in the session
3. Denoise using the selected truncation length and determine ASVs with DADA2 (output = rep-seqs.qza, table.qza, stats.qza)
    a. p-trim-left = 0
    b. p-trunc-len = 410
7. Generate a visualization for the DADA2 stats (output = stat.qzv).
8. Allow DADA2 to run overnight (approx. 12 hours total).

Results and Observations:

No results generated. This processing step took longer than expected.

Discussion:

All data processing steps were carried out according to the protocol described in Methods.

Conclusion:

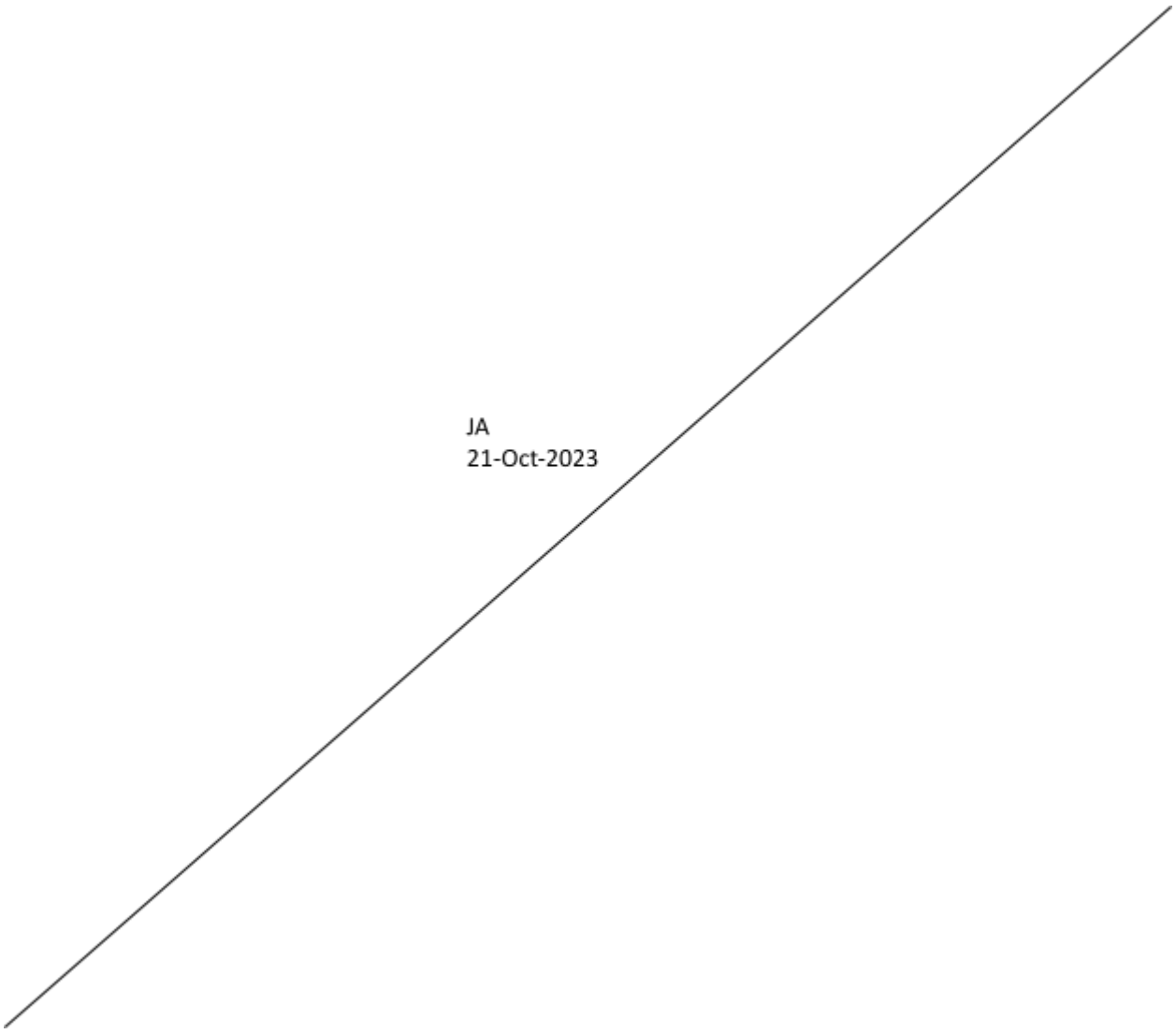
This experiment completed the initial data processing steps of the project (steps 1-2, 2-2, 3-2). The following files were generated and can be used for downstream analysis in QIIME 2 or R:

- table.qza

- rep-seqs.qza
- stats.qza

Future Directions:

The project will continue on at steps 1-3, 2-3, 3-3.



Date: 21-Oct-2023 Signature/Initial: JA END

START

JAAFV4723-017: Initial data processing with QIIME 2 Continued

Purpose:
To continue the initial data processing by retraining the classifier for taxonomic analysis.

Materials:
Laptop with Mac iOS

Methods:

1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
2. Resume detached screen session using the command: screen -r class
3. Generate a visualization for the ASVs stats (output = table.qzv, rep-seqs.qzv) and scp the files to the laptop.
4. Terminated detached screen session (class)
5. Start a new detached screen session (classifier) using the command: screen -S classifier
    a. detach by using [Ctrl + a] followed by [d]
    b. resume screen session using the command: screen -r class
    c. terminate session by typing exit when in the session
6. Attempted to re-train classifier with silva-138-99-nb-classifier.qza following the code from Module 6.
    a. Did not work; was met with error:

```
                    There was a problem with the command:
(1/1) Invalid value for '--i-sequences': Expected an artifact of at least
 type FeatureData[Sequence]. An artifact of type TaxonomicClassifier was
 provided.
(qiime2-2023.7) root@148702ecf078:/data/project_2#
```

7. Attempt to re-train classifier using the original raw classifier files:
    a. Downloaded the original raw classifier files from: https://docs.qiime2.org/2020.6/data-resources/
        i. Silva 138 SSURef NR99 full-length sequences: https://data.qiime2.org/2020.6/common/silva-138-99-seqs.qza
        ii. Silva 138 SSURef NR99 full-length: taxonomy<https://data.qiime2.org/2020.6/common/silva-138-99-tax.qza
    b. Follow the classifier training tutorial: https://docs.qiime2.org/2020.6/tutorials/feature-classifier/#train-the-classifier

8. Classify representative sequences using re-trained classifier (silva-138-99-515-806-nb-classifier.qza) by aligning the sequences to reference 16S rRNA V1 and V3 regions (output = ref-seqs.qza).
    a. The forward primer used: AGAGTTTGATCMTGGCTCAG
    b. The reverse primer used: GWATTACCGCGGCKGCTG
9. Train classifier with new ref-seq file (output = classifier.qza).
10. Use the trained classifier to assign taxonomy to reads (output = taxonomy.qza)
11. Allow this to run overnight (approx. 13 hours).

Results and Observations:

No results generated. This processing step took longer than expected.

Discussion:

All data processing steps were carried out according to the protocol described in Methods.

Conclusion:

The following files were generated and can be used for downstream analysis in QIIME 2 or R:

- table.qzv
- rep-seqs.qzv
- taxonomy.qza

Future Directions:

The project will continue on at steps 1-4.

JA
22-Oct-2023

Date: 22-Oct-2023 Signature/Initial: JA END

START

JAAFV4723-019: Initial data processing with QIIME 2 Continued

Purpose:
To continue the initial data processing by viewing the taxonomic analysis and taxa barplots to ensure classifier retraining was run properly, filter the features table for no mitochondria and chloroplast sequences, and only BC Douglas Fir stands.

Materials:
Laptop with Mac iOS

Methods:

1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
2. Resume detached screen session using the command: screen -r classifier
3. Visualize the classifications assigned and generate taxa bar plots (output = taxonomy.qzv, taxa-bar-plots.qzv).
4. View the taxa-bar-plots.qzv file using the QIIME 2 View website (https://view.qiime2.org/) to ensure the classifier step ran correctly.
5. Filter out mitochondria and chloroplast sequences and create visualization file (output = table-no-mitochondria-no-chloroplast.qza, table-no-mitochondria-no-chloroplast.qzv)
9. Create a subset of the filtered features table by:
    a. filtering to include only BC Douglas Fir stands (output = tree-cover-filtered-table.qza) (for Question 1, 2, and 3)
10. Generate visualizations for the filtered features table.
11. Generate trees in preparation for phylogenetic diversity analyses.
    a. Output = aligned-rep-seqs.qza, masked-aligned-rep-seqs.qza, unrooted-tree.qza, rooted-tree.qza

Discussion:

All data processing steps were carried out according to the protocol described in Methods.

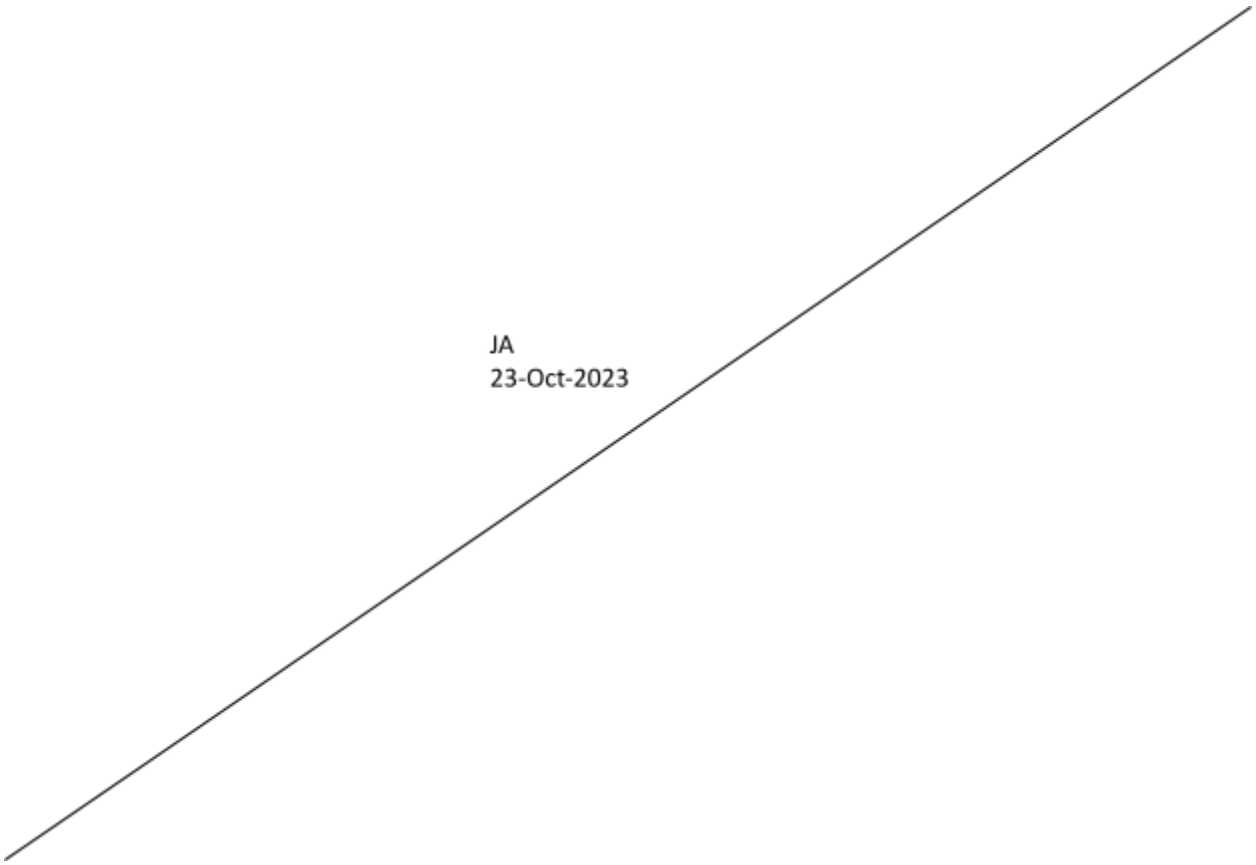
Conclusion:

This experiment completed the initial data processing steps of the project (steps 1-3, 2-3, 3-3). The following files were generated and can be used for downstream analysis in QIIME 2 or R:

- taxonomy.qzv
- taxa-bar-plots.qzv
- table-no-mitochondria-no-chloroplast.qza
- table-no-mitochondria-no-chloroplast.qzv
- tree-cover-filtered-table.qza
- aligned-rep-seqs.qza
- masked-aligned-rep-seqs.qza
- unrooted-tree.qza
- rooted-tree.qza

Future Directions:

The project will continue on at steps 1-4.



Date: 23-Oct-2023 Signature/Initial: JA END

START

JAAFV4723-021: Diversity metrics generation in QIIME2

Purpose:
To generate the alpha-rarefaction curve and determine the sampling depth.

Materials:
Laptop with Mac iOS

Methods:

1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
2. Open the tree-cover-filtered-table.qzv file generated in JAAFV4723-016 in QIIME 2 View (https://view.qiime2.org/).
3. Using the "Interactive Sample Detail" tab, determine an appropriate sampling depth for the filtered tree cover table after visualizing the data according to the appropriate metadata category (see Figure 1 for details):
    a. Metadata category - Horizon; selected sampling depth - 4092
4. Generate alpha-rarefaction curves (Figure 2) using the table's max sampling depth (19000).
5. Review each alpha rarefaction curve to determine whether the selected sampling depth in step 4 is appropriate. If the selected sampling depths are appropriate, proceed to step 7, otherwise, re-select sampling depths.

Results and Observations:

Figure 1. Interactive Sample Detail of samples against the 'Horizon' Metadata Category. The Sampling depth 4092 was selected to retain 401,016 (71.51%) features in 98 (96.08%) samples.

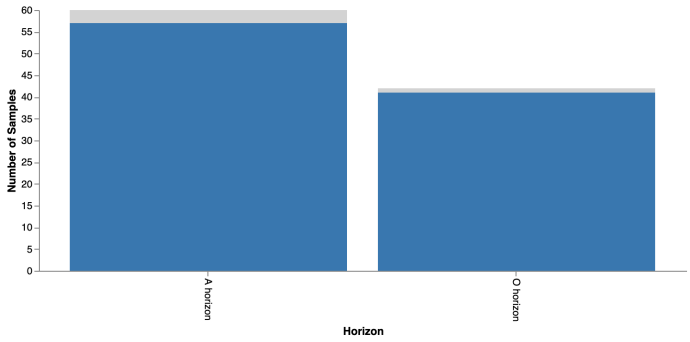
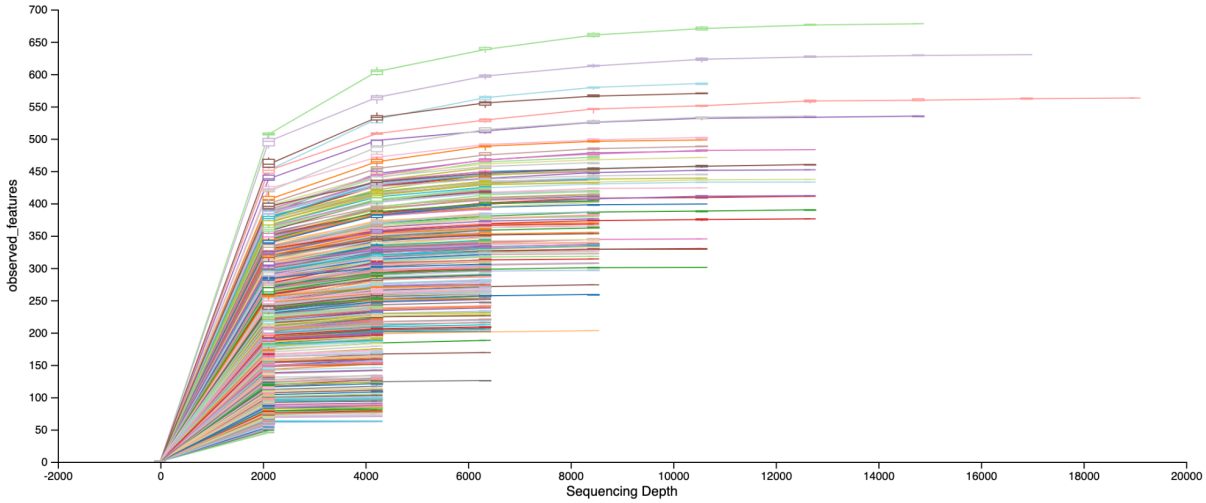


Figure 2. Alpha rarefaction curve.



Discussion:

After visualizing the horizon metadata category for the tree-cover-filtered-table.qzv file, we selected a sampling depth of 4092. This decision was based on a depth that retained the highest number of features. This value was then analyzed against the rarefaction curve that showed the sampling depth is within the plateau where the ASVs were saturated.

Conclusion:

This experiment completed the initial data processing steps of the project (step 1-4).

Future Directions:

The project will continue on at steps 1-5, 2-4, and 3-4.

Filtered feature tables and phylogenetic trees will be used for further analyses in QIIME 2 (alpha and beta diversity metrics) or R (differential abundance analysis).

JA
23-Oct-2023

Date: 23-Oct-2023 Signature/Initial: JA END

START

W8-TM: Week 8 Team Meeting

Agenda

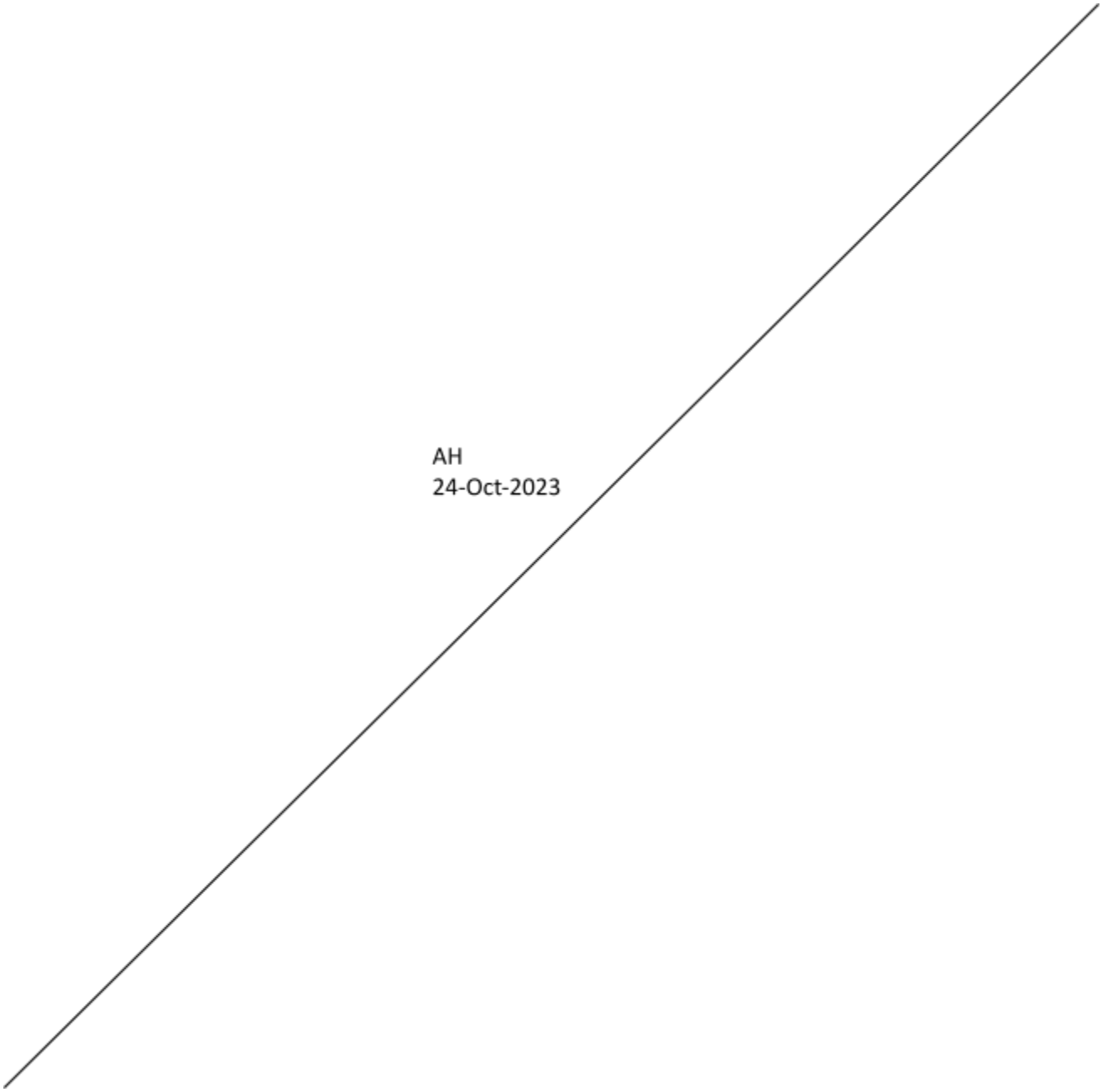
1. Go over the project proposal
2. What are next steps? How should we structure the next part of this project?

Meeting Minutes

- Discussed next steps
- Dividing who wants to do what going forward
    - Start doing R
    - If anyone loves QIIME2, can start looking at PICRUSt
- Set up github with R (or use github desktop app)
    - Make sure to pull and push files when working on your desktop
- One person on each separate thing (alpha / beta diversity, indicator, PICRUSt)
    - For indicator, think about how to visualize data
    - People who finish first can help others, start compiling manuscript
    - Only difference between DESeq2 and PICRUSt is literally just running PICRUSt
    - 2 people work on this together, based on stuff in QIIME2
- One person on alpha / beta, 2 on indicator, 2 on PICRUStR 1 setting up indicator (making it match up with dataset), 1 trouble shooting plots / making plot
    - This is the option we'll do
- Check out ggplot2 thing Avril sent for plot formatting
    - e.g. rotating bacteria names
    - supplementary plots too
    - bigger bar = more significant (log 2 fold change bar plots)
    - for the bar plot: figure out how to re-order so it goes smallest to biggest
    - for PICRUSt:
        - QIIME2 --> 3 files (metapsych pathways, called "pathway")
    - Pathway = group of reads
- **ROLES:**
    - Jastina: indicator plot visualization
    - Ally: PICRUSt
    - Fatima: alpha / beta diversity
    - Vera: PICRUSt
    - Anu: indicator code

Action Items

- Get started!!
- Monday evening meet as a group



Date: 24-Oct-2023 Signature/Initial: AH END

START

W9-TM: Week 9 Team Meeting

Agenda

- Notebooks: we have 2 right now, one on Docs and the readme (the Docs one is more updated!) Should we keep going with the doc or push to GitHub?
- Installing PICRUSt2 difficulties (error: The current user does not have write permissions to the target environment. environment location: /opt/conda/envs/qiime2-2023.7) 2b13b2888ca5c51f1df3aafbf6c44a8ac30c152b
- Walk thru comments for clarification:

1. Dataset Overview
    - Q: What did the distribution of lengths generally look like?
        - A: Does this mean long/short? How should we describe this?
    - Q: This is definitely desirable, but this won't affect sample filtering. Sample filtering = metadata filtering.
        - A: Do we delete this sentence?
2. Proposed Approach Table
    - Q: This is a little off – your reads will be inputted into PICRUSt2, which will then produce outputs in the form of functional datasets (KEGG as you note, and also MetaCyc pathways and enzyme data (EC), just fyi). The OTU table is the table of taxonomic counts, which isn't a consideration for PICRUSt2 – PICRUSt2 outputs and OTU tables are basically different ways of annotating the same data.
        - A: Plz help with walk thru :(
    - Q: Permutation test
        - A: Is using the p value preferred over this permutation test for significance? What should we change in this table?

- Questions about coding with R:

1. Aim 2: What species indicators do we see arise in A vs O horizons?
    - Is the workflow: assignment 5 (create phyloseq object; rarefy samples; ordinaton figure) > assignment 6 (indicator species analysis)
    - Do we use the filtered metadata table in R or the original table.qza?

Meeting Minutes

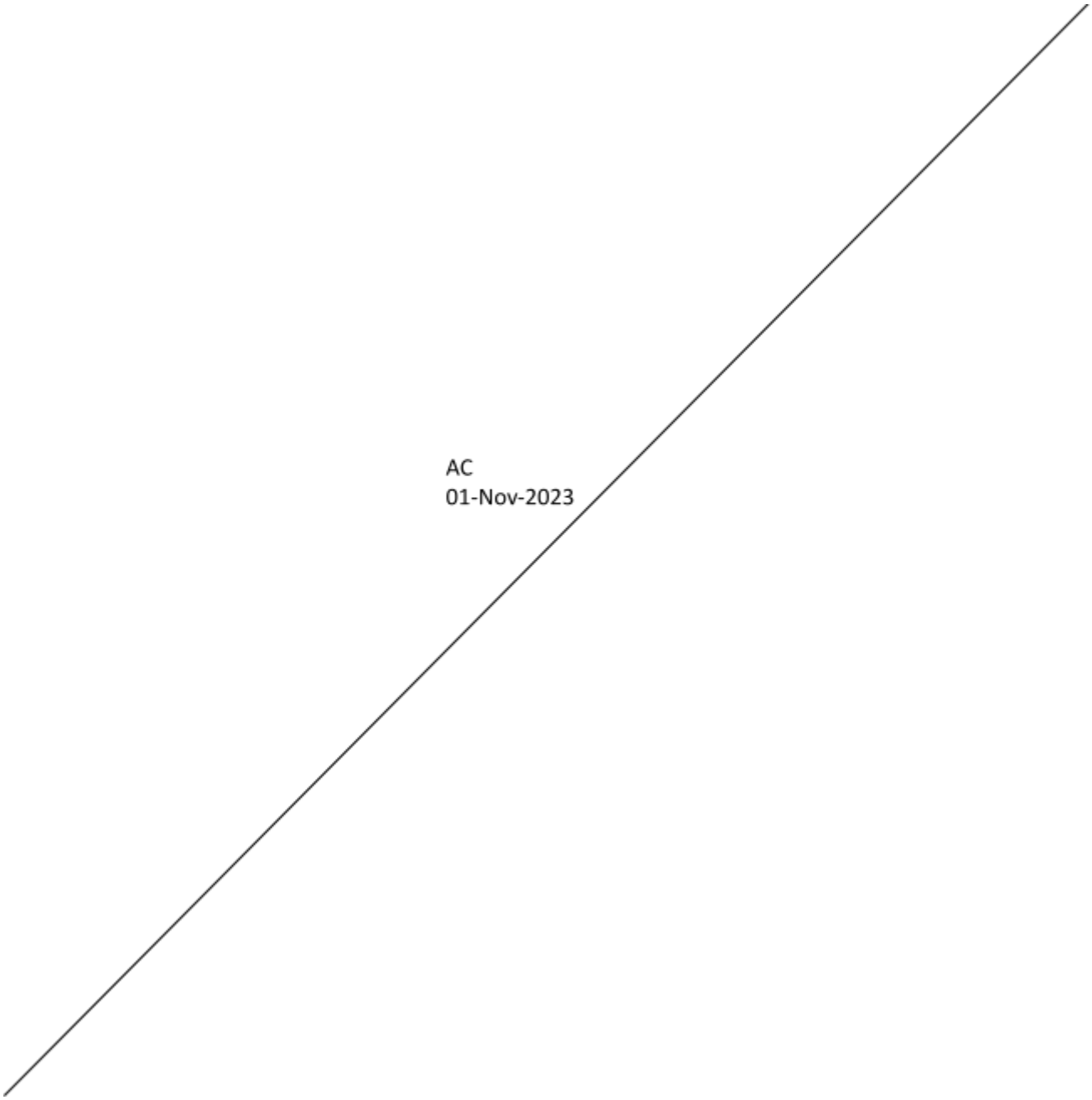
- Notebook
    - stick with Docs for notebook
    - include notes for R as well
    - discussions just for important things

        - don't spend too much time on this, just show that you're thinking through things
    - more emphasis on methods so projects are recreatable
    - make an effort to make it readable and backed up by rationale
    - Docs doesn't record history, but for purposes of this class thats ok
        - don't delete portions!
- Downloading PICRUSt2
    - should be able to just download to directory
    - Avril will check it out
- resubmitting proposal
    - Distribution of lengths (dataset overview)
    - logic is a bit circular (sounds like "shorter because its shorter")
    - Do most of the samples have maximum read length? Are they outliers?
    - We can delete sentence about chloroplast filtering
    - We get half-mark for flowchart (formatted properly in the PDF!)
- SUBMIT BOTH WORD AND PDF FILES
    - PICRUSt2 generated in QIIME2 but analyzed in R
    - Resources for PICRUSt2:
        - look up "PICRUSt2 code example"
        - stack exchange / stack overflow
    - add cell to export PICRUSt2 to R

```
qiime picrust2 full-pipeline
--i-table ld67-table-dada2.qza
--i-seq ld67-rep-seqs-dada2.qza \
```

- Permutation test
    - stat test
    - permutation tests robustness of p-value
    - repeats indicator test 999 times (calculates p-values each time)
    - in resubmission "we'll use 999 mutations and use a p-value cut off of ____"
    - indicator taxa gives table of p-value and stat value regardless of number of permutations
- Questions coding with R
    - DON'T need to rarefy for indicator species (that is only for diversity)
    - everything autoconverted for you to relative abundance (don't need to do anything extra) --> look at example code
    - use filtered metadata tables
- Files too large (from github desktop)

- can ignore full metadata tables (only need filtered ones)



Date: 01-Nov-2023 Signature/Initial: AC END

START

JAAFV4723-028: Indicator Species Analysis R Processing

Purpose:
To begin R processing of Aim 2.

Materials:
Surface laptop with Windows 11

Methods:

1. Installed the indicspecies package.
2. Load in the Load in the 'tidyverse', 'dplyr', 'phyloseq', 'ape', 'vegan', 'indicspecies' and 'microbiome' package
3. Load in the exported files: soil metadata, OTU table, taxonomy file, and phylogenetic tree
4. Adjust exported files to be read into phyloseq objects
    a. Format OTU table as a matrix with row names and column name as OTUs and sampleIDs
    b. Format metadata into phyloseq sample data
    c. Format taxonomy by separating into taxa rank columns
5. Create phyloseq object adjusted files

Conclusion:

This experiment generated the phyloseq object for further use in the project.

Future Directions:

Perform indicator species analysis

AC
31-Oct-2023

W9-TM: Week 9 Team Meeting

Agenda

- Notebooks: we have 2 right now, one on Docs and the readme (the Docs one is more updated!) Should we keep going with the doc or push to GitHub?
- Installing PICRUSt2 difficulties (error: The current user does not have write permissions to the target environment. environment location: /opt/conda/envs/qiime2-2023.7) 2b13b2888ca5c51f1df3aafbf6c44a8ac30c152b
- Walk thru comments for clarification:

2. Dataset Overview
    - Q: What did the distribution of lengths generally look like?
        - A: Does this mean long/short? How should we describe this?
    - Q: This is definitely desirable, but this won't affect sample filtering. Sample filtering = metadata filtering.
        - A: Do we delete this sentence?
3. Proposed Approach Table
    - Q: This is a little off – your reads will be inputted into PICRUSt2, which will then produce outputs in the form of functional datasets (KEGG as you note, and also MetaCyc pathways and enzyme data (EC), just fyi). The OTU table is the table of taxonomic counts, which isn't a consideration for PICRUSt2 – PICRUSt2 outputs and OTU tables are basically different ways of annotating the same data.
        - A: Plz help with walk thru :(
    - Q: Permutation test
        - A: Is using the p value preferred over this permutation test for significance? What should we change in this table?

- Questions about coding with R:

2. Aim 2: What species indicators do we see arise in A vs O horizons?
    - Is the workflow: assignment 5 (create phyloseq object; rarefy samples; ordinaton figure) > assignment 6 (indicator species analysis)
    - Do we use the filtered metadata table in R or the original table.qza?

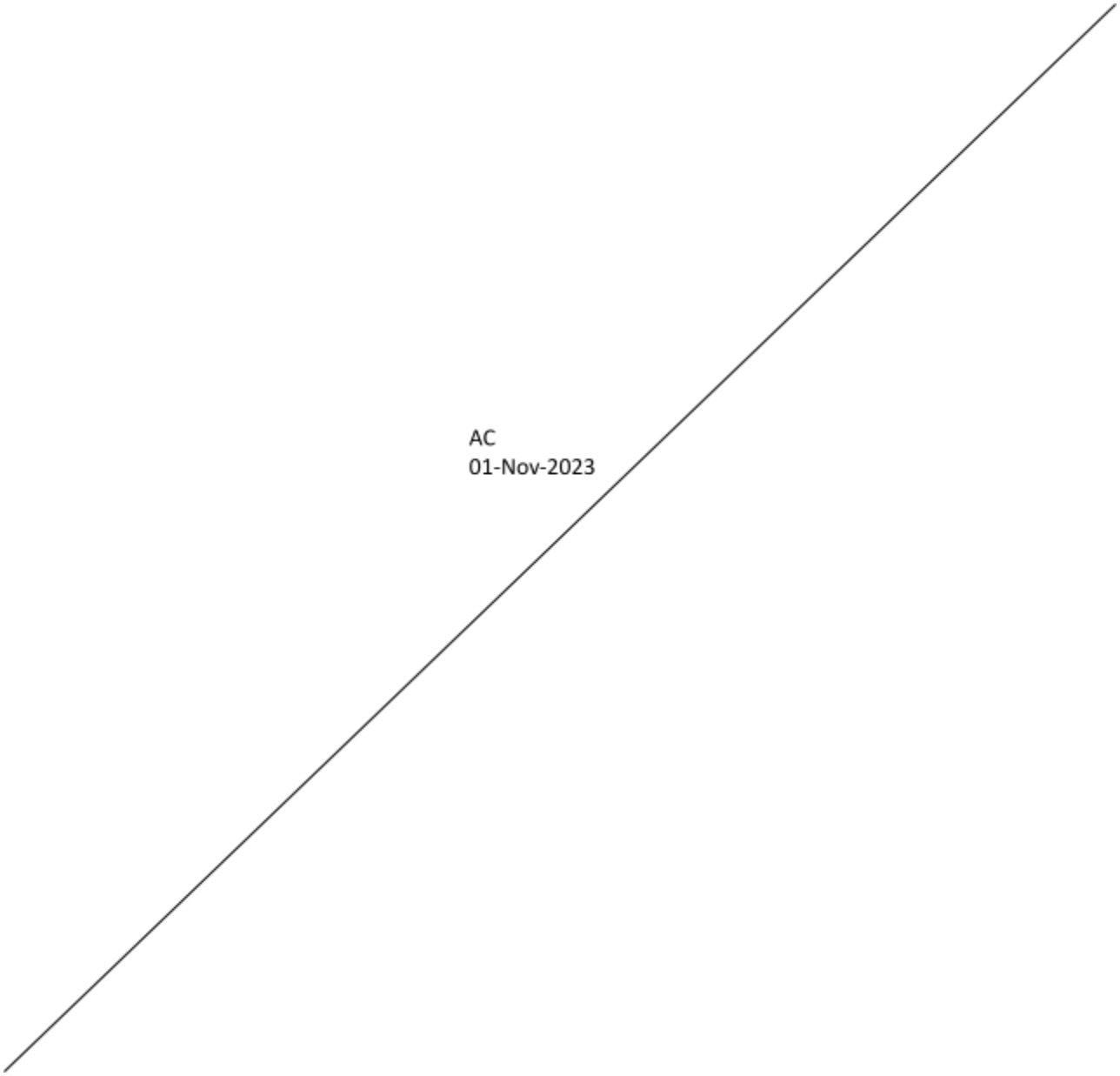
Meeting Minutes

- Notebook
    - stick with Docs for notebook
    - include notes for R as well
    - discussions just for important things
        - don't spend too much time on this, just show that you're thinking through

things
    - more emphasis on methods so projects are recreatable
    - make an effort to make it readable and backed up by rationale
    - Docs doesn't record history, but for purposes of this class thats ok
        - don't delete portions!
- Downloading PICRUSt2
    - should be able to just download to directory
    - Avril will check it out
- resubmitting proposal
    - Distribution of lengths (dataset overview)
    - logic is a bit circular (sounds like "shorter because its shorter")
    - Do most of the samples have maximum read length? Are they outliers?
    - We can delete sentence about chloroplast filtering
    - We get half-mark for flowchart (formatted properly in the PDF!)
- SUBMIT BOTH WORD AND PDF FILES
    - PICRUSt2 generated in QIIME2 but analyzed in R
    - Resources for PICRUSt2:
        - look up "PICRUSt2 code example"
        - stack exchange / stack overflow
    - add cell to export PICRUSt2 to R

qiime picrust2 full-pipeline
--i-table ld67-table-dada2.qza
--i-seq ld67-rep-seqs-dada2.qza \

- Permutation test
    - stat test
    - permutation tests robustness of p-value
    - repeats indicator test 999 times (calculates p-values each time)
    - in resubmission "we'll use 999 mutations and use a p-value cut off of ____"
    - indicator taxa gives table of p-value and stat value regardless of number of permutations
- Questions coding with R
    - DON'T need to rarefy for indicator species (that is only for diversity)
    - everything autoconverted for you to relative abundance (don't need to do anything extra) --> look at example code
    - use filtered metadata tables
- Files too large (from github desktop)
    - can ignore full metadata tables (only need filtered ones)



Date: 01-Nov-2023 Signature/Initial: AC END

START

W10-TM: Week 10 Team Meeting

Agenda

- still don't know how to install PICRUSt2 :((
- getting following error each time we try to load any of the files required for phyloseq: Error in file(file, "rt") : cannot open the connection In addition: Warning message: In file(file, "rt") : cannot open file '../mpt_export/filtered_table_export/filtered_feature-table.txt': No such file or directory
    - tried creating new R project and changing working directory to R processing folder (Folder path currently: micb 475 -> code -> R processing -> mpt_export (with exported filtered table, rooted tree, table, and taxonomy folders)

Meeting Minutes

- to install picrust:

wget https://github.com/picrust/picrust2/archive/v2.5.2.tar.gz
tar xvzf v2.5.2.tar.gz
cd picrust2-2.5.2/

conda env create -f picrust2-env.yaml
conda activate picrust2
pip install --editable .

- to run picrust: picrust2_pipeline.py [-i, -o]
- result from indicator taxa to be used in picrust
- last team meeting: anticipate half of the time investing on practice presentation (10 mins)

Action Items

- try finishing all the tests by next meeting!!
- same time for next week's meeting (Wednesday 1pm)

VP
08-Nov-2023

Date: 08-Nov-2023 Signature/Initial: VP END
START

JAAFV4723-033: Indicator Species Analysis R Processing Continued

Purpose:
To complete the R processing of Aim 2.

Materials:
Surface laptop with Windows 11

Methods:
1. Edit phyloseq code
2. Group the phyloseq object based on genus level
3. Flip OTU row and column names
4. Calculate indicator values for ASV's 999 times as per permutation hypothesis test
5. Perform indicator species analysis
6. Produce summary table of ISA results

Conclusion:

This experiment completed the initial ISA steps of the project.

Future Directions:

Generate visualization of indicator species analysis in R

AC
09-Nov-2023

Date: 09-Nov-2023 Signature/Initial: AC Go to page 52
START

JAAFV4723-034: PICRUSt2 Processing with QIIME 2

Purpose:
To process data with the PICRUSt2 package on QIIME2 for further visualization on R.

Materials:
Laptop with Mac iOS

Methods:

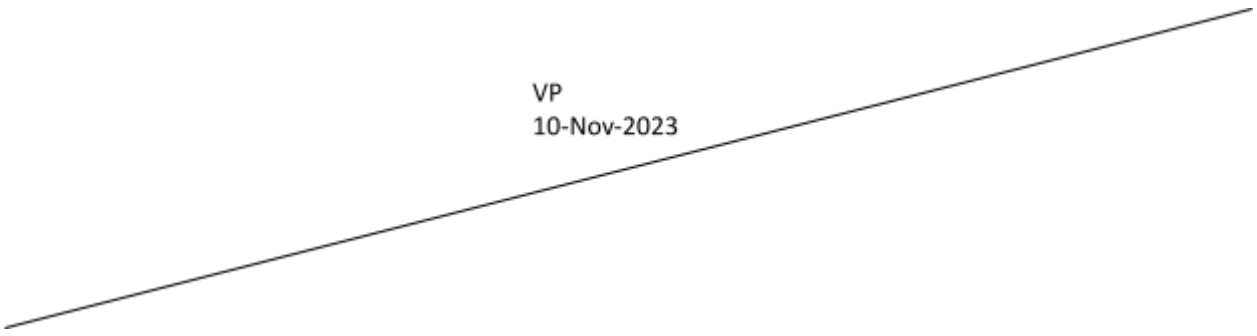
1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
2. Activate PICRUSt2 environment on QIIME2.
3. Download 'rep-seqs.qzv' from GitHub, insert to QIIME2 View, and download the raw FASTA file and transfer it to the remote server.
4. Prepare 'feature-table.biom', 'sequences.FASTA' for PICRUSt2.
5. Generate PICRUSt2 output directory with *picrust2_pipeline.py* command.

Conclusion:

This experiment successfully produced PICRUSt2 output for AIM 3.

Future Directions:

The project will continue to process on R to generate error bar plots, heat maps and PCA plots.



Date: 10-Nov-2023 Signature/Initial: VP END

START

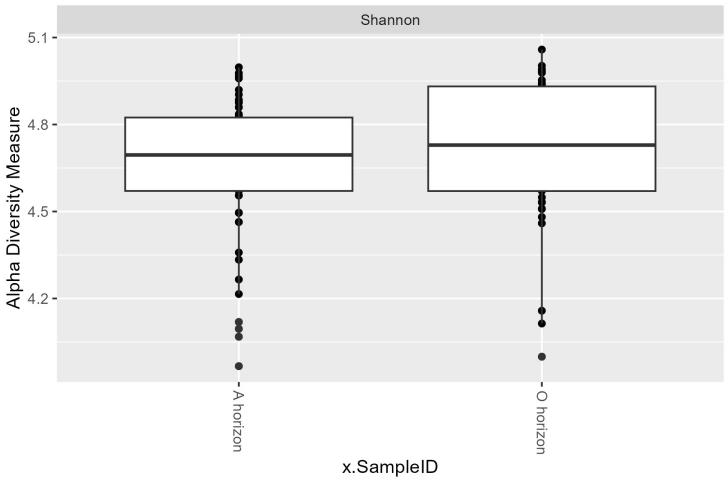
JAAFV4723-035: Diversity Metrics Analysis Through R

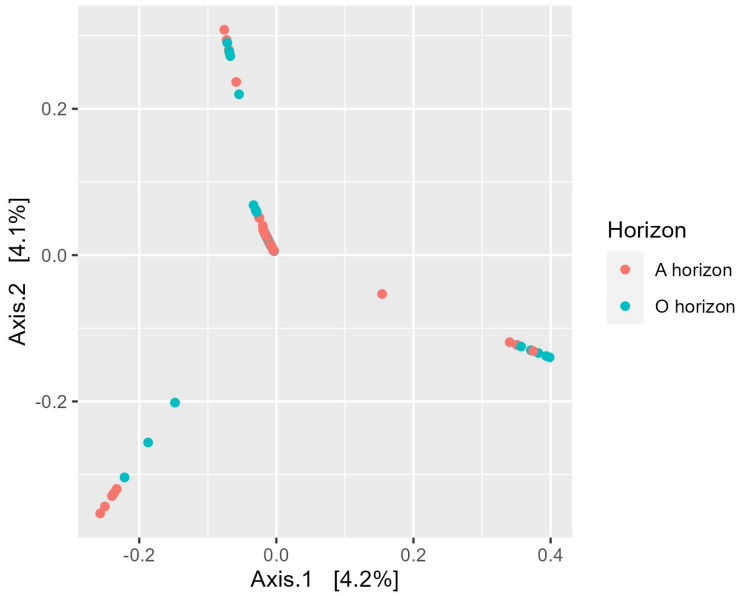
Purpose:
To determine soil richness and beta diversity and create interpretable plots.

Materials:
Dell laptop with Windows 11

Methods:
1. Install indicspecies package.
2. Load in tidyverse, dplyr, phyloseq, ape, vegan, indicspecies, ggpubr packages.
3. Get data ready by creating a phyloseq object using metadata, otu table, taxonomy file and the phylogenetic tree.
4. Rarefy data to 4092 - depth determined using qiime2 data processing.
5. Save rarefied and non-rarefied files.
6. Use the plot_richness() function on all data to view all alpha diversity metrics.
7. Use plot_richness() specifically with the Shannon measure.
8. Use the plot_richness() function under the measure Shannon against the variable horizon.
9. Save the new plot as soil_richness_shannon.
10. Use the distance() function with the "bray" method to determine distances between species.
11. Use this calculated value with the ordinate() function and color=Horizon to plot beta diversity with the Bray Curtis metrics
12. Save new beta diversity plot

Results and Observations:
The alpha diversity box plot seems to have overlapping horizons which therefore suggests no significance. The beta diversity plot is quite confusing and we are not able to understand what it is showing. It is difficult to ascertain any possible patterns as there appears to be no clustering.





Conclusion:

The Shannon alpha diversity plot suggests no significant differences between A and O horizon
The Bray Curtis beta diversity plot suggests no significant patterns in species diversity.

Future Directions:

We aim to run a statistical significance check using the Wilcox T to determine significance and try other beta diversity metrics.

FN
13-Nov-2023

START

W11-TM: Week 11 Team Meeting

Agenda

- AIM 3: picrust2
    - help with picrust2 visualization (how do we incorporate indicator taxa?)
    - is the metadata file just the original from the paper? or a filtered one?
- AIM 2: ISA
    - we found some example papers of what we want our data to look like (https://www.nature.com/articles/ismej201379)
    - what are on the x and y axis?
    - we want to make a bubble plot where the size of the dot represents relative abundance
        - can't access link you sent Anu

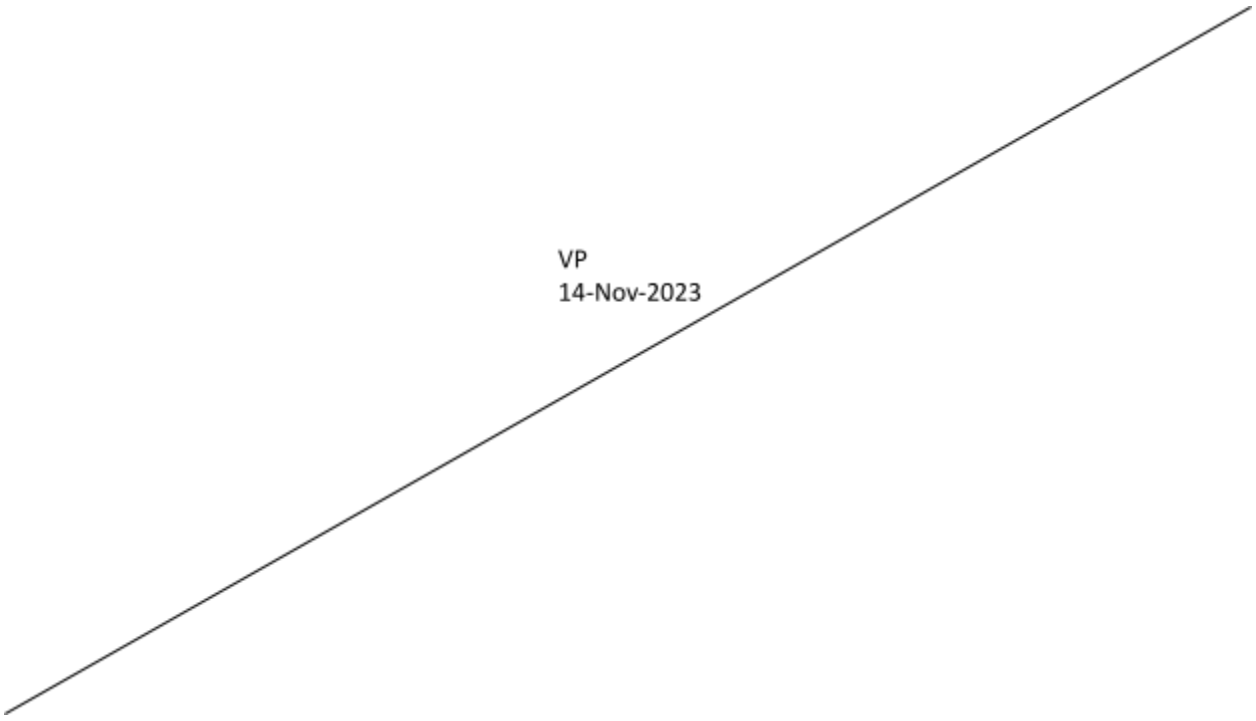
Meeting minutes

- AIM 3
    - use filtered metadata
    - heat map, all are significant results; may have to state that "everything is very different"
        - suggestion: check the core microbiome and see which specific species present prevalently
    - PCA looks great
    - alternative methods: enrichment analysis to see which pathway is overarching, run indicator species to see which group is more distinctive over another (make a phyloseq object, create a "fake" taxonomy table with one column with all pathway names)
- AIM 2
    - focus on the 3*** ones
    - in microbiome pkg - "microbiome::transform('compositional')" - transformed phyloseq objects - then do "clr" & "psmelt()"
    - create a minimum threshold to remove barely present species
        - phyloseq_filter_prevalence(physeq, prev.trh = 0.1, abund.trh = 0.01, threshold_condition = “AND”, abund.type = "total")
        - check if abundance is in percentage or add to 1
        - https://rdrr.io/github/vmikk/metagMisc/man/phyloseq_filter_prevalence.html
    - stick with one visualization - use ggplot

        - color indicating what group of species is in the indicator for; determine association b/w condition and species
    - from Avril: average>1e-4; function(x) mean(x) > 1e-4; https://github.com/armetcal/Metcalfe-Roach_PD_2023/blob/main/R%20Scripts/999.%20Differential%20Abundance%20Functions.R
- AIM 1
    - alpha plot looks okay
        - wilcox test: remove quotes; sample ID should be replaced with alpha diversity
    - beta plot needs to fix
        - send Avril R code
- REMINDERS
    - 2 meetings left: 1.5 full meeting, 0.5 presentation prep
    - finish presentation by last week's meeting (11/29)
    - after that, start drafting manuscript and practice presentation
        - send manuscript to Avril early for good feedback

Action item

- complie all results in one google slides (story board for manuscript - what you want to show and talk about)



Date: 14-Nov-2023 Signature/Initial: VP END

START

JAAFV4723-039: PICRUSt2 Visualization in R

Purpose:

To generate error bar plots, heat maps and PCA plots for visualization of PICRUSt2 data. These plots will be used to analyze PICRUSt2 data in the writing of our final manuscript.

Materials:

Laptop with Mac iOS

Methods:

1. Connect to UBC's VPN. Open Windows Terminal and login to MICB475 class server through secure shell ssh using the provided login credentials.
2. Secure copy PICRUSt2 outputs to the local environment.
3. Open a new project in RStudio.
4. Load in necessary data.
5. Compute and annotate KEGG orthology pathway and enzyme commission pathway.
6. Generate plots.

Discussion:

There was some difficulty in annotating the MetaCyc data.

Conclusion:

May need to double check that metadata and abundance files match up.

Future Directions:

Finish annotating MetaCyc data and generating plots.

AH
14-Nov-2023

Date: 14-Nov-2023 Signature/Initial: AH END

START

JAAFV4723-040: Diversity Metrics Analysis Through R Continued

Purpose:
To determine significance for soil richness and beta diversity plots.

Materials:
Dell laptop with Windows 11

Methods:
1. Load in tidyverse, dplyr, phyloseq, ape, vegan, indicspecies, ggpubr packages.
2. Create a variable to filter out the soil richness data with the Shannon variable.
3. Use the wilcox.test() function with the Horizon value and variable created above as data
4. Use the adonis2() function to determine significance using the PERMANOVA statistical test with Weighted and Unweighted Unifrac measures of beta diversity

Results and Observations:
The Wilcoxan rank sum test with continuity correction gives a p-value = 0.1999. Both beta diversity plots return a p-value <0.001.

Conclusion:
The Shannon alpha diversity plot has no significant differences between the A and O horizons as the returned p-value is greater than 0.05. Beta diversity plots with weighted and unweighted Unifrac measures are significant.

Future Directions:
We will make new beta diversity plots using the weighted and unweighted Unifrac metrics and make soil richness plots with other diversity metrics to see if others are possibly significant.

FN
15-Nov-2023

START

JAAFV4723-041: PICRUSt2 MetaCyc Visualization in R

Purpose:

To generate error bar plots, heat maps and PCA plots for visualization of MetaCyc PICRUSt2 data. These plots will be used to analyze PICRUSt2 data in the writing of our final manuscript.

Materials:

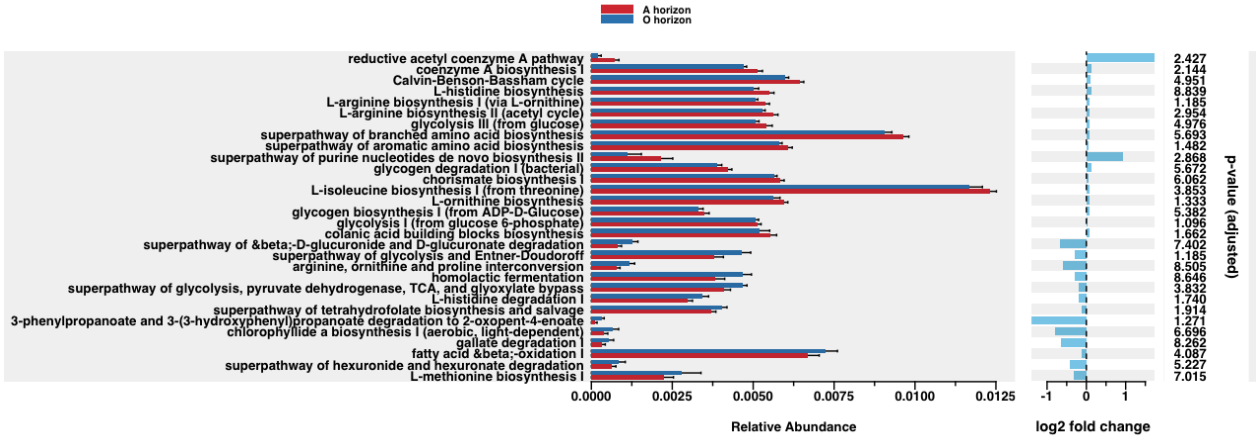
Laptop with Mac iOS

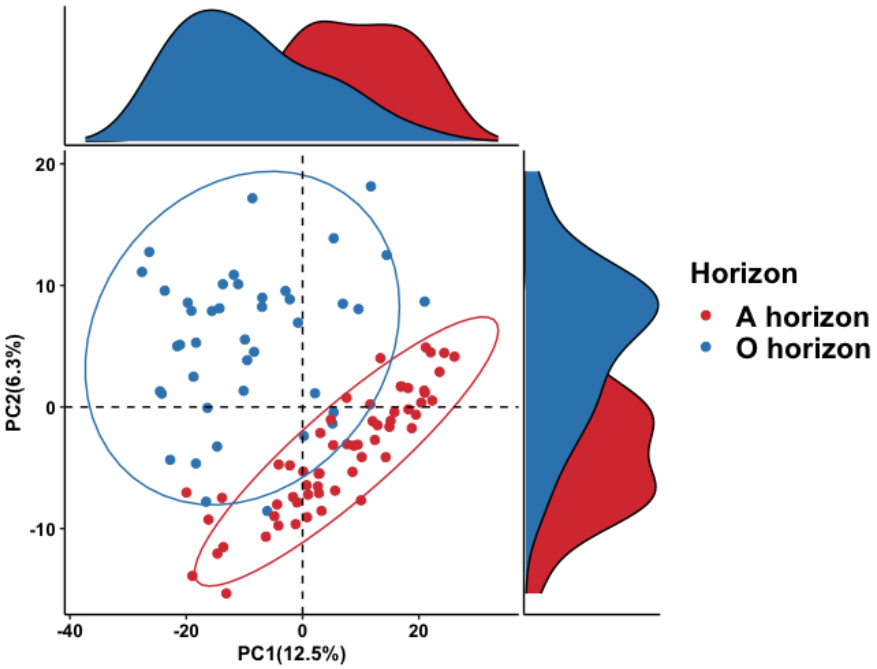
Methods:

1. Redownloaded path_abun_unstrat.tsv from PICRUSt2 outputs (last time the file was corrupted which is why MetaCyc wasn't working)
2. Load in necessary data to RStudio
3. Compute and annotate MetaCyc pathway.
4. Filter for significant results and subset because there are still too many significant results
5. Generate plots (error bar, PCA and heatmap)
6. Upload code to GitHub (code/R Processing/picrust_analysis.R)

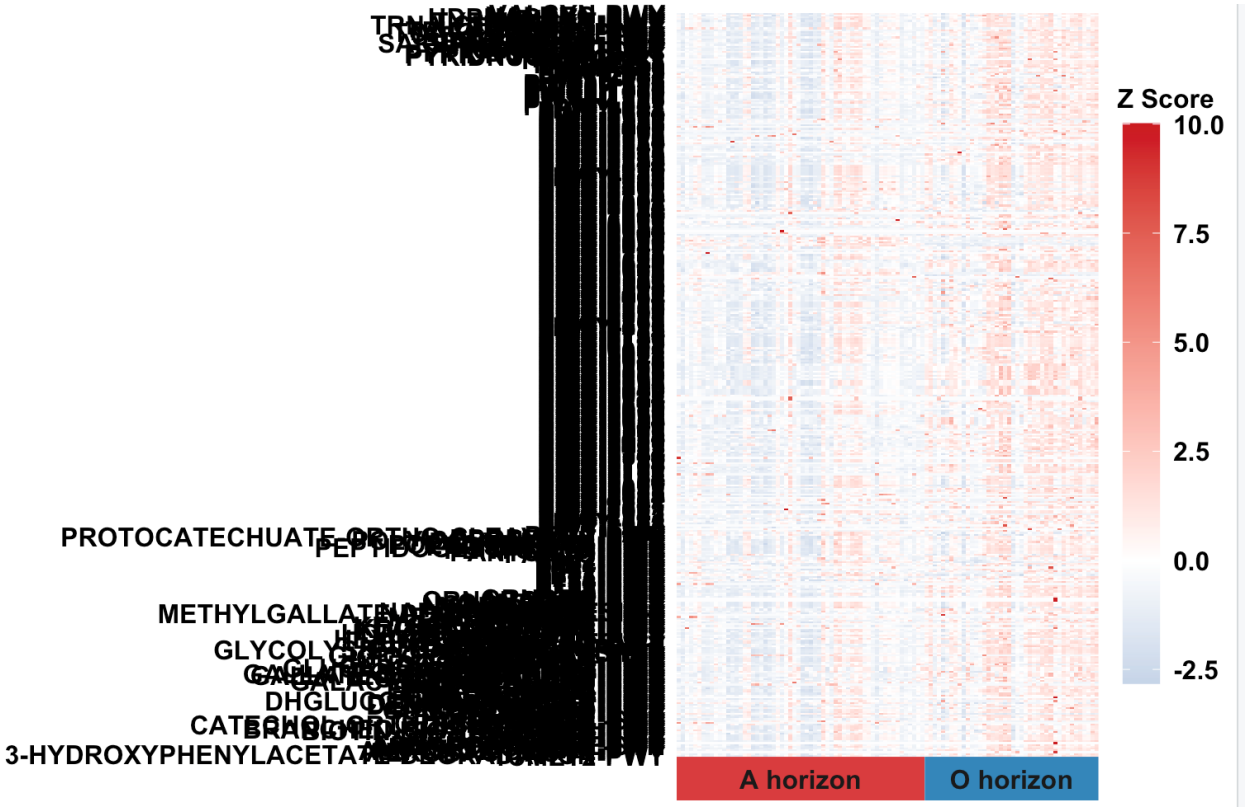
Discussion:

The error bar plot and PCA plots look good, although error bar plot is only a random subset of the MetaCyc data.



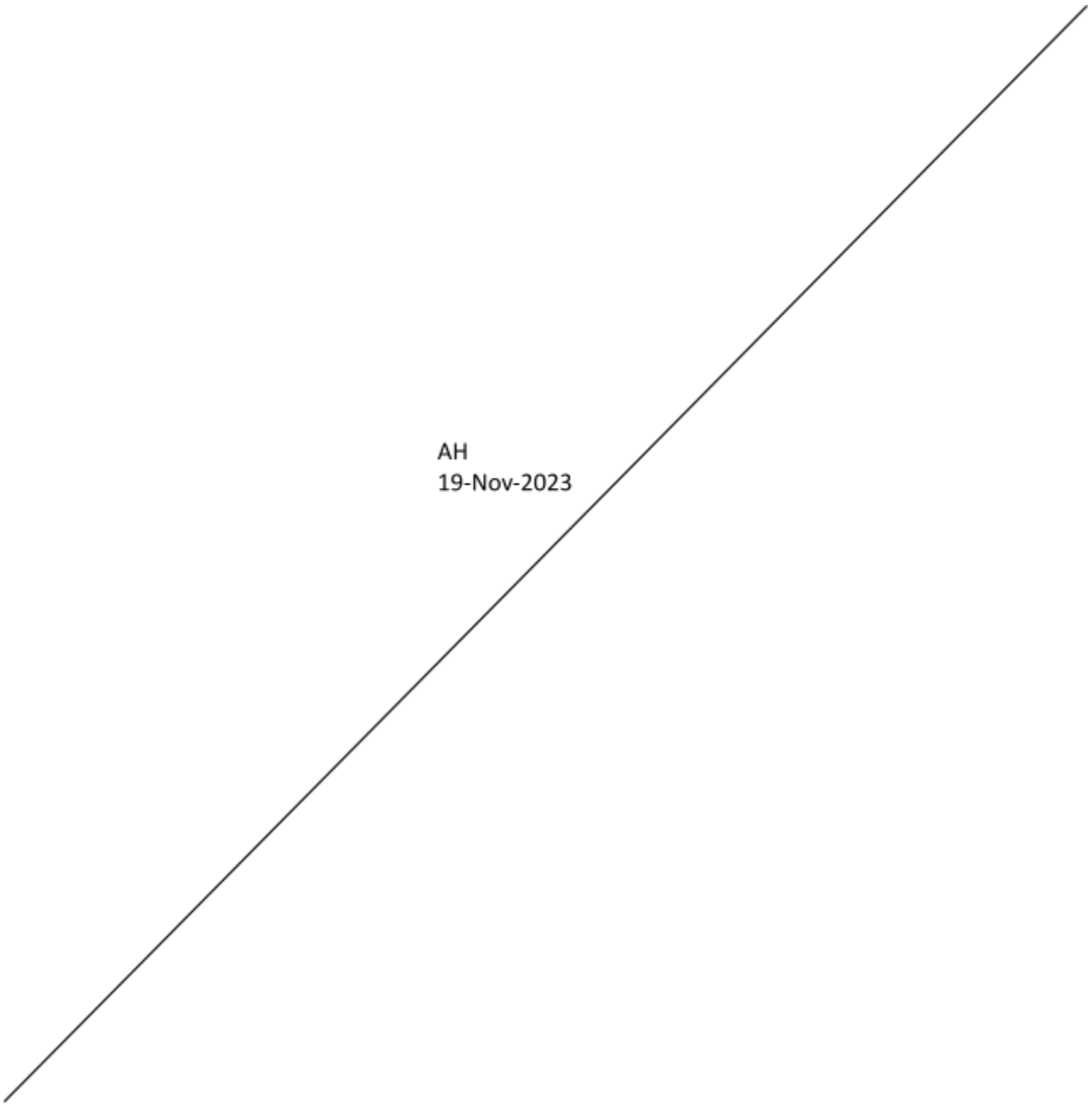


Everything is significantly different so we need to find a way to decide what to talk about in the manuscript. The heat map is hard to interpret because so many pathways are included.



Conclusion:
We have usable plots for our manuscript! The error bar plot could be filtered rather than using a random subset to be able to talk more concretely about the data.

Future Directions:
Filter error bar plot by log2fold change and / or prevalence. Decide whether the heat map is necessary for our manuscript.



Date: 19-Nov-2023 Signature/Initial: AH END

START

JAAFV4723-044: Diversity Metrics Analysis Through R Continued

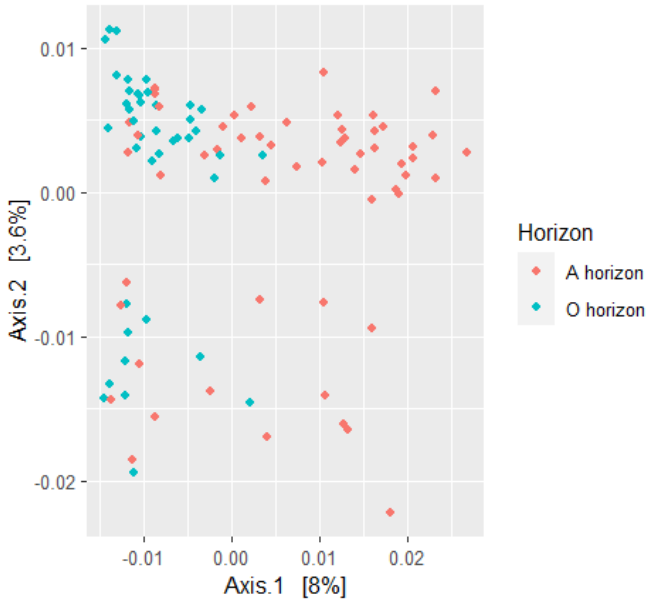
Purpose:
To create beta diversity plots with Weighted and Unweighted Unifrac distance.

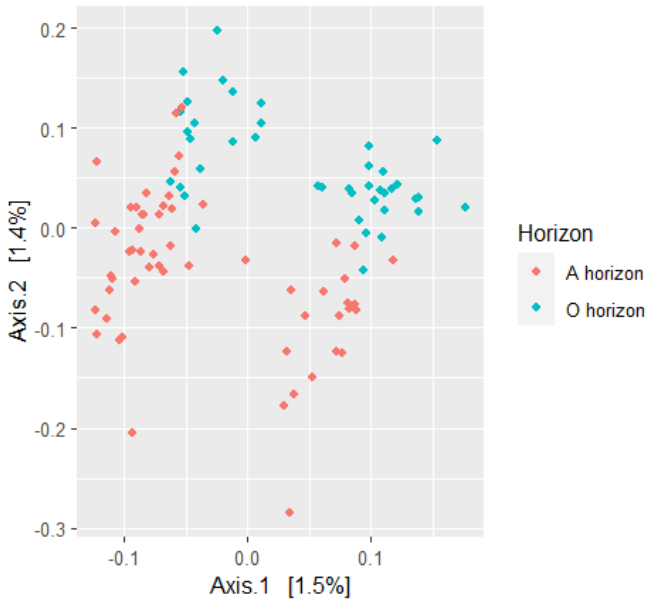
Materials:
Dell laptop with Windows 11

Methods:

1. Load in tidyverse, dplyr, phyloseq, ape, vegan, indicspecies, and ggpubr packages.
2. Create a weighted variable and using the Unifrac() function and weighted=TRUE determine the distance.
3. Create an unweighted variable and use the Unifrac() function and weighted=FALSE to determine the distance.
4. Use both distances determined with the ordinate() function and color=Horizon to create plots.
5. Save both plots.

Results and Observations:





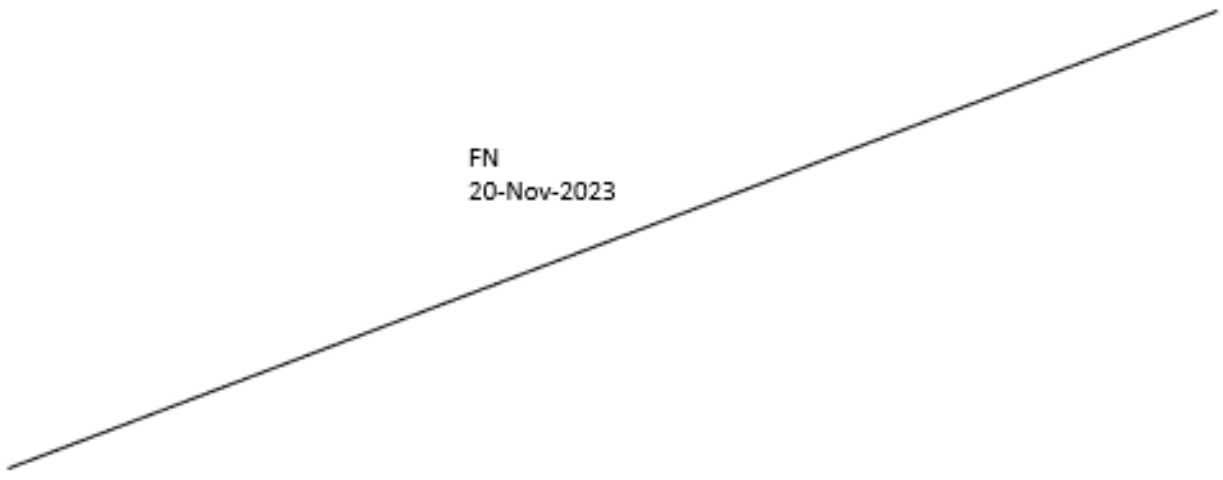
Both beta diversity plots, weighted (left) and unweighted (right) have clear clustering patterns for their perspective horizons.

Conclusion:

Beta diversity plots with weighted and unweighted Unifrac measures show that species diversity differs according to each horizon.

Future Directions:

Alter beta plots by adding ellipses and corresponding theme colours.



START

JAAFV4723-046: Diversity Metrics Analysis Through R Continued

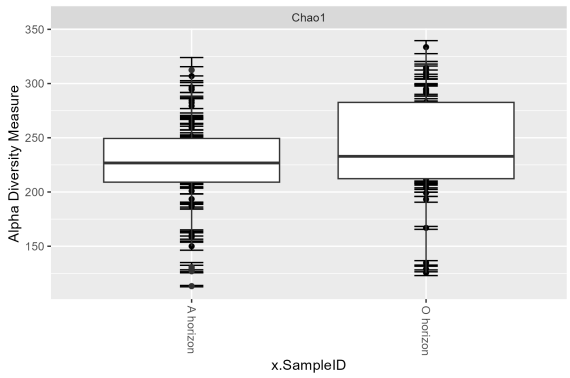
Purpose:
To create soil richness diversity plots with different metrics.

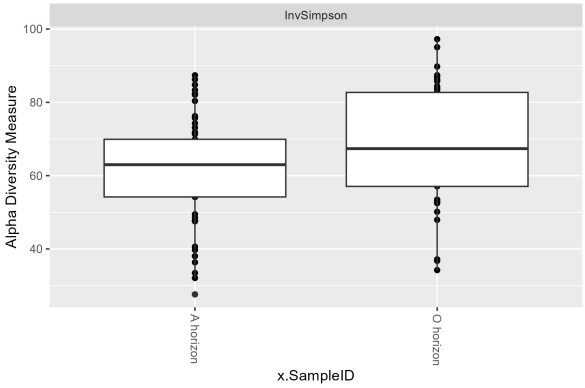
Materials:
Dell laptop with Windows 11

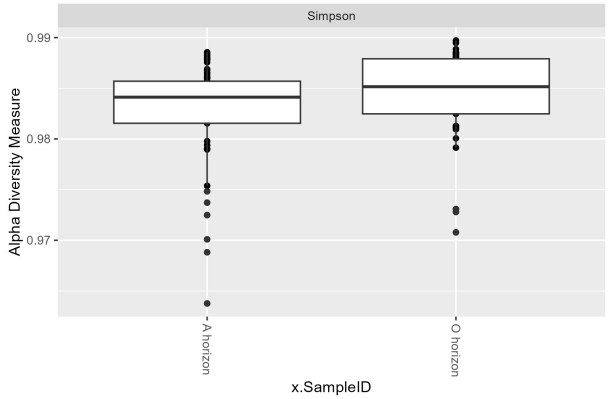
Methods:

1. Load in tidyverse, dplyr, phyloseq, ape, vegan, indicspecies, ggpubr packages
2. Use the plot_richness() function on all data to view all alpha diversity metrics.
3. Use plot_richness() specifically with the Chao1, Simpson and Inverse Simpson measures and the variable horizon.
4. Save all new plots.

Results and Observations:
All new alpha diversity plots have overlapping medians.





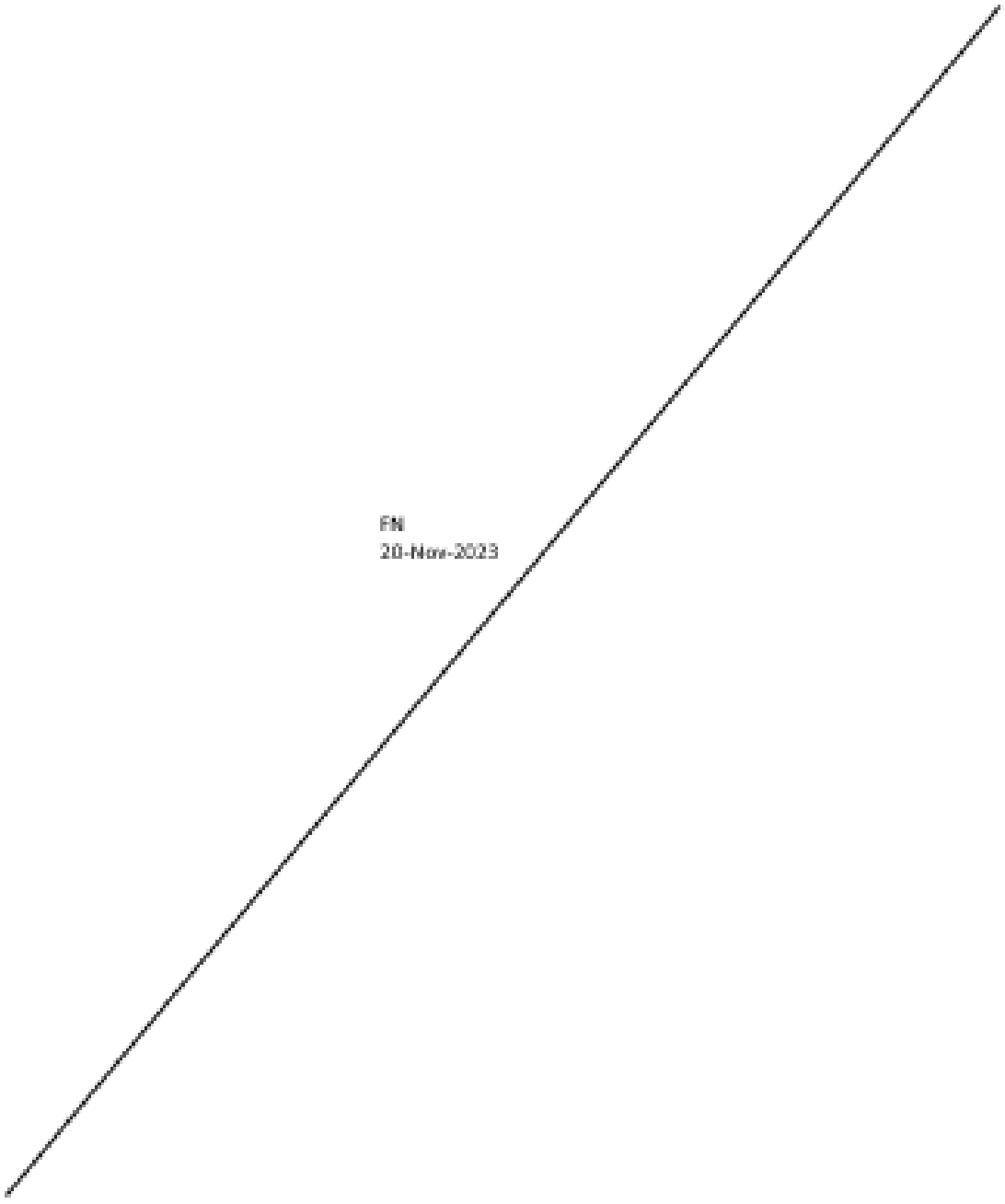


Conclusion:

Alpha diversity is not significantly distinct across the two horizons with any metric.

Future Directions:

Add significance to the Shannon diversity plot.



START

W12-TM: Week 12 Team Meeting

Agenda

- Go through the data summary
- Alpha and Beta Diversity Plots - technically complete but alpha significance tests are slightly concerning...
    - trying it with different metrics? But then the t-test code doesn't run???
    - put all the stuff into slides - would appreciate everyone's opinion!
- Downstream errors and issues after ISA filtering
    - could it be due to NA and 0?
    - should we remove the x/sum(x) code?
- Core microbiome and heat map

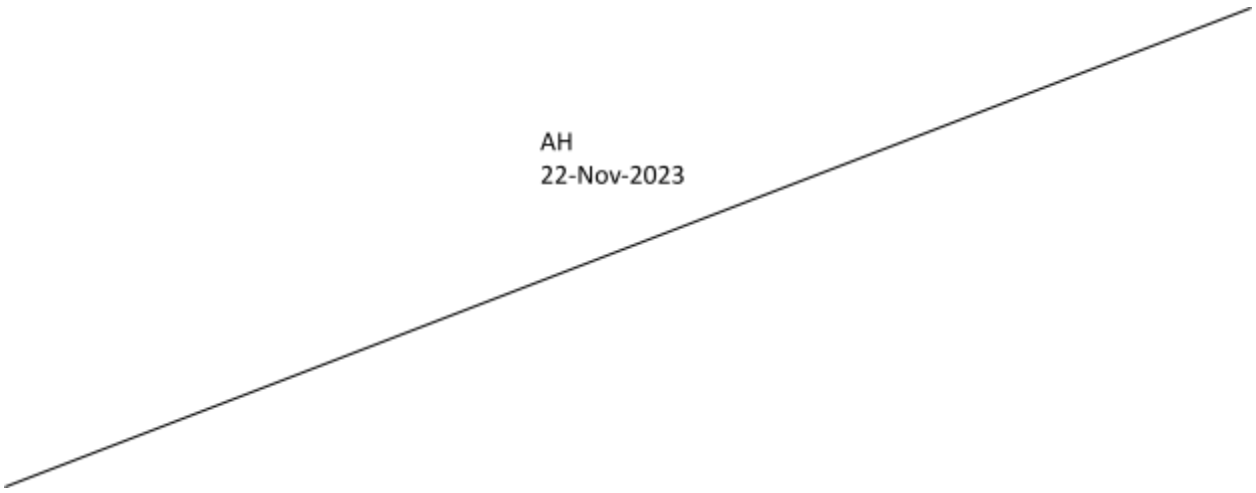
Meeting Minutes

- overview of project for Evelyn
- Alpha diversity:
    - not significant
    - only seems to be including one horizon??
    - send code to Avril!!
    - If medians fall within box of the other, not significant
    - good to show Shannon ones that diversity doesn't differ between then
- Beta diversity:
    - Weighted unifrac is better
    - can get rid of unweighted unifrac
    - less clustering, richness and phylogenetic distance are more likely to be driving diversity
    - difference in evenness is a bit less
    - can present BOTH weighted and unweighted
    - only 1.4% variance
    - draw ellipses in
    - MAKE 3 PANNELS: alpha, weighted, unweighted
    - f-statistic is used to calculate p-value
- ISA
    - everything very significant
    - core microbiome??
    - unable to do ISA with filtered data

    - error might be a dividing by 0 problem
    - already in relative abundance, don't need to do it again
    - can fully delete that line
    - don't need the tree so that warning is fine (error tree replaced with NULL)
    - use view() to look at data
    - prevalence is too stringent
    - tax_glom before running filtering
    - SET stat > 0.8 threshold and build a table (abundance 0.001, prevalence 0.1)
- MetaCyc error bar plot
    - log2fold change cut off to filter (0.58) 1.5 fold change in log2
    - crop out p-value
        - labs to edit titles and customize
- MetaCyc PCA plot
    - can use to show that everything is significant
    - don't need heat map
- Core microbiome
    - use tax_glom to genus level
    - set_1 is AH_core
    - after doing genus level, can put in paper
- Bubble plot:
    - do in ggplot
    - email Avril if we need help

Next week

    - practice presentation!!!
    - outline for manuscript



Date: 22-Nov-2023 Signature/Initial: AH END

START

JAAFV4723-050: Core Microbiome R Processing

Purpose:
To investigate key microorganisms in A/O Horizons with specified levels of abundance and prevalence.

Materials:
Laptop with Mac iOS

Methods:

1. Load the necessary packages and required data, including soil metadata, OTU table, taxonomy file, and phylogenetic tree.
2. Adjust the files for reading into a phyloseq object and create the phyloseq object.
3. Convert the phyloseq object to the Genus level.
4. Transform the phyloseq object to represent relative abundance.
5. Create separate phyloseq objects for each group of interest. "Group1" corresponds to the A Horizon, and "group2" corresponds to the O Horizon.
6. Utilize the core_members function to identify core microbiome members in each group, with a detection (abundance) threshold set at 1e-4 and a prevalence threshold at 0.1.
7. Generate a Venn diagram to visualize the results.

Results and Observations:

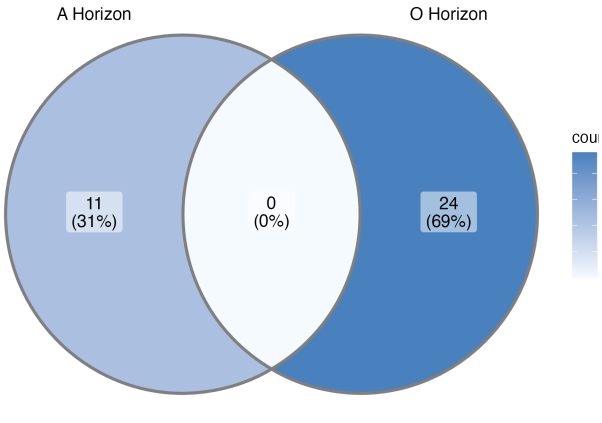


Figure 1. Core Microbiome at ASV level

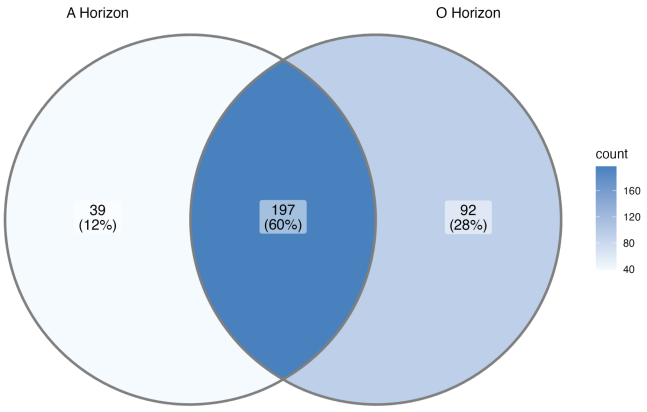


Figure 2. Core Microbiome at Genus level

Discussion:

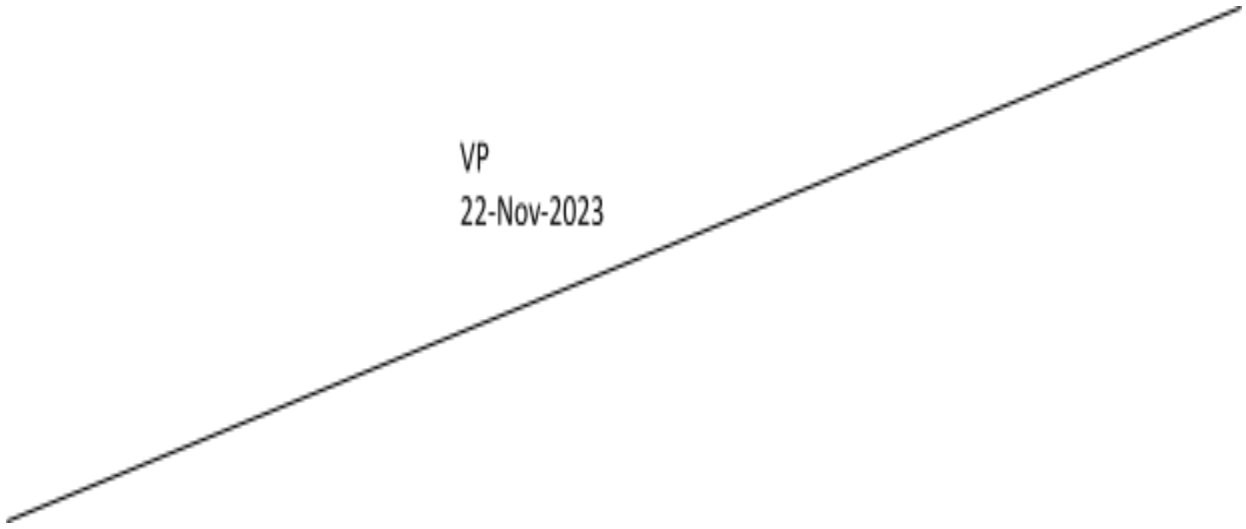
At the ASV level, there was no overlap observed between the A and O horizons, with the O horizon harbouring 38% more species than the A Horizon. Conversely, at the genus level, 60% of the species were shared across both horizons, and the O horizon continued to encompass a larger proportion of species.

Conclusion

The O horizon exhibits a larger portion of the microbiome than the A horizon because it is located above the A horizon. This positioning provides more opportunities for exposure to different environmental conditions and this soil layer contains a high percentage of organic matter, resulting in greater proportion for various microorganisms to thrive.

Future Directions:

The results from the core microbiome analysis will be integrated into the subsequent PICRUSt2 analysis.



Date: 22-Nov-2023 Signature/Initial: VP END

START

JAAFV4723-052: Indicator Species Analysis R Processing Continued

Purpose:
To filter the Aim 2 data in R.

Materials:
Surface laptop with Windows 11

Methods:

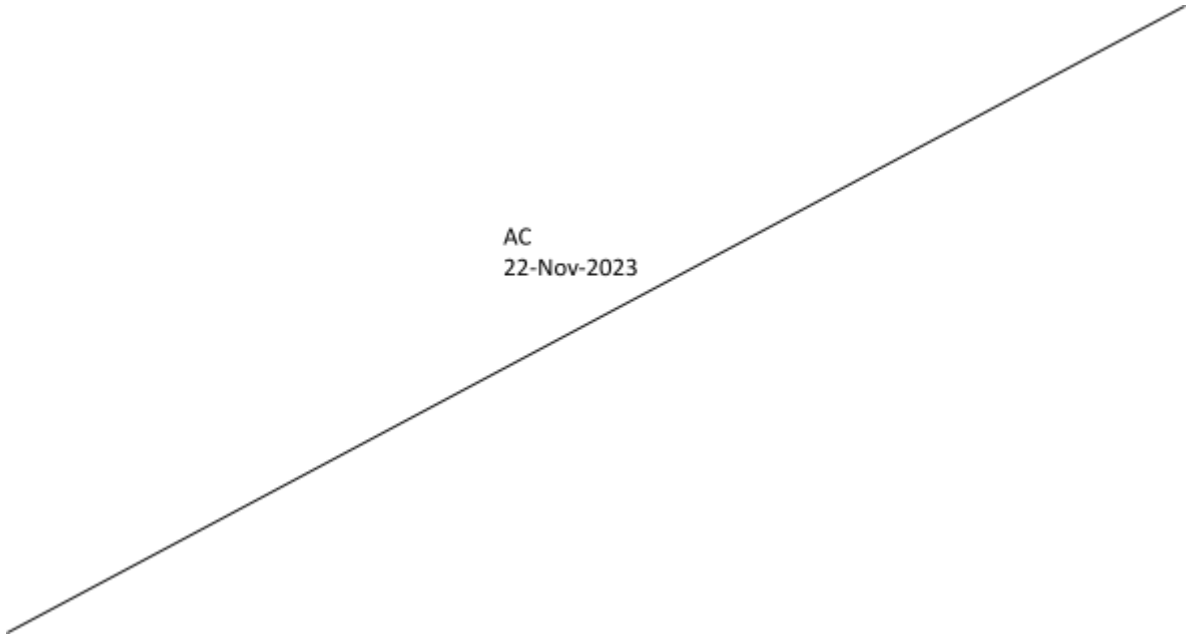
1. Convert phyloseq table reads (from page 28)to relative abundance
2. Filter resultant phyloseq object based on prevalence and abundance thresholds
3. Proceed with steps 2-6 on page 33

Conclusion:

This experiment completed the ISA steps of the project.

Future Directions:

Generate visualization of indicator species analysis in R



Date: 22-Nov-2023 Signature/Initial: AC END

START

JAAFV4723-053: MetaCyc error bar editing

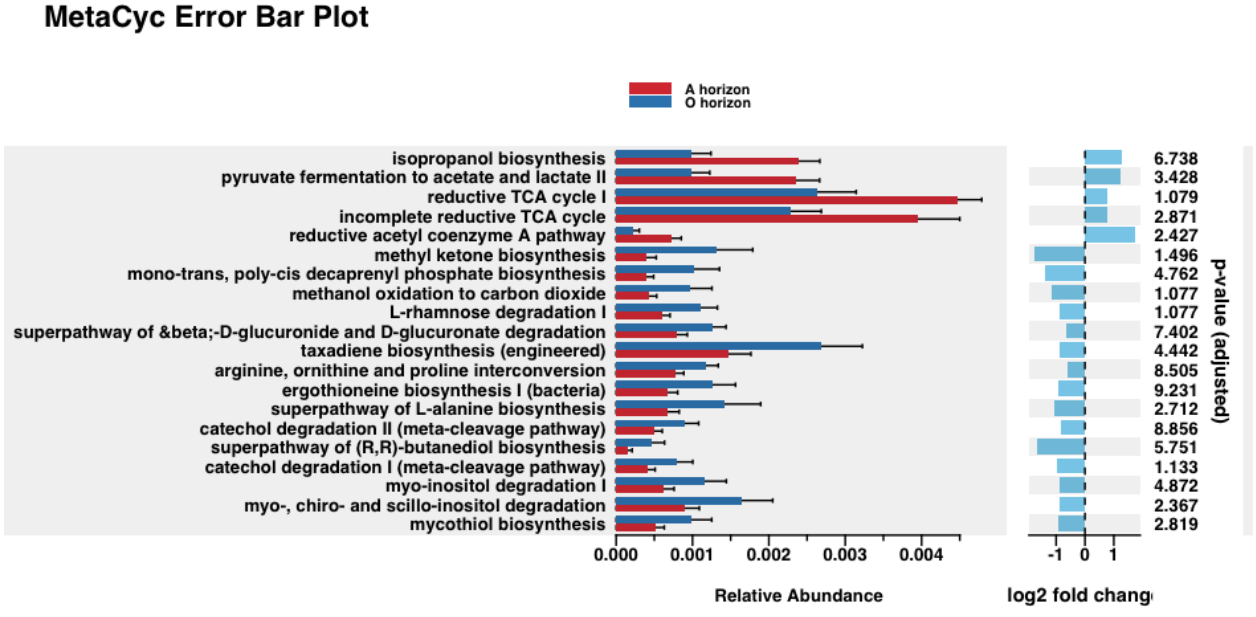
Purpose:
To improve the MetaCyc error bar plots so that they are more easy to interpret for the final manuscript.

Materials:
Laptop with Mac iOS

Methods:

1. Open picrust_analysis.R script in RStudio
2. Sort MetaCyc abundance by p-value
3. Load new error bar function with the ability to do a log2foldchange cutoff
4. Filter by log2fold change of 0.58 (1.5 log 2)
5. Increase subset number of MetaCyc data to 70 to include more data
6. Generate error bar plot

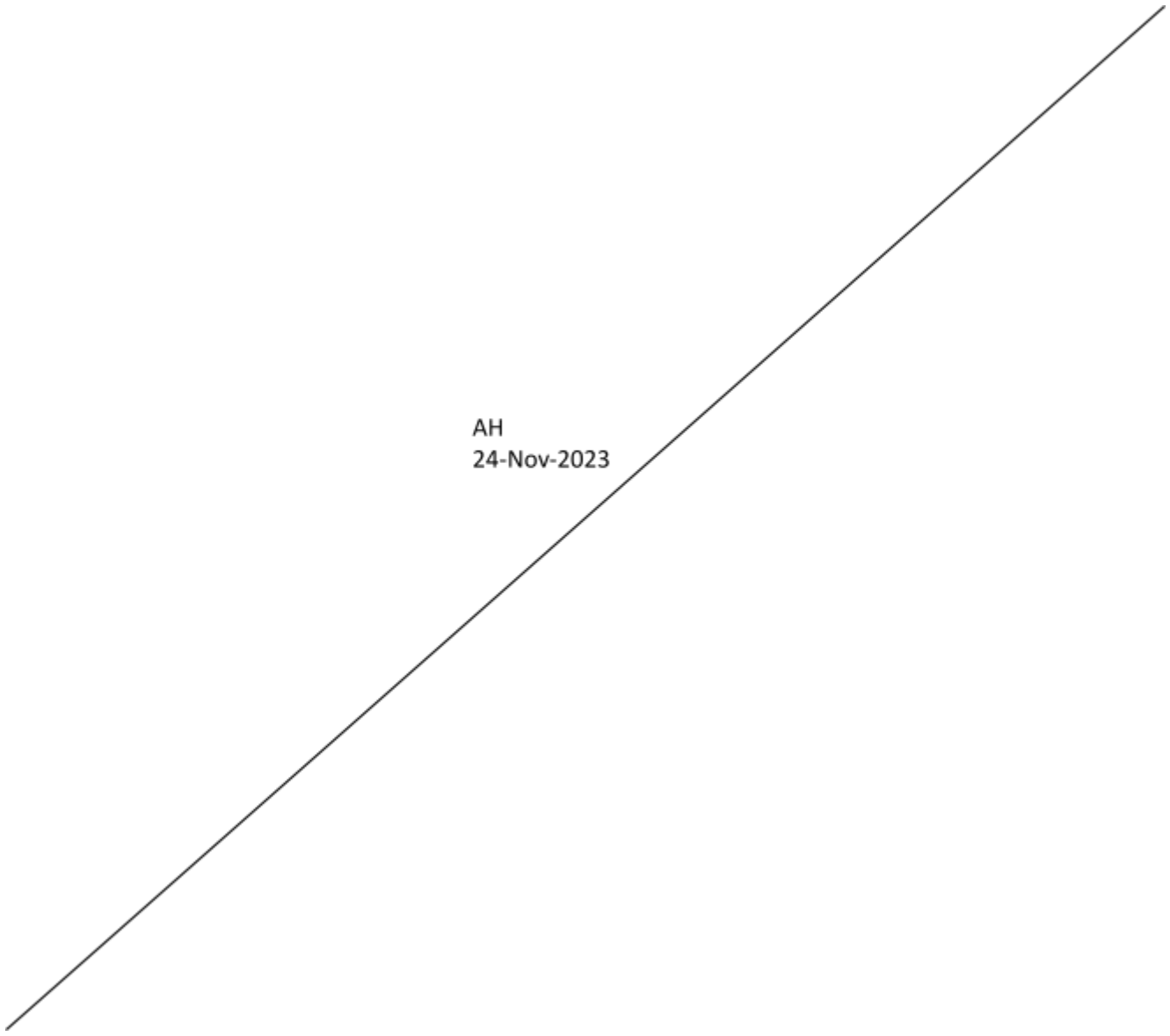
Discussion:
The error bar plot now shows data with higher log2 fold changes so it is easier to focus on functional pathways that could be important.



However, it could still be made more easy to interpret by sorting by the log2 fold change.

Conclusion:
The error bar plot is almost ready for presentation in the manuscript but could still be edited slightly further.

Future Directions:
Sort error bar plot by log2fold change.



Date: 24-Nov-2023 Signature/Initial: AH END

START

JAAFV4723-055: Indicator Species Visualization in R

Purpose:
To generate a scatter plot for visualization of indicator species analysis data.

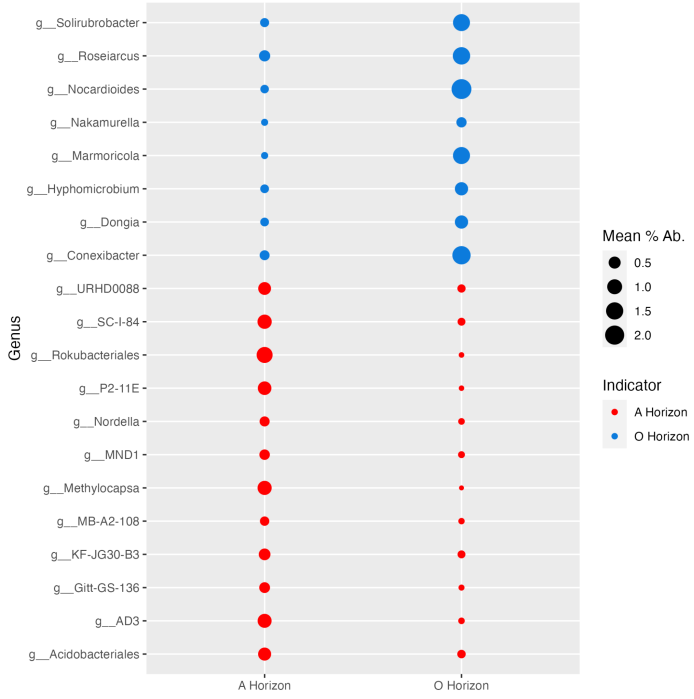
Materials:
Laptop with Mac iOS

Methods:

1. Open Aim_2.R script in RStudio.
2. Load in data from the isa_RA table.
3. Generate scatter plot, with horizon columns side by side.
4. Represent abundance by the size of the plotted points and the colour by whether the point is an indicator for the respective horizon.
5. Save file.

Results:

Figure 1. A graph where the genus is plotted against the horizons with the abundance represented by size.

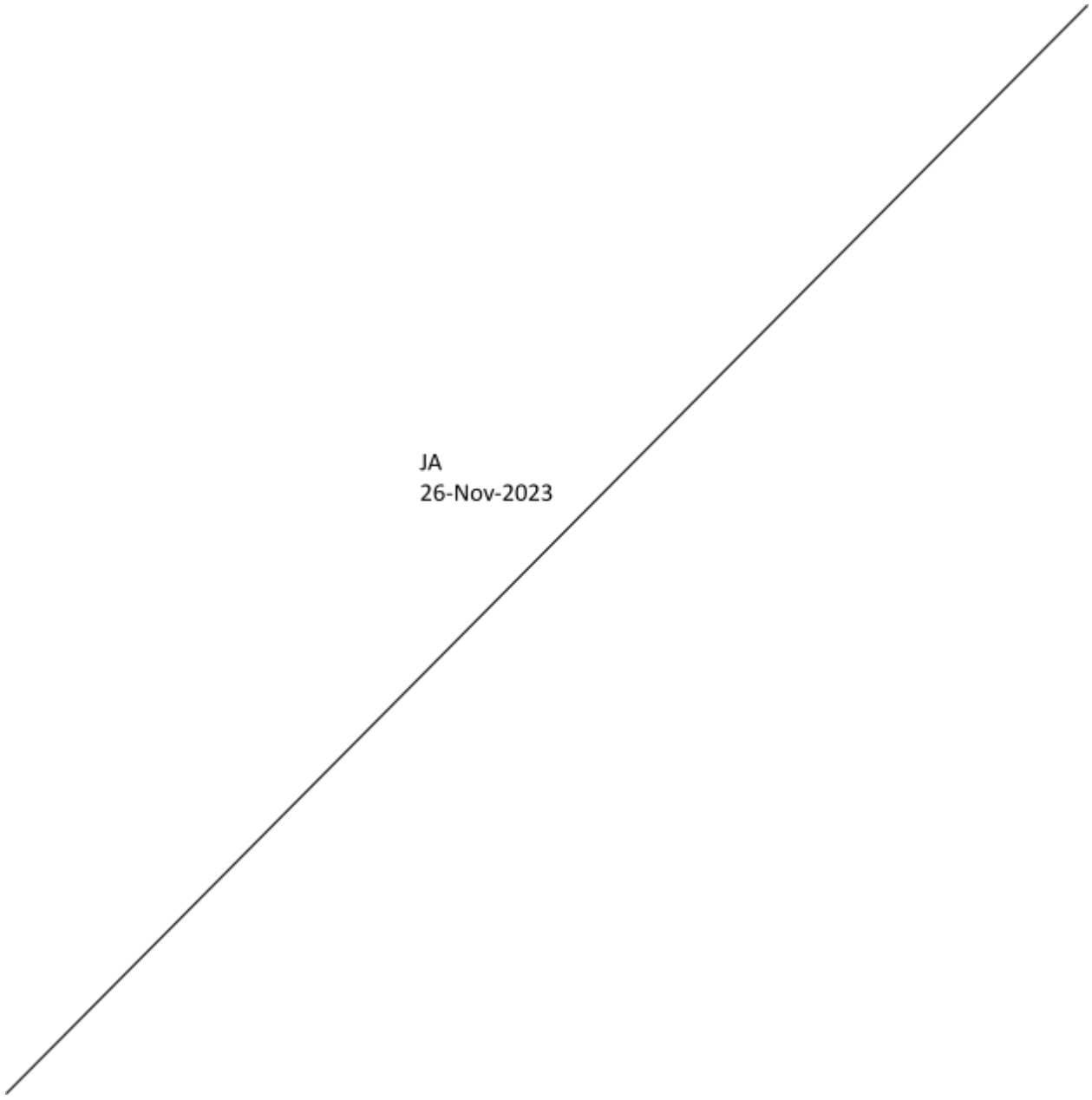


Discussion:

Figure 1 shows us which genus are most abundant for each horizon.

Conclusion:

This experiment completed Aim 2 of the project. The following files were generated to visualize the main indicator species of each group.



Date: 26-Nov-2023 Signature/Initial: JA END

START

W13-TM: Week 13 Team Meeting

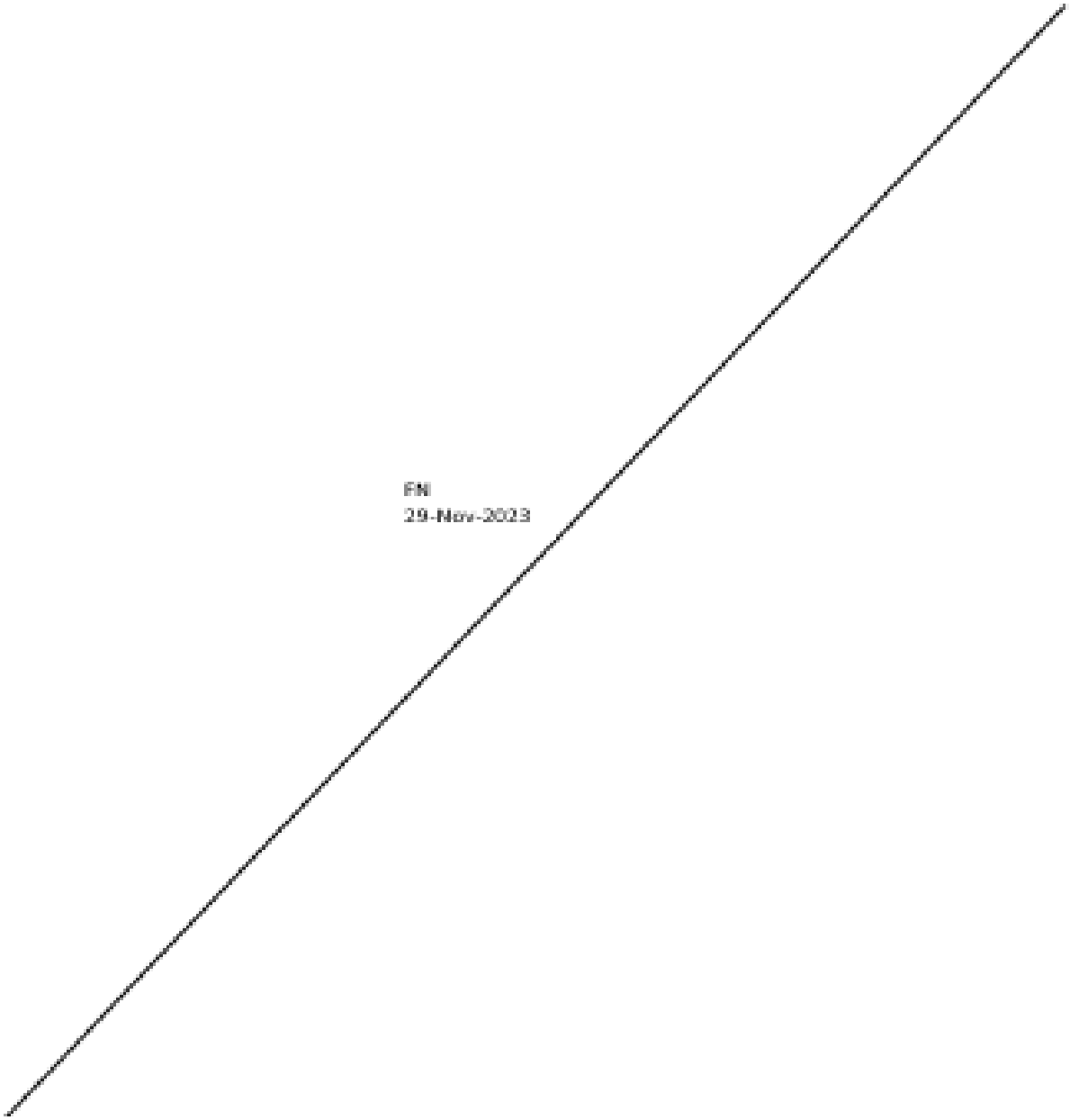
Agenda

- PCA plot: should we talk about the low PC1 and PC2 percentages? Does that matter?
- Error bar plot:
    - what is the relative abundance relative to (lol)
    - looks like O horizon is on average more downregulated and A horizon upregulated?
    - should we look at individual pathways in more detail?
- Manuscript
    - UJEMI style?
    - ideal day to get you a draft for feedback?
- PCoA plots - how to add clustering elipses :(
- Thoughts on future directions? does it really address a "problem" like the rubric says?
- Practice Presentation
- Jastina:
    - fixing plot
    - discussion of data

Meeting Minutes:

- PCA plot: Don't worry about the percentages, don't even mention it - Avril
- Error bar plot:
    - relative to the entire microbiome out of a percentage of the total counts
    - don't put too much weight on the RA for unctional data
    - not so much downregulated but rather there is less of that particular function in that particular horizon
    - log2 is relative to A/O
    - maybe take a glance at the major pathways? but if it doesn't give you anything is 'cool' then don't bother
- Manuscript:
    - search up UJEMI -> for authors -> submission guidelines
    - guidelines are summarized in Module 17
    - try to get in draft manuscript by Tuesday night for guaranteed feedback from Avril
- PCoA plots:
- ISA plot:

    - ask for help about figure legends after, for now edit axis in PP
- Future direction:
    - put it into a wider context: do other authors address which horizon they collect data from? if not, our results show a significant difference so perhaps it would be worthwhile for that to be mentioned
- Notes on presentation:
    - emailed later



Date: 29-Nov-2023 Signature/Initial: FN END

START

JAAFV4723-059: Diversity Metrics Analysis Through R Continued

Purpose:

To add clustering ellipses, significance value to alpha box plot and colours to both diversity plots.

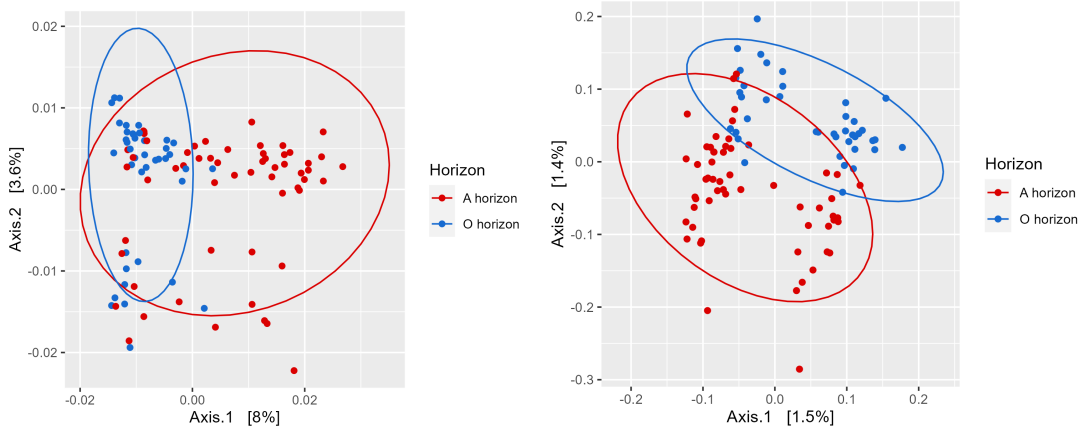
Materials:

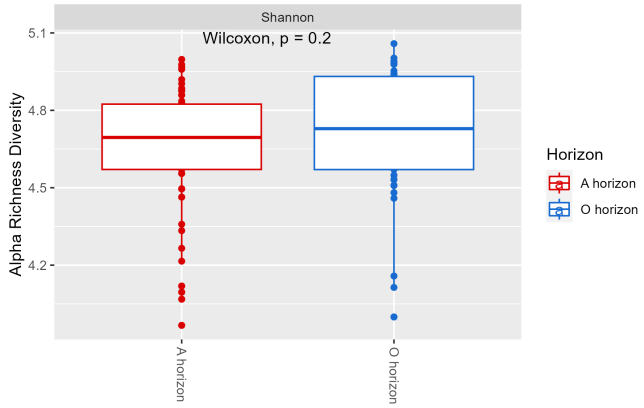
Dell laptop with Windows 11

Methods:

1. Load in tidyverse, dplyr, phyloseq, ape, vegan, indicspecies, ggpubr packages
2. Use the stat_compare_means() function to add p-value to the alpha plot.
3. Set up a variable with colour palette of red and blue.
4. Use the color_palette() function to change the colours of the alpha plot
5. Use stat_elipse() on the beta plots to add ellipses
6. Use the color_palette() function to change the colours of both beta plots
7. Save all new plots.

Results and Observations:



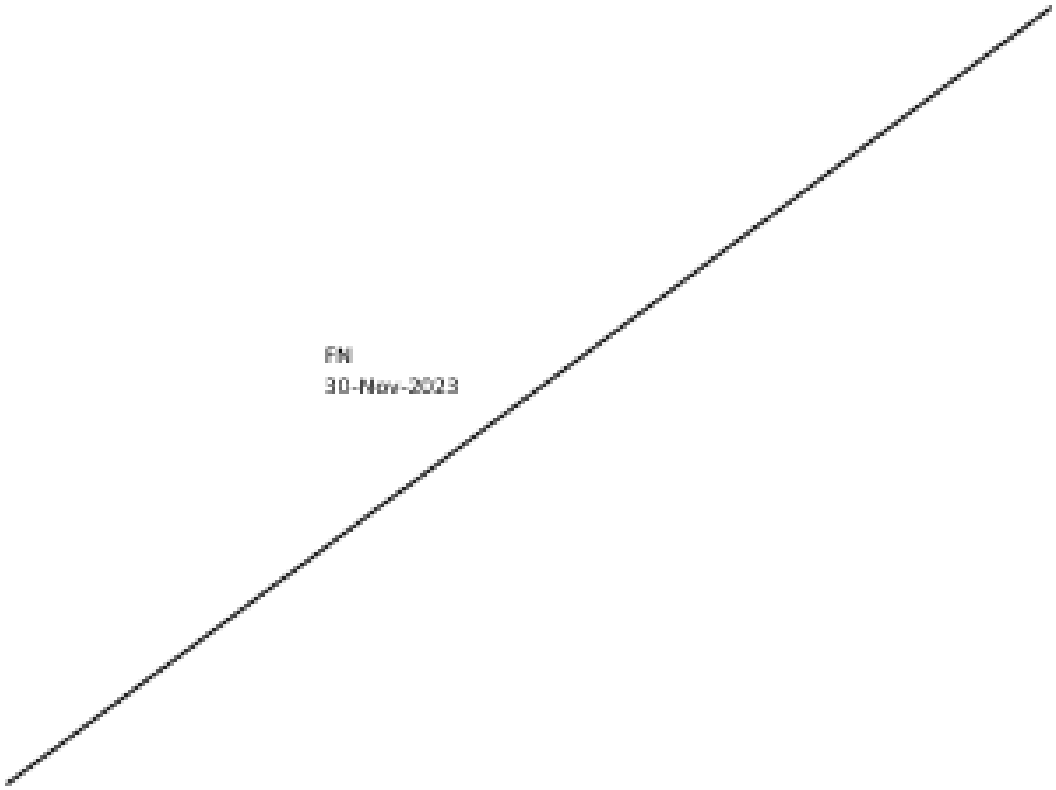


Conclusion:

Clustering in both beta plots is visible

Future Directions:

Add plots to manuscript and presentation.



Date: 29-Nov-2023 Signature/Initial: FN END