Fecal Microbiome Bioinformatics Analysis GEN711 Final Project

Riley Wilson, Zachariah Gravel-Blaney, John Kelleher

Background: Original Data

- This data was taken from a study on the correlation between the human gut microbiome and autism spectrum disorder (ASD) gastrointestinal issues

- The main goal of this study was to evaluate the effect of Microbiota Transfer Therapy (MTT) on the GI of ASD-diagnosed children and overall how that impacted ASD symptoms

Kang, DW., Adams, J.B., Gregory, A.C. et al. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome* 5, 10 (2017). https://doi.org/10.1186/s40168-016-0225-7

Background: Data Collection

- One group received the MTT fecal transplant (treatment group) and one group did not receive the transplant (control group)
- Stool collections and fecal swab were collected bi-weekly for 12 weeks

- DNA was isolated from both sample collection avenues with PowerSoil® DNA Isolation Kit
- A 16S rRNA library prep from Illumina MiSeq was performed to amplify specifically bacterial and archaeal DNA
  - Primer tag: 515f-806r
- 10% of total data was analyzed

Methods: Flowchart
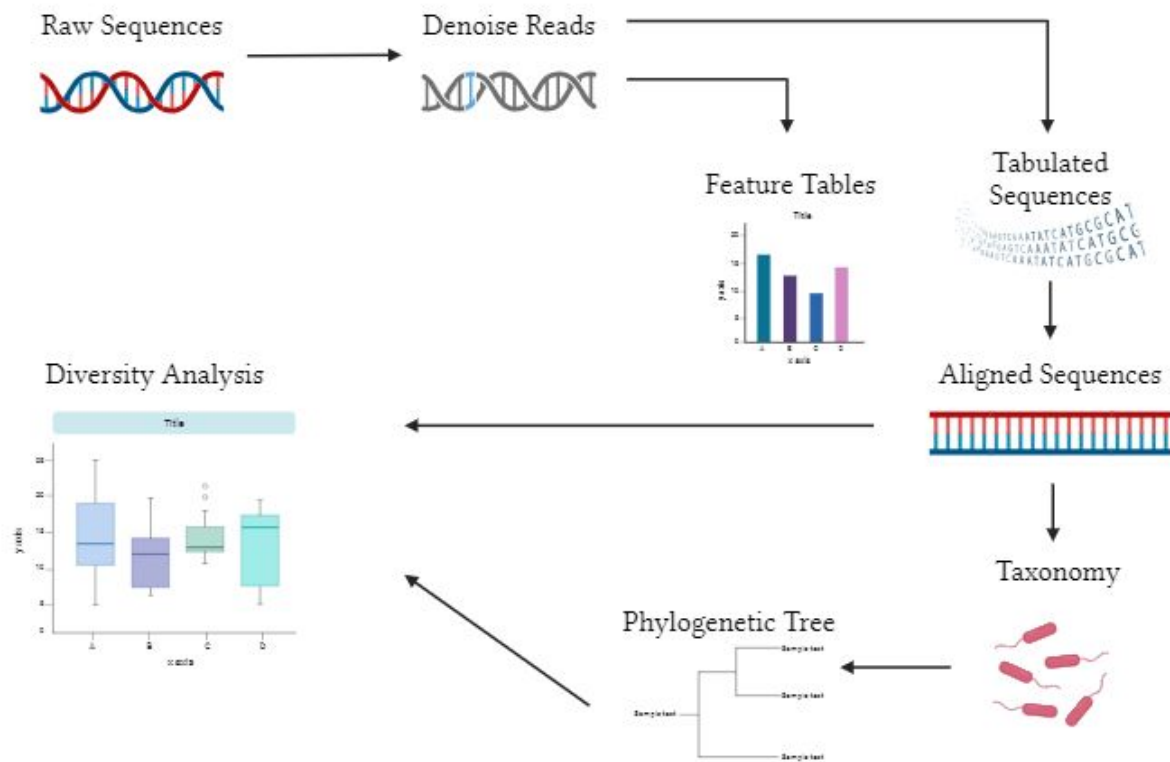
Methods: Data Import and Qiime Installation

- Performed "wget" command to download metadata and raw reads
  - Downloaded forward and reverse reads

- Installed "qiime2-2022.2" onto local device using "wget" and "conda env create"
  - Enabled all "qiime2" analysis to be done in terminal

Methods: Denoising

- "qiime demux summarize"
  - first 13 bases will be trimmed
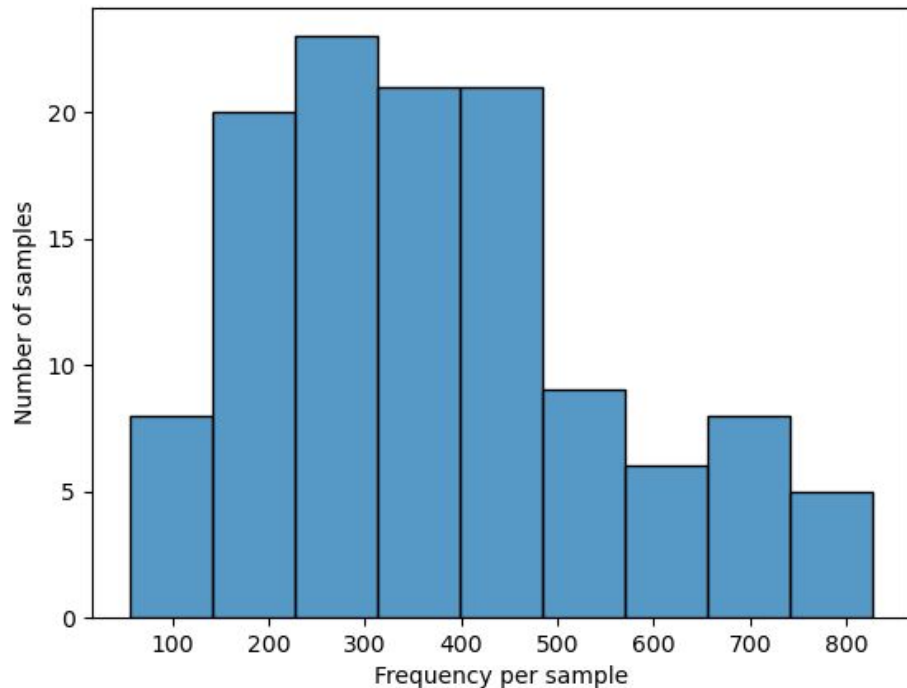 - 150 bases total length
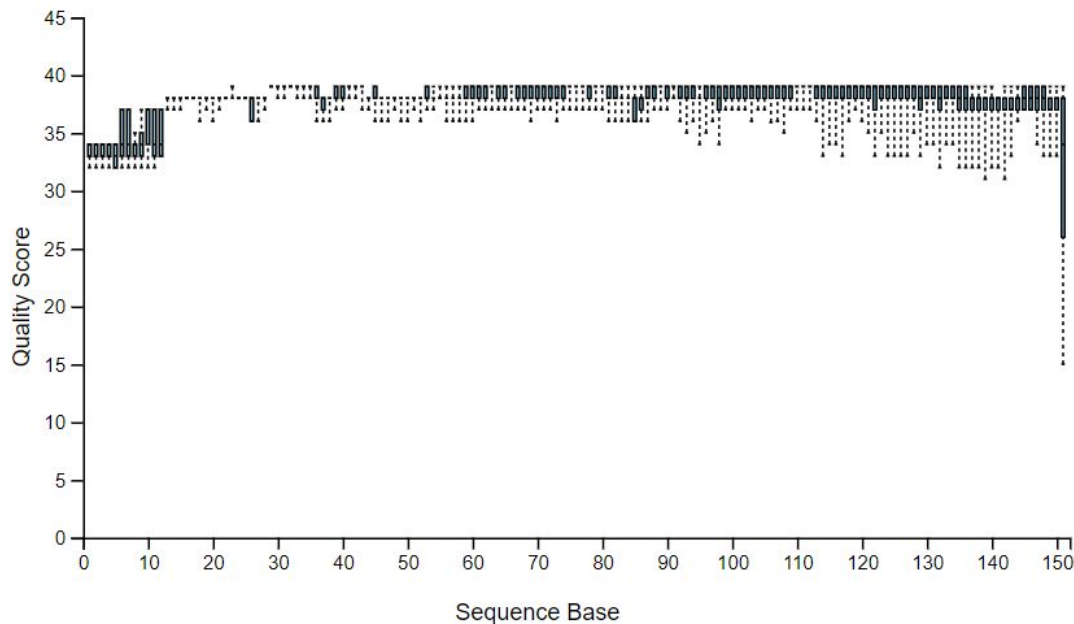
Methods: Denoising continued

- "qiime dada2 denoise-single"
  - trimmed and truncated both raw reads

- "qiime metadata tabulate"
  - output table containing all read information (feature ID, filtering statistics, how much of the data was denoised/passed filter)

sample-id #q2:types	input numeric	filtered numeric	percentage of input passed filter numeric	denoised numeric	non-chimeric numeric	percentage of input non-chimeric numeric
0bb5e24a.aa34.48e7.b1f9.e761ac2dc6b4	6621	6488	97.99	6394	6141	92.75
101c02ce.4f8d.4394.83a3.0c406831d934	6634	6545	98.66	6473	6473	97.57
104e5902.1c3e.417f.bb22.124bfca61a61	7952	7795	98.03	7663	7040	88.53

Methods: Merging Reads and Alignment

- "qiime feature-table merge"
  - merge denoised sequences together and create new aligned clean DNA
- "qiime feature-table summarize"
  - create frequency table from merged sequences
- "qiime feature-table tabulate-seqs"
  - show table of merged DNA

Methods: Merging Reads and Alignment continued

"qiime feature-table summarize" (left) and "qiime feature-table tabulate-seqs" (right)

Methods: Taxonomy Assignment

- Downloaded a reference classifier for human stool from SILVA rRNA database using "wget"

- "qiime feature-classifier classify-sklearn" and "qiime metadata tabulate"
  - uses reference classifier to compare sequences to known bacteria in human stool and create a list of taxa found

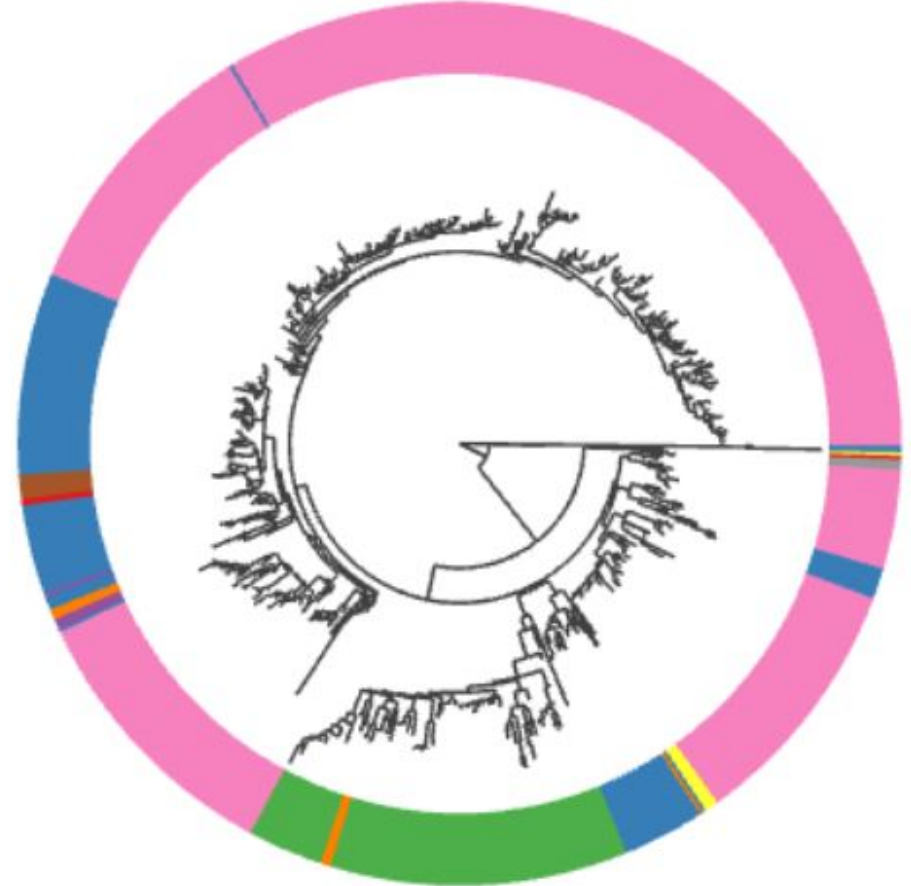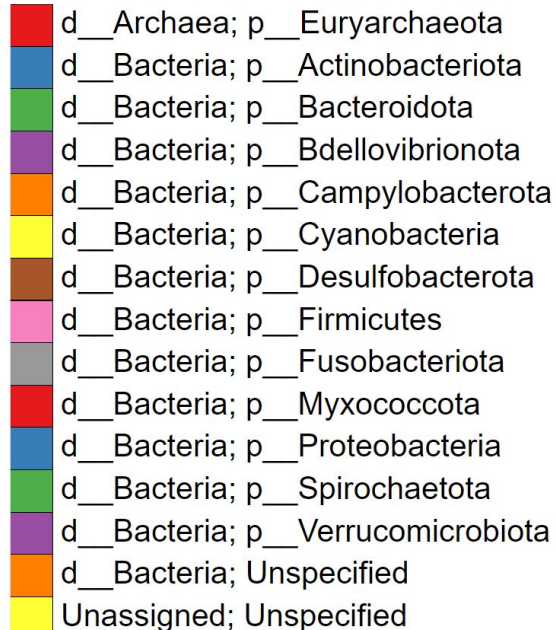| Feature ID #q2:types	Taxon categorical	Confidence categorical
002e78333d6cf2b11aa7a5ba03dd2c68	d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Marinifilaceae; g__Butyricimonas	0.9997788497367014
0046913ae6f9e12dbd889671ed26c09d	d__Bacteria; p__Firmicutes; c__Negativicutes; o__Veillonellales-Selenomonadales; f__Veillonellaceae; g__Megasphaera; s__uncultured_bacterium	0.9685055485925435

Methods: Phylogenetic Tree Assembly

- "qiime phylogeny align-to-tree-mafft-fasttree"
  - utilizes all of the aligned sequences and taxonomic analysis to construct data for a phylogenetic tree

- "qiime empress tree-plot"
  - had to install the "empress" extension for qiime
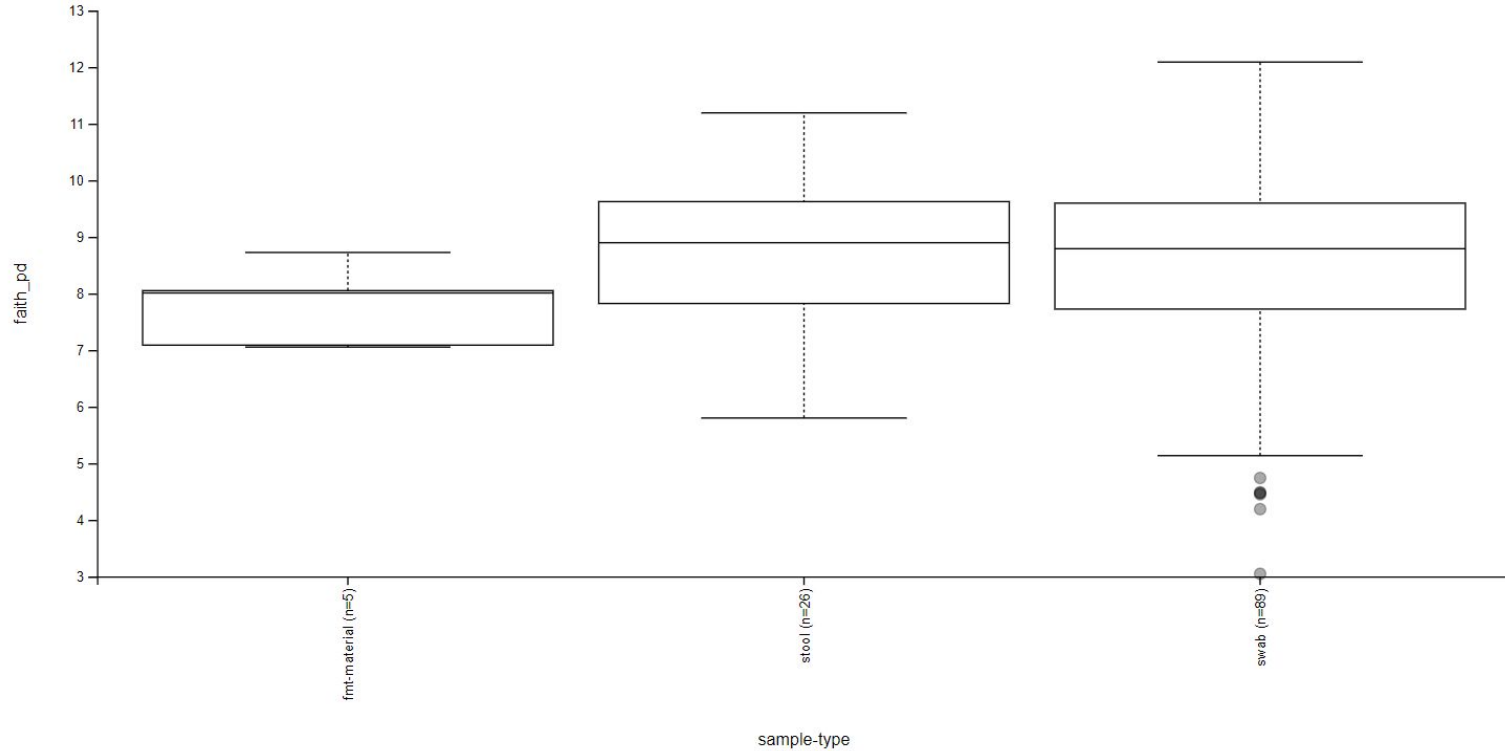 - creates a physical phylogenetic tree for the data

Methods: Alpha and Beta Diversity Analysis

- "qiime diversity core-metrics-phylogenetic"
  - set the initial parameters for alpha and beta analysis
- "qiime diversity alpha-group-significance"
  - produced graphs regarding the data richness and evenness
- "qiime diversity beta-group-significance"
  - produced several graphs to compare how the treatment groups and sample types are associated
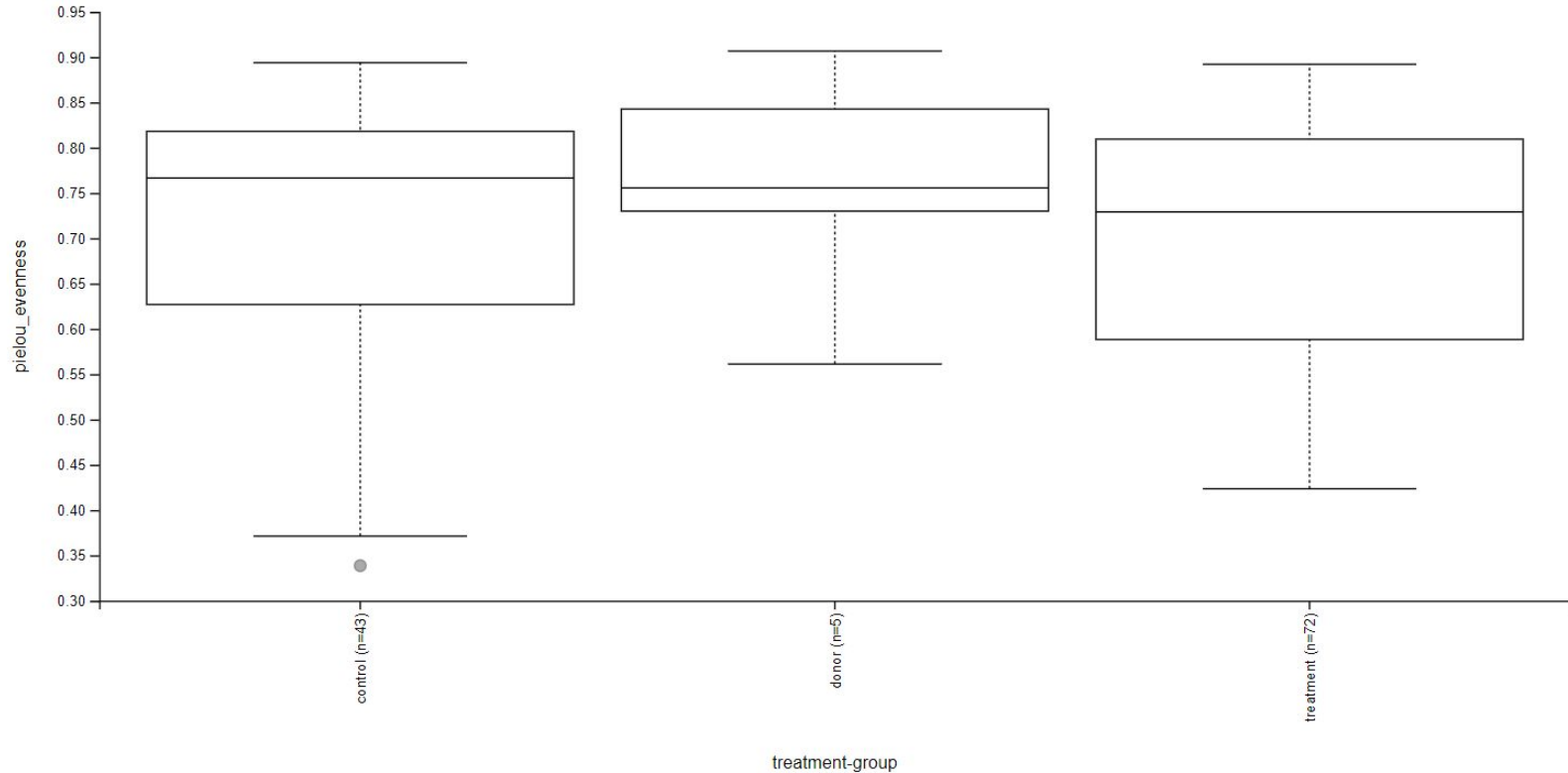
Results: Phylogenetic Tree

Legend:
- d__Archaea; p__Euryarchaeota
- d__Bacteria; p__Actinobacteriota
- d__Bacteria; p__Bacteroidota
- d__Bacteria; p__Bdellovibrionota
- d__Bacteria; p__Campylobacterota
- d__Bacteria; p__Cyanobacteria
- d__Bacteria; p__Desulfobacterota
- d__Bacteria; p__Firmicutes
- d__Bacteria; p__Fusobacteriota
- d__Bacteria; p__Myxococcota
- d__Bacteria; p__Proteobacteria
- d__Bacteria; p__Spirochaetota
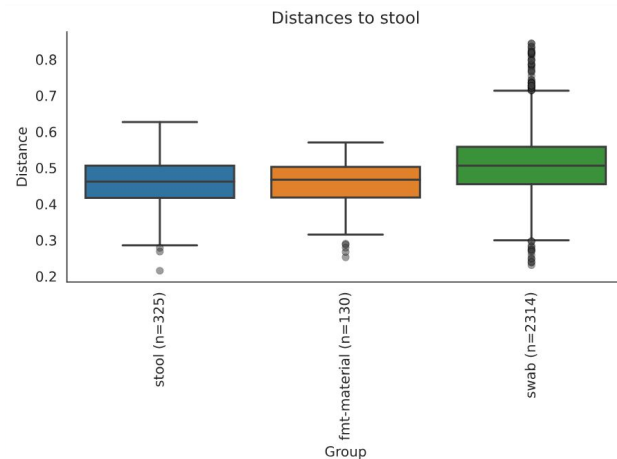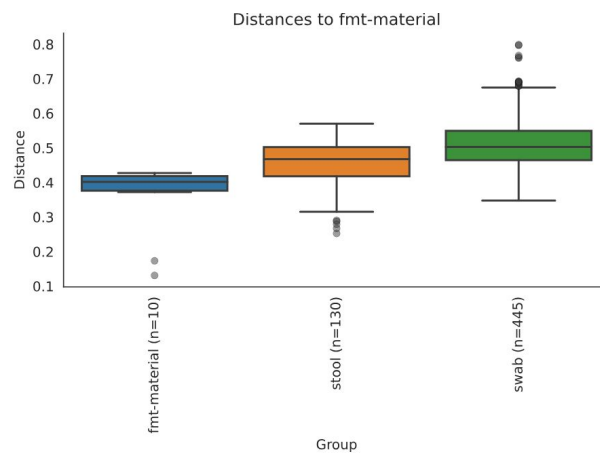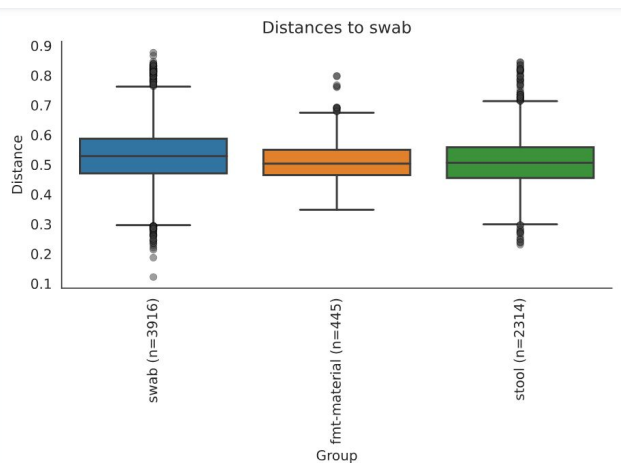- d__Bacteria; p__Verrucomicrobiota
- d__Bacteria; Unspecified
- Unassigned; Unspecified
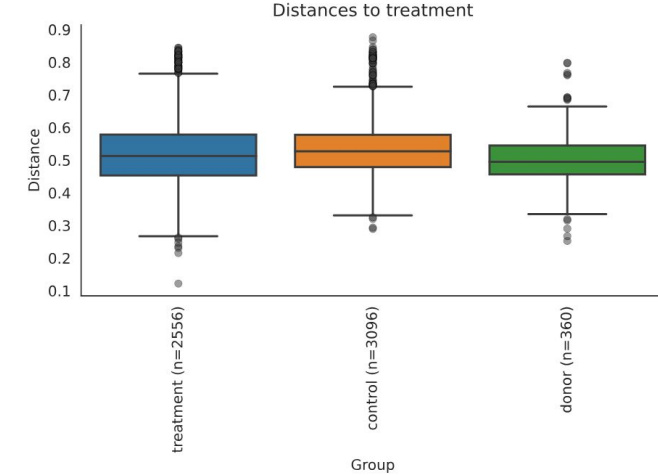
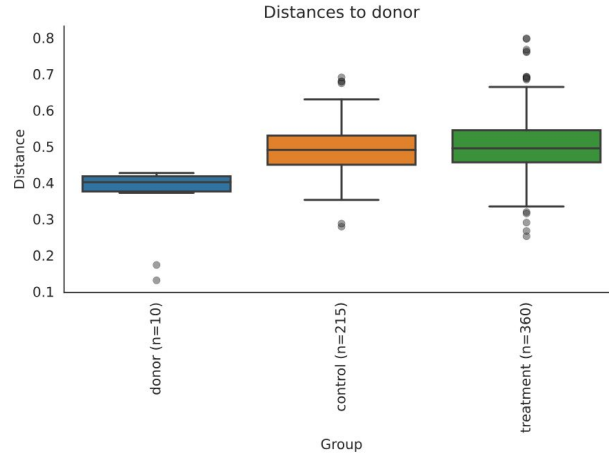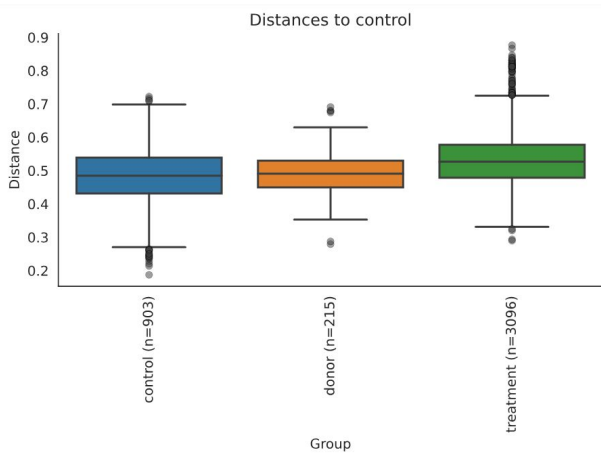# Results: Alpha Diversity (Richness)

Results: Alpha Diversity (Evenness)

Results: Beta Diversity (Sample Group)

Results: Beta Diversity (Treatment Group)

Major Takeaways:

- qiime2 is an incredibly useful tool to analyze raw reads and transform them into visual statistics
- The most difficult part of this project was trying to download all files onto a local device to be viewed
- Best analysis and result: taxonomic assignment and the phylogenetic tree
  - Best visualization of bacteria present

References

- "Fecal Microbiota Transplant (FMT) Study: An Exercise¶." *QIIME 2 Docs*, docs.qiime2.org/2022.2/tutorials/fmt/. Accessed 10 May 2024.
- Kaehler BD, Bokulich NA, McDonald D, Knight R, Caporaso JG, Huttley GA. (2019). Species-level microbial sequence classification is improved by source-environment information. Nature Communications 10: 4643.
- Kang, DW., Adams, J.B., Gregory, A.C. et al. Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. Microbiome 5, 10 (2017). https://doi.org/10.1186/s40168-016-0225-7
- Bokulich, N.A., Kaehler, B.D., Rideout, J.R. et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6, 90.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucl. Acids Res. 41 (D1): D590-D596
- Robeson, M. S., O'Rourke, D. R., Kaehler, B. D., Ziemski, M., Dillon, M. R., Foster, J. T., & Bokulich, N. A. (2021). RESCRIPt: Reproducible sequence taxonomy reference database management. PLoS Comp. Bio., 17(11).