

Periodontitis

Jaewoong Lee

Seunghoon Kim

Semin Lee

2020-11-27

Contents

| | | |
|----------|------------------------------------|----------|
| 1 | Introduction | 4 |
| 1.1 | Microbiome | 4 |
| 1.2 | Ribosomal RNA | 4 |
| 1.3 | 16S rRNA Gene Sequencing | 4 |
| 1.4 | Periodontitis | 4 |
| 2 | Materials | 4 |
| 2.1 | 16S rRNA Gene Sequencing | 4 |
| 3 | Methods | 4 |
| 3.1 | QIIME2 Workflow | 4 |
| 3.1.1 | Denoising techniques | 4 |
| 3.1.2 | Taxonomy Classification | 4 |
| 3.1.3 | Rarefaction | 4 |
| 3.1.4 | Alpha-diversity | 7 |
| 3.1.5 | Beta-diversity | 7 |
| 3.1.6 | ANCOM | 7 |
| 3.2 | Python Packages | 7 |
| 3.2.1 | Pandas | 7 |
| 3.2.2 | Scikit-learn | 7 |
| 3.2.3 | Matplotlib | 7 |
| 3.2.4 | Seaborn | 7 |
| 3.3 | t-SNE | 8 |
| 4 | Results | 8 |
| 4.1 | Quality Filter | 8 |
| 4.2 | Rarefaction | 8 |
| 4.3 | Alpha-diversity | 8 |
| 4.4 | Beta-diversity | 8 |
| 4.5 | ANCOM | 8 |
| 5 | Discussion | 8 |
| 6 | References | 8 |

List of Tables

| | | |
|----|---|----|
| 1 | Kruskal-Wallis Tests among All Group with DADA2 | 9 |
| 2 | Kruskal-Wallis Tests from Evenness Index with DADA2 | 10 |
| 3 | Kruskal-Wallis Tests from Faith PD Index with DADA2 | 11 |
| 4 | Kruskal-Wallis Tests from Observed Features Index with DADA2 | 11 |
| 5 | Kruskal-Wallis Tests from Shannon's Diversity Index with DADA2 | 11 |
| 6 | Kruskal-Wallis Tests among All Group with Deblur | 11 |
| 7 | Kruskal-Wallis Tests from Evenness Index with Deblur | 13 |
| 8 | Kruskal-Wallis Tests from Faith PD Index with Deblur | 13 |
| 9 | Kruskal-Wallis Tests from Observed Features Index with Deblur | 13 |
| 10 | Kruskal-Wallis Tests from Shannon's Diversity Index with Deblur | 13 |
| 11 | Bray-Curtis Distance Index with DADA2 | 15 |
| 12 | Jaccard Distance Index with DADA2 | 15 |
| 13 | Unweighted UniFrac Distance Index with DADA2 | 15 |
| 14 | Weighted UniFrac Distance Index with DADA2 | 15 |
| 15 | Bray-Curtis Distance Index with Deblur | 15 |
| 16 | Jaccard Distance Index with Deblur | 15 |
| 17 | Unweighted UniFrac Distance Index with Deblur | 18 |
| 18 | Weighted UniFrac Distance Index with Deblur | 18 |
| 19 | ANCOM Significant Taxa with DADA2 and Greengenes | 20 |
| 20 | ANCOM Significant Taxa with DADA2 and SILVA | 22 |
| 21 | ANCOM Significant Taxa with Deblur and Greengenes | 23 |
| 22 | ANCOM Significant Taxa with DADA2 and SILVA | 24 |

List of Figures

| | | |
|----|---|----|
| 1 | Concept of a Core Human Microbiome (Turnbaugh et al., 2007) | 5 |
| 2 | A Theoretic Overview of QIIME2 Workflow (Bolyen et al., 2019, 2018) | 5 |
| 3 | Denoising Techniques which provided by QIIME2 | 6 |
| 4 | Taxonomy Classification which provided by QIIME2 | 6 |
| 5 | Example ANCOM Volcano Plot which Provided by QIIME2 (Bolyen et al., 2019, 2018) | 9 |
| 6 | Visualization by t-SNE (Maaten & Hinton, 2008) | 9 |
| 7 | Sequence Quality Plot | 9 |
| 8 | Frequency per Sample by DADA2 | 10 |
| 9 | Frequency per Sample by DADA2 | 10 |
| 10 | Evenness Index from DADA2 | 11 |
| 11 | Faith PD Index from DADA2 | 12 |
| 12 | Observed Features Index from DADA2 | 12 |
| 13 | Shannon's Diversity Index from DADA2 | 12 |
| 14 | Evenness Index from Deblur | 13 |
| 15 | Faith PD Index from Deblur | 14 |
| 16 | Observed Features Index from Deblur | 14 |
| 17 | Shannon's Diversity Index from Deblur | 14 |
| 18 | Bray-Curtis Distance Index with DADA2 | 16 |
| 19 | Jaccard Distance Index with DADA2 | 16 |
| 20 | Unweighted Unifrac Distance Index with DADA2 | 17 |
| 21 | Weighted Unifrac Distance Index with DADA2 | 17 |
| 22 | Bray-Curtis Distance Index with Deblur | 18 |
| 23 | Jaccard Distance Index with Deblur | 19 |
| 24 | Unweighted Unifrac Distance Index with Deblur | 19 |
| 25 | Weighted Unifrac Distance Index with Deblur | 20 |
| 26 | ANCOM Volcano Plot with DADA2 and Greengenes | 21 |
| 27 | ANCOM Volcano Plot with DADA2 and SILVA | 21 |
| 28 | ANCOM Volcano Plot with Deblur and Greengenes | 21 |
| 29 | ANCOM Volcano Plot with Deblur and SILVA | 26 |

1 Introduction

1.1 Microbiome

Microbiome is consist of microbiota, the micro-organisms which live inside and on humans (Turnbaugh et al., 2007). Microbiome is also about 10^{13} micro-organisms whose which collective genome (Gill et al., 2006).

1.2 Ribosomal RNA

Ribosomal RNA (rRNA) is well-known as a key to phylogeny (Olsen & Woese, 1993).

1.3 16S rRNA Gene Sequencing

1.4 Periodontitis

Periodontitis is an inflammatory conditions which effecting periodontium, tissues which surround and support teeth. Major components of periodontitis are clinical attachment loss and bone loss (Flemmig, 1999). Previous study found risk factors of periodontitis such as smoking, diabetes, genetic factors and host response (Van Dyke & Dave, 2005).

2 Materials

2.1 16S rRNA Gene Sequencing

- 100 Healthy samples
- 50 Chronic Early Periodontitis Sample
- 50 Chronic Moderate Periodontitis Sample
- 50 Chronic Severe Periodontitis Sample

3 Methods

3.1 QIIME2 Workflow

QIIME2 is a capable, expandable and distributed microbiome analysis package with transparent analysis (Bolyen et al., 2019, 2018). A theoretic overview of QIIME2 workflow is shown as figure 2.

3.1.1 Denoising techniques

There are two denoising techniques provided by QIIME2: DADA2 (Callahan et al., 2016) and Deblur (Amir et al., 2017). Major difference between DADA2 and Deblur, as shown as figure 3, is a strategy, the strategy used to divide as different species. DADA2 uses amplicon sequence variants (ASVs), strictly divides sequences even one-base mismatch. However, Deblur uses operational taxonomic units (OTUs), considers as same sequence when sequences are 97 % or more matched.

3.1.2 Taxonomy Classification

There are two taxonomy classification databases which provided by QIIME2: Greengenes (GG) (DeSantis et al., 2006) and SILVA (Pruesse et al., 2007). Major difference between Greengenes and SILVA is resolution. Resolution of Greengenes is from kingdom to species; however, resolution of SILVA is from domain to genus. Note that a higher accuracy at taxonomic levels above genus level; but accuracy drops at species level (Gihawi et al., 2019).

3.1.3 Rarefaction

Rarefaction is a statistical method of estimating the number of species expected in a random sample which taken from a collection (James & Rathbun, 1981). Moreover, rarefaction allows comparisons of the species richness among communities. Thus, rarefaction is a good choice for normalization (Weiss et al., 2017).

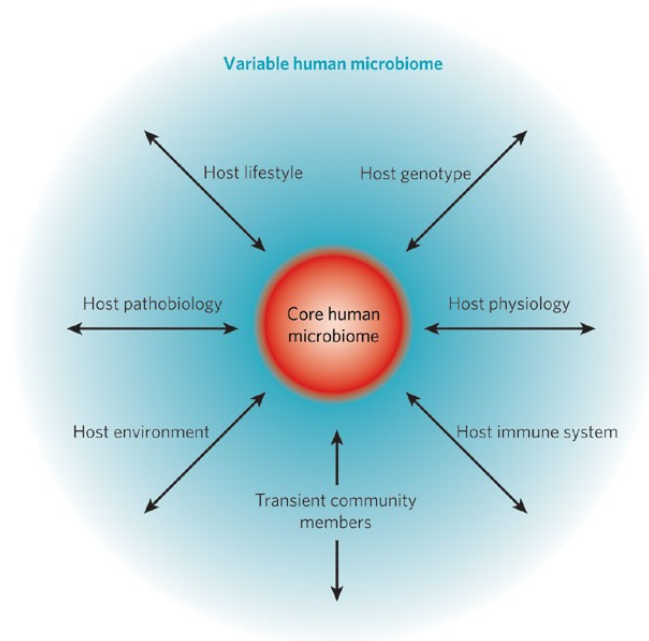


Figure 1: Concept of a Core Human Microbiome (Turnbaugh et al., 2007)

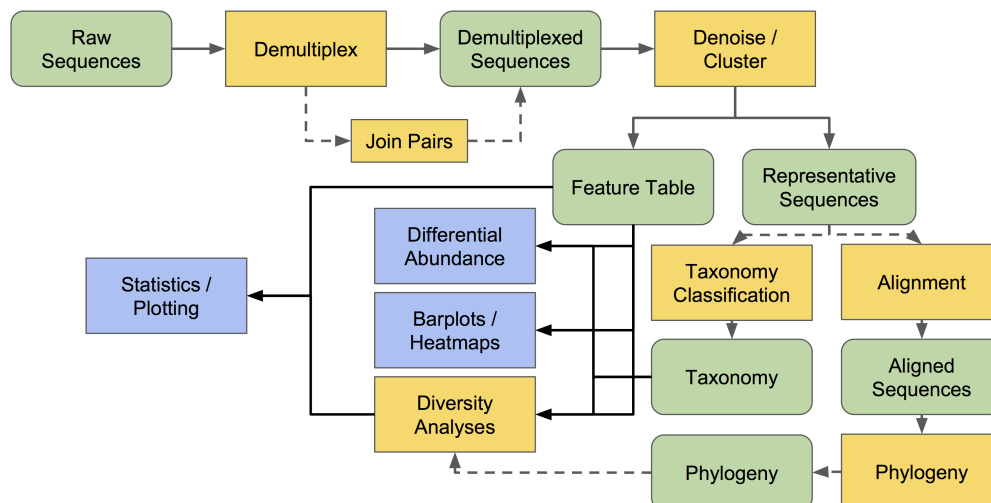


Figure 2: A Theoretic Overview of QIIME2 Workflow (Bolyen et al., 2019, 2018)

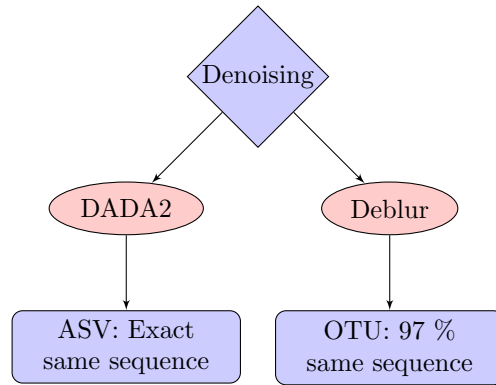


Figure 3: Denoising Techniques which provided by QIIME2

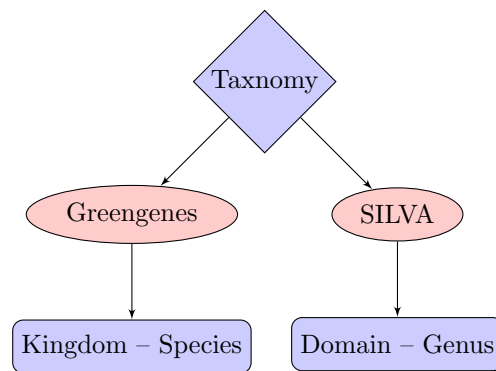


Figure 4: Taxonomy Classification which provided by QIIME2

3.1.4 Alpha-diversity

Alpha-diversity is a metric which shows the richness of taxa at a single community. There are four alpha-diversity indices which provided from QIIME2:

- Evenness index.
- Faith's phylogenetic diversity (Faith PD).
- Observed features.
- Shannon's diversity index.

Shannon's diversity index shows a quantitative measure of community richness; Observed features, however, is a qualitative measure of community richness. Faith's phylogenetic diversity index indicates a qualitative measure of community richness which assimilates phylogenetic relationship among features. Finally, evenness index, as its name, shows a measure of community evenness.

3.1.5 Beta-diversity

Beta-diversity is a metric which indicates the taxonomic differentiation between multiple communities. There are four beta-diversity indices which provided from QIIME2:

- Bray-Curtis distance.
- Jaccard distance.
- Unweighted UniFrac distance.
- Weighted UniFrac distance.

Bray-Curtis distance shows a quantitative of community dissimilarity; Jaccard distance, however, indicates a qualitative measure of community dissimilarity. UniFrac distances reveal a measure of community dissimilarity which consolidates phylogenetic relationship among features. Difference between unweighted UniFrac distance and weighted UniFrac distance is a qualitative and a quantitative, respectively.

3.1.6 ANCOM

ANCOM (Analysis of composition of microbiomes) can be used for analyzing the composition of microbiome in multiple populations (Mandal et al., 2015). Example ANCOM volcano plot is shows as figure 5. In figure 5, two metrics are clearly shown: clr and W. clr stands for centered log ratio, and W is a count of the number of sub-hypothesis which have passed for given species.

3.2 Python Packages

3.2.1 Pandas

Pandas is a Python package of rich data structures and tools for analyzing with structured data sets (McKinney et al., 2011).

3.2.2 Scikit-learn

Scikit-learn grants state-of-the-art implementation of many machine learning algorithms, while controlling an easy-to-use interface tightly integrated the Python code (Pedregosa et al., 2011).

3.2.3 Matplotlib

Matplotlib is a Python graphics package which used for application development, interactive scripting and publication quality image generation (Barrett, Hunter, Miller, Hsu, & Greenfield, 2005). Matplotlib, also, is designed to create simple plots with a few commands (Hunter, 2007).

3.2.4 Seaborn

Seaborn is a Python data visualization package which based on matplotlib, allows a high-level interface for displaying engaging and descriptive statistical graphics (Waskom & the seaborn development team, 2020).

3.3 t-SNE

t-SNE (t-distributed stochastic neighbor embedding) reveals high-dimensional data a location in two-dimensional map (Maaten & Hinton, 2008).

4 Results

4.1 Quality Filter

Longer sequences have more fallen sequence quality than shorter. Thus, sequences which longer than threshold should be trimmed out due to their low quality. However, gold-standard strategy for deciding the threshold does not exist; the threshold is set as longest sequence length which have half of sequences have greater than 30 quality score. Hence, sequence quality plot is shown as figure 7; trimmed length in forward reads is 300, and trimmed length in reverse reads is 265.

4.2 Rarefaction

Sampling depth should be decided for rarefaction. Gold-standard method for determining sampling depth is minimum frequency in the samples. Hence, sampling depth with DADA2 is 3786 (Figure 8), and sampling depth with Deblur is 7253 (Figure 9).

4.3 Alpha-diversity

Alpha-diversity analysis with DADA2 was done: Evenness index (Table 2 and Figure 10), Faith PD (Table 3 and Figure 11), observed feature index (Table 4 and Figure 12) and Shannon's diversity index (Table 5 and Figure 13). Also, alpha-diversity analysis with DADA2 was done: Evenness index (Table 7 and Figure 14), Faith PD (Table 8 and Figure 15), observed feature index (Table 9 and Figure 16) and Shannon's diversity index (Table 10 and Figure 17). Moreover, Kruskal-Wallis tests among all groups are shown as table 1 (with DADA2) and table 6 (with Deblur).

4.4 Beta-diversity

Beta-diversity analysis with DADA2 was done: Bray-Curtis distance (Table 11 and Figure 18), Jaccard distance (Table 12 and Figure 19), unweighted UniFrac distance (Table 13 and Figure 20) and weighted UniFrac distance (Table 14 and Figure 20). Also, beta-diversity analysis with Deblur was done: Bray-Curtis distance (Table 15 and Figure 22), Jaccard distance (Table 16 and Figure 23), unweighted UniFrac distance (Table 17 and Figure 24) and weighted UniFrac distance (Table 18 and Figure 24).

4.5 ANCOM

Statistically significant different taxa and volcano plots by ANCOM were derived: DADA2 and Greengenes (Table 19 and Figure 26), DADA2 and SILVA (Table 20 and Figure 27), Deblur and Greengenes (Table 21 and Figure 28) and Deblur and SILVA (Table 22 and Figure 29).

5 Discussion

6 References

- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Xu, Z. Z., ... others (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, 2(2).
- Barrett, P., Hunter, J., Miller, J. T., Hsu, J.-C., & Greenfield, P. (2005). matplotlib—a portable python plotting package. In *Astronomical data analysis software and systems xiv* (Vol. 347, p. 91).
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., ... others (2018). *Qiime 2: Reproducible, interactive, scalable, and extensible microbiome data science* (Tech. Rep.). PeerJ Preprints.
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... others (2019). Reproducible, interactive, scalable and extensible microbiome data science using qiime 2. *Nature biotechnology*, 37(8), 852–857.
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). Dada2: high-resolution sample inference from illumina amplicon data. *Nature methods*, 13(7), 581–583.

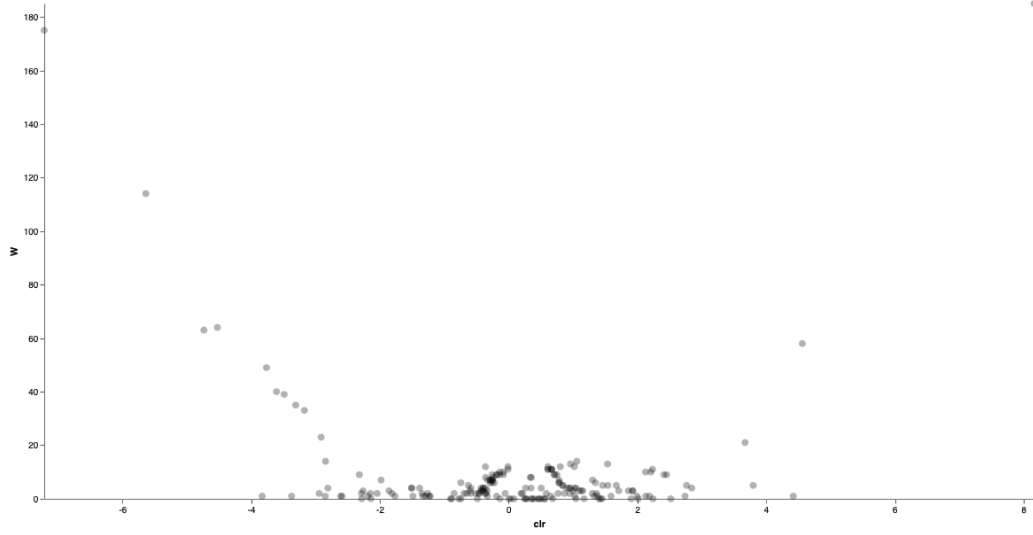


Figure 5: Example ANCOM Volcano Plot which Provided by QIIME2 (Bolyen et al., 2019, 2018)

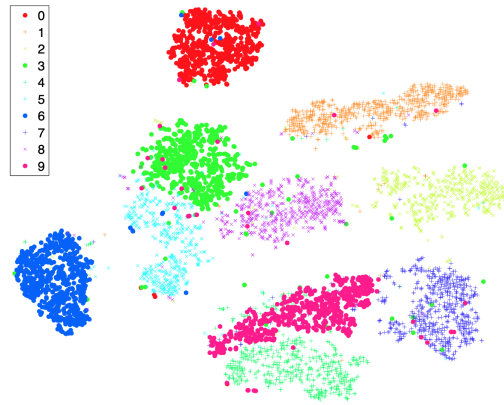


Figure 6: Visualization by t-SNE (Maaten & Hinton, 2008)

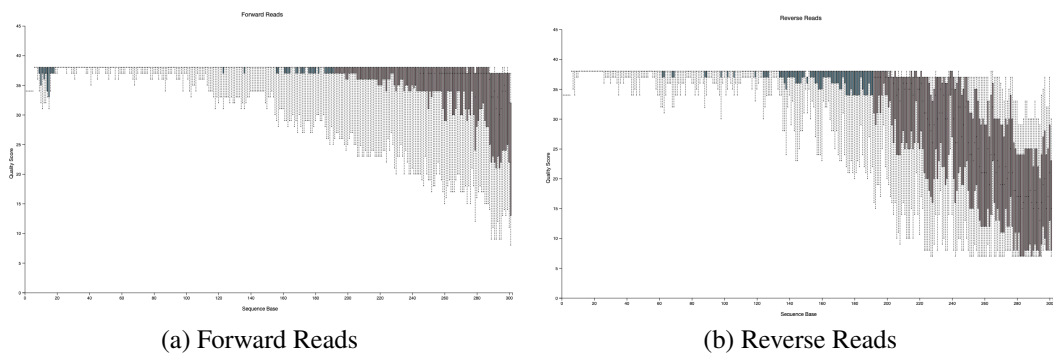


Figure 7: Sequence Quality Plot

Table 1: Kruskal-Wallis Tests among All Group with DADA2

| Alpha-Diversity | H | p-value |
|---------------------|--------------------|-----------------------|
| Evenness | 12.185457848605665 | 0.006774123738087294 |
| Faith PD | 33.42272318725111 | 2.6227945981005624e-7 |
| Observed Features | 21.019370066584198 | 0.0001043055436502384 |
| Shannon's Diversity | 7.311350438247132 | 0.06260902704190516 |

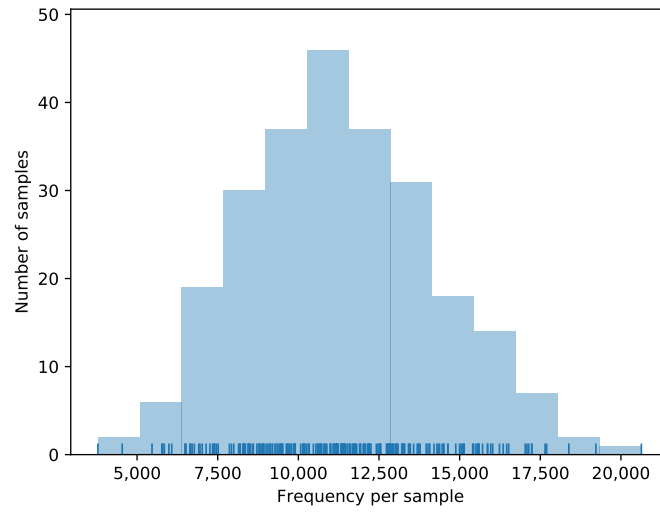


Figure 8: Frequency per Sample by DADA2

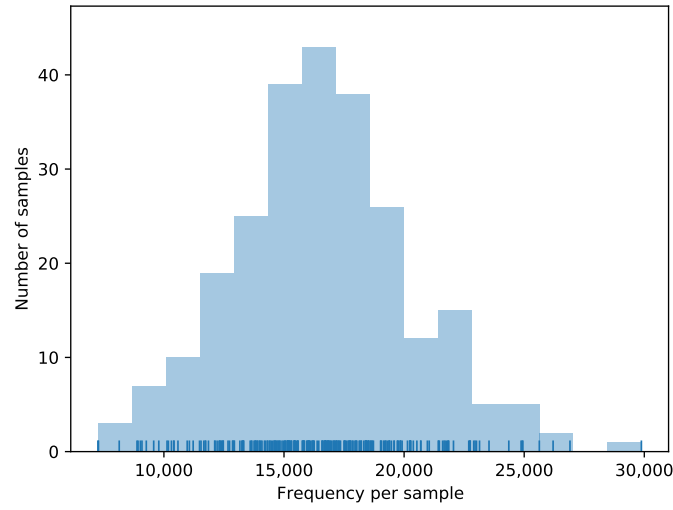


Figure 9: Frequency per Sample by DADA2

Table 2: Kruskal-Wallis Tests from Evenness Index with DADA2

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|----------------------|----------------------|---------------------|
| Early (n=50) | Healthy (n=100) | 0.003576158940404639 | 0.9523141335184352 | 0.9523141335184352 |
| Early (n=50) | Moderate (n=50) | 5.112902970297 | 0.02374855135702787 | 0.03562282703554181 |
| Early (n=50) | Severe (n=50) | 5.206859405940577 | 0.022497939047433364 | 0.03562282703554181 |
| Healthy (n=100) | Moderate (n=50) | 6.591830463576116 | 0.01024477815032801 | 0.03073433445098403 |
| Healthy (n=100) | Severe (n=50) | 6.756619867549659 | 0.0093400517403089 | 0.03073433445098403 |
| Moderate (n=50) | Severe (n=50) | 0.01216633663364064 | 0.9121705706341857 | 0.9523141335184352 |

Table 3: Kruskal-Wallis Tests from Faith PD Index with DADA2

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|--------------------|-----------------------|-----------------------|
| Early (n=50) | Healthy (n=100) | 0.3434543046357703 | 0.557842085850555 | 0.557842085850555 |
| Early (n=50) | Moderate (n=50) | 7.833790099009889 | 0.005127846488653557 | 0.0076917697329803355 |
| Early (n=50) | Severe (n=50) | 19.832839603960394 | 8.451807369366e-06 | 2.5355422108098e-05 |
| Healthy (n=100) | Moderate (n=50) | 8.964254304635801 | 0.0027531304578610103 | 0.005506260915722021 |
| Healthy (n=100) | Severe (n=50) | 24.32056688741727 | 8.156352492752821e-07 | 4.893811495651693e-06 |
| Moderate (n=50) | Severe (n=50) | 5.461592079207946 | 0.019438927334967618 | 0.02332671280196114 |

Table 4: Kruskal-Wallis Tests from Observed Features Index with DADA2

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|---------------------|-----------------------|-----------------------|
| Early (n=50) | Healthy (n=100) | 9.559750209810552 | 0.001988901703187571 | 0.005966705109562713 |
| Early (n=50) | Moderate (n=50) | 0.01069480203811357 | 0.9176330712208788 | 0.9176330712208788 |
| Early (n=50) | Severe (n=50) | 1.8918489487993617 | 0.1689935259025544 | 0.20279223108306527 |
| Healthy (n=100) | Moderate (n=50) | 16.280824652808626 | 5.461383546704547e-05 | 0.0003276830128022728 |
| Healthy (n=100) | Severe (n=50) | 6.9139163882453465 | 0.008552745576573654 | 0.017105491153147308 |
| Moderate (n=50) | Severe (n=50) | 2.1161415616917054 | 0.145753334857958 | 0.20279223108306527 |

Table 5: Kruskal-Wallis Tests from Shannon's Diversity Index with DADA2

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|---------------------|----------------------|---------------------|
| Early (n=50) | Healthy (n=100) | 5.291586754966886 | 0.021428686619934936 | 0.11394854365524665 |
| Early (n=50) | Moderate (n=50) | 1.3095920792079028 | 0.2524685249140654 | 0.3029622298968785 |
| Early (n=50) | Severe (n=50) | 4.305790099009869 | 0.037982847885082216 | 0.11394854365524665 |
| Healthy (n=100) | Moderate (n=50) | 2.223194701986756 | 0.13595148461788642 | 0.27190296923577284 |
| Healthy (n=100) | Severe (n=50) | 0.06109668874171348 | 0.8047709009969876 | 0.8047709009969876 |
| Moderate (n=50) | Severe (n=50) | 1.3573544554455452 | 0.2439965042398798 | 0.3029622298968785 |

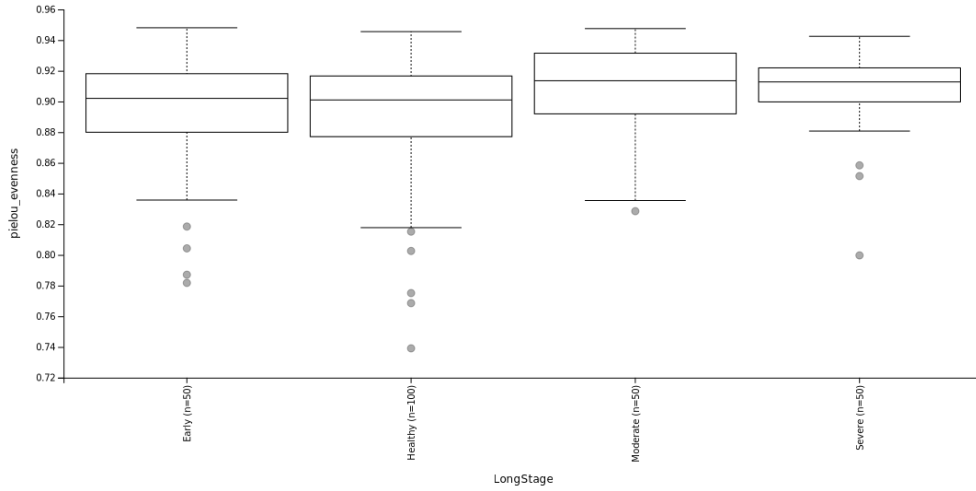


Figure 10: Evenness Index from DADA2

Table 6: Kruskal-Wallis Tests among All Group with Deblur

| Alpha-Diversity | H | p-value |
|--------------------|--------------------|-------------------------|
| Evenness | 9.242885737051779 | 0.026229960554059864 |
| Faith PD | 87.83605864541846 | 6.386769940789011e-19 |
| Observed Features | 59.59138364929631 | 7.186872791755095e-13 |
| Shnnon's Diversity | 24.823351075697246 | 0.000016810908296023026 |

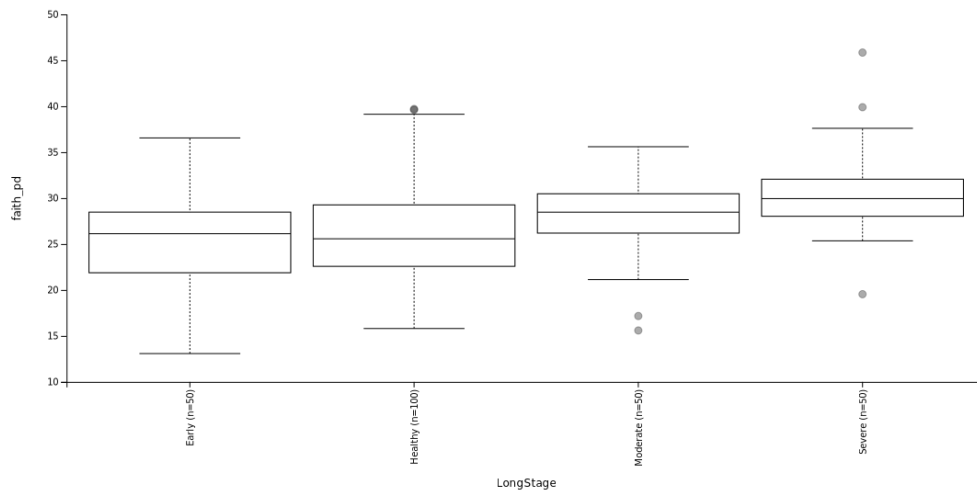


Figure 11: Faith PD Index from DADA2

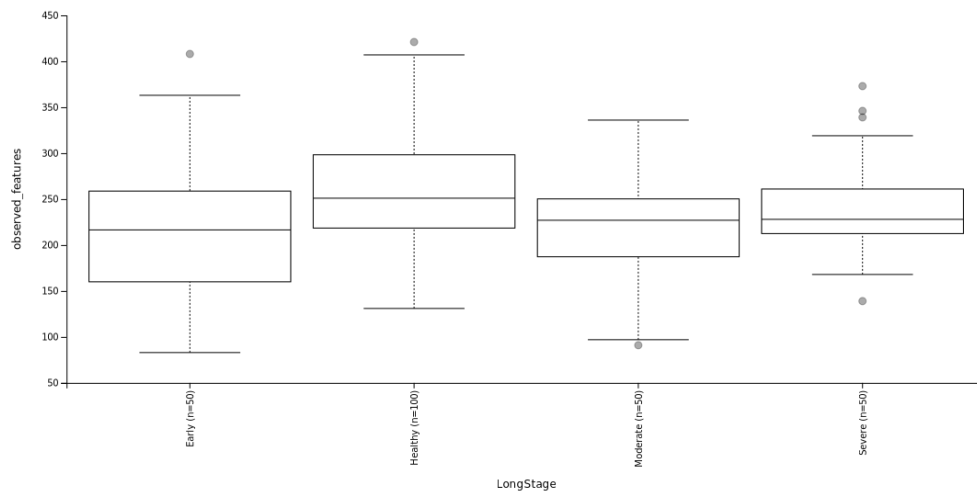


Figure 12: Observed Features Index from DADA2

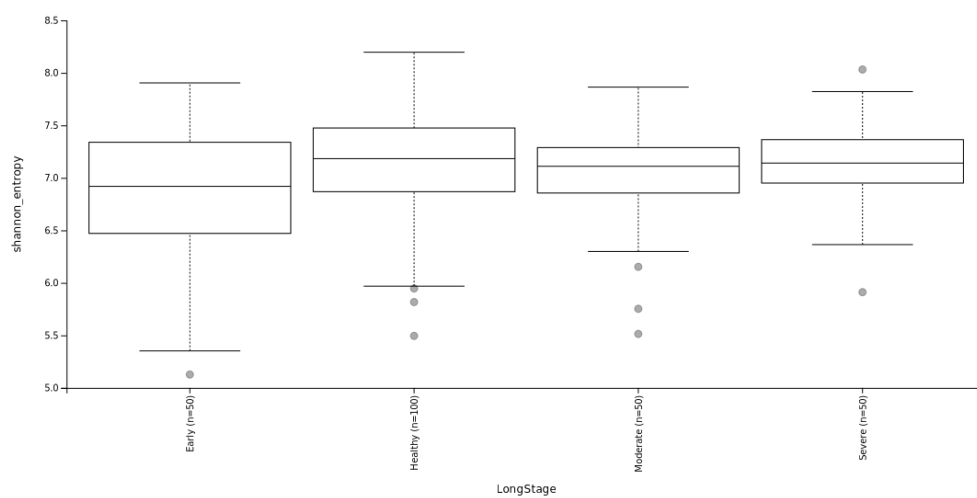


Figure 13: Shannon's Diversity Index from DADA2

Table 7: Kruskal-Wallis Tests from Evenness Index with Deblur

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|--------------------|----------------------|----------------------|
| Early (n=50) | Healthy (n=100) | 2.884386754966897 | 0.0894420544121846 | 0.15829564582637523 |
| Early (n=50) | Moderate (n=50) | 4.392047524752456 | 0.03610692636685824 | 0.10832077910057474 |
| Early (n=50) | Severe (n=50) | 8.828245544554477 | 0.002966034055389358 | 0.017796204332336148 |
| Healthy (n=100) | Moderate (n=50) | 0.6168317880794802 | 0.43222705558822094 | 0.43597874518665736 |
| Healthy (n=100) | Severe (n=50) | 2.6199099337748066 | 0.1055304305509168 | 0.15829564582637523 |
| Moderate (n=50) | Severe (n=50) | 0.6068435643564385 | 0.43597874518665736 | 0.43597874518665736 |

Table 8: Kruskal-Wallis Tests from Faith PD Index with Deblur

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|----------------------|-----------------------|------------------------|
| Early (n=50) | Healthy (n=100) | 2.7110304635762077 | 0.09965659889456922 | 0.11958791867348306 |
| Early (n=50) | Moderate (n=50) | 26.80400792079206 | 2.251698564500841e-07 | 3.3775478467512613e-07 |
| Early (n=50) | Severe (n=50) | 29.06252673267329 | 7.007948881210323e-08 | 1.4015897762420645e-07 |
| Healthy (n=100) | Moderate (n=50) | 51.153949668874134 | 8.539868055189094e-13 | 2.5619604165567283e-12 |
| Healthy (n=100) | Severe (n=50) | 54.86883178807949 | 1.288482355374052e-13 | 7.730894132244311e-13 |
| Moderate (n=50) | Severe (n=50) | 0.005750495049483106 | 0.9395527422741722 | 0.9395527422741722 |

Table 9: Kruskal-Wallis Tests from Observed Features Index with Deblur

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|----------------------|------------------------|------------------------|
| Early (n=50) | Healthy (n=100) | 0.4675226919952207 | 0.49412905906624816 | 0.5929548708794977 |
| Early (n=50) | Moderate (n=50) | 18.684815977243918 | 1.542055834477253e-05 | 2.31308375171588e-05 |
| Early (n=50) | Severe (n=50) | 20.703272962949605 | 5.362426456004328e-06 | 1.0724852912008657e-05 |
| Healthy (n=100) | Moderate (n=50) | 35.26606516292951 | 2.875998708064018e-09 | 8.627996124192055e-09 |
| Healthy (n=100) | Severe (n=50) | 37.015293460828644 | 1.1720632904898772e-09 | 7.032379742939263e-09 |
| Moderate (n=50) | Severe (n=50) | 0.003849966992737873 | 0.9505245257136643 | 0.9505245257136643 |

Table 10: Kruskal-Wallis Tests from Shannon's Diversity Index with Deblur

| Group 1 | Group 2 | H | p-value | q-value |
|-----------------|-----------------|---------------------|------------------------|-----------------------|
| Early (n=50) | Healthy (n=100) | 0.38679735099333357 | 0.5339876723058008 | 0.6407852067669609 |
| Early (n=50) | Moderate (n=50) | 10.767968316831627 | 0.0010327180791227218 | 0.0020654361582454436 |
| Early (n=50) | Severe (n=50) | 14.428562376237608 | 0.00014557751137778065 | 0.000627545643904027 |
| Healthy (n=100) | Moderate (n=50) | 10.172185430463571 | 0.0014257517732722547 | 0.002138627659908382 |
| Healthy (n=100) | Severe (n=50) | 13.746754966887409 | 0.0002091818813013423 | 0.000627545643904027 |
| Moderate (n=50) | Severe (n=50) | 0.15987326732670226 | 0.6892732232396639 | 0.6892732232396639 |

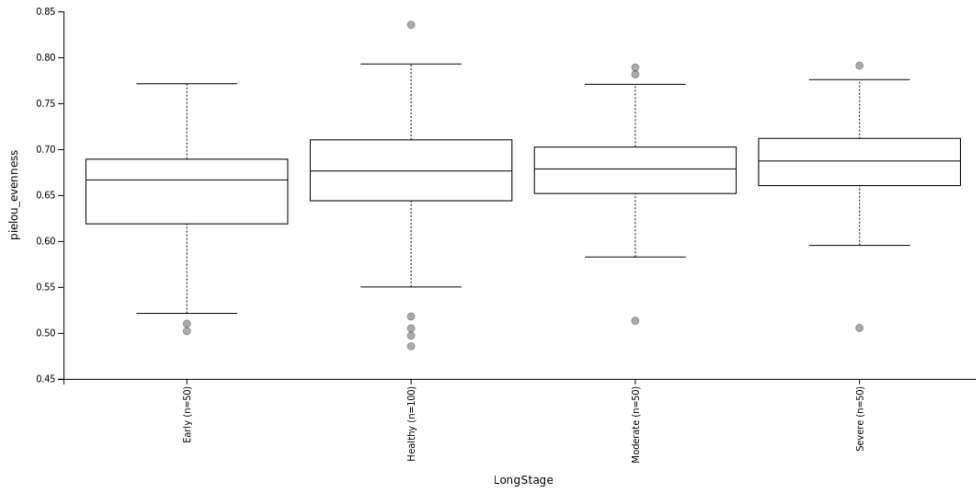


Figure 14: Evenness Index from Deblur

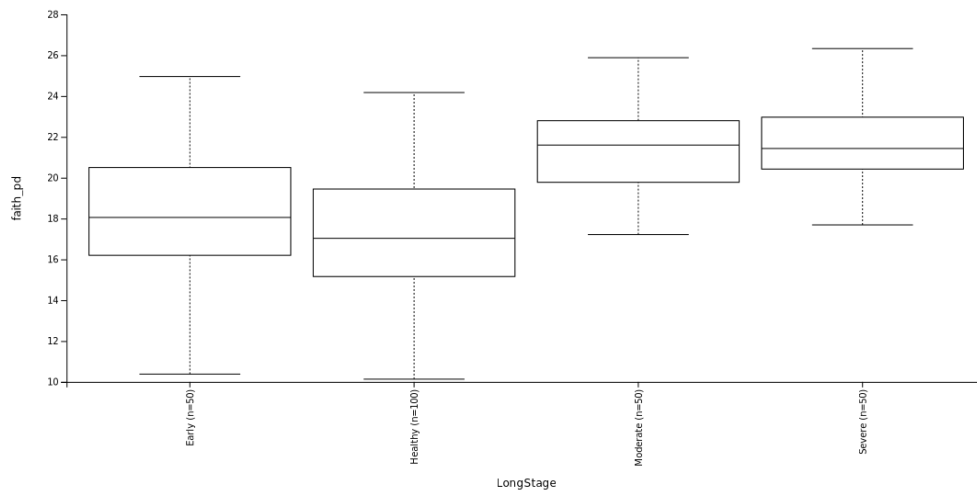


Figure 15: Faith PD Index from Deblur

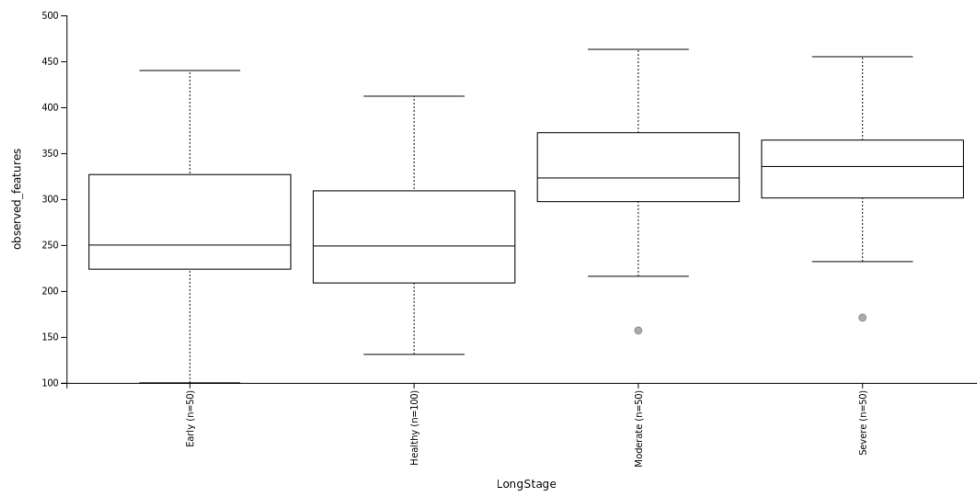


Figure 16: Observed Features Index from Deblur

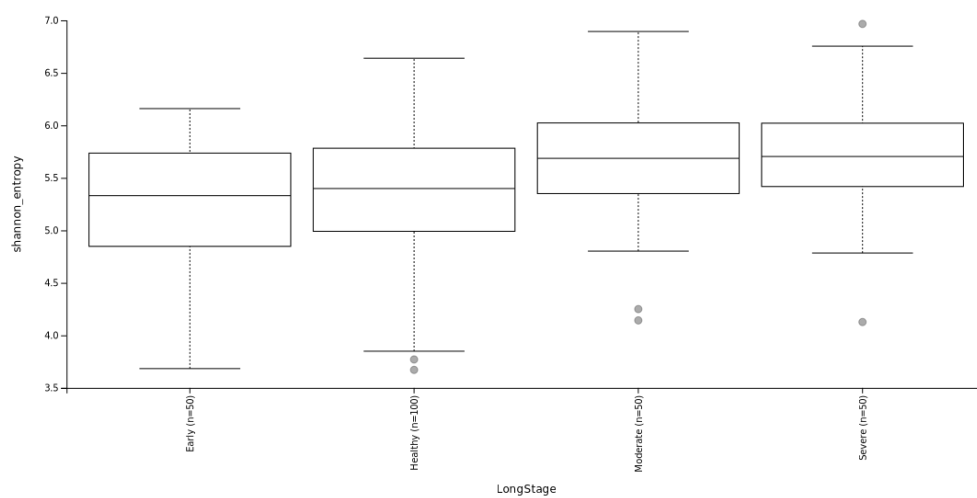


Figure 17: Shannon's Diversity Index from Deblur

Table 11: Bray-Curtis Distance Index with DADA2

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|---------|
| Early | Healthy | 150 | 999 | 1.8288671026193992 | 0.004 | 0.0048 |
| Early | Moderate | 100 | 999 | 2.4738348324475568 | 0.001 | 0.0015 |
| Early | Severe | 100 | 999 | 3.3691960533567005 | 0.001 | 0.0015 |
| Healthy | Moderate | 150 | 999 | 5.602936565444328 | 0.001 | 0.0015 |
| Healthy | Severe | 150 | 999 | 6.325447306476738 | 0.001 | 0.0015 |
| Moderate | Severe | 100 | 999 | 1.1018815494184453 | 0.219 | 0.219 |

Table 12: Jaccard Distance Index with DADA2

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|---------|
| Early | Healthy | 150 | 999 | 1.5875955458962276 | 0.001 | 0.0012 |
| Early | Moderate | 100 | 999 | 1.7486415070626309 | 0.001 | 0.0012 |
| Early | Severe | 100 | 999 | 1.8371794988000507 | 0.001 | 0.0012 |
| Healthy | Moderate | 150 | 999 | 3.9547515710373635 | 0.001 | 0.0012 |
| Healthy | Severe | 150 | 999 | 3.8380356039546784 | 0.001 | 0.0012 |
| Moderate | Severe | 100 | 999 | 0.9700395015774723 | 0.62 | 0.62 |

Table 13: Unweighted UniFrac Distance Index with DADA2

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|---------|
| Early | Healthy | 150 | 999 | 2.414078271406213 | 0.002 | 0.0024 |
| Early | Moderate | 100 | 999 | 4.941256726696032 | 0.001 | 0.0015 |
| Early | Severe | 100 | 999 | 6.184322196061149 | 0.001 | 0.0015 |
| Healthy | Moderate | 150 | 999 | 12.484494695636283 | 0.001 | 0.0015 |
| Healthy | Severe | 150 | 999 | 13.432593034368626 | 0.001 | 0.0015 |
| Moderate | Severe | 100 | 999 | 1.2428267228930112 | 0.084 | 0.084 |

Table 14: Weighted UniFrac Distance Index with DADA2

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|----------------------|
| Early | Healthy | 150 | 999 | 2.6584441800971716 | 0.019 | 0.022799999999999997 |
| Early | Moderate | 100 | 999 | 8.702906307484113 | 0.001 | 0.0015 |
| Early | Severe | 100 | 999 | 14.068214366598513 | 0.001 | 0.0015 |
| Healthy | Moderate | 150 | 999 | 22.059259782524673 | 0.001 | 0.0015 |
| Healthy | Severe | 150 | 999 | 31.310013450629775 | 0.001 | 0.0015 |
| Moderate | Severe | 100 | 999 | 1.7543213081828324 | 0.115 | 0.115 |

Table 15: Bray-Curtis Distance Index with Deblur

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|----------------------|
| Early | Healthy | 150 | 999 | 1.7634974220433302 | 0.019 | 0.022799999999999997 |
| Early | Moderate | 100 | 999 | 3.203442604434298 | 0.001 | 0.0015 |
| Early | Severe | 100 | 999 | 4.192790849454974 | 0.001 | 0.0015 |
| Healthy | Moderate | 150 | 999 | 6.953487468508356 | 0.001 | 0.0015 |
| Healthy | Severe | 150 | 999 | 7.5433379986347155 | 0.001 | 0.0015 |
| Moderate | Severe | 100 | 999 | 1.0959020597220626 | 0.313 | 0.313 |

Table 16: Jaccard Distance Index with Deblur

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|---------|
| Early | Healthy | 150 | 999 | 1.3701039884255466 | 0.001 | 0.0012 |
| Early | Moderate | 100 | 999 | 2.198029993855521 | 0.001 | 0.0012 |
| Early | Severe | 100 | 999 | 2.237738583770674 | 0.001 | 0.0012 |
| Healthy | Moderate | 150 | 999 | 4.528432929980079 | 0.001 | 0.0012 |
| Healthy | Severe | 150 | 999 | 4.374635292015638 | 0.001 | 0.0012 |
| Moderate | Severe | 100 | 999 | 1.0036296853126103 | 0.429 | 0.429 |

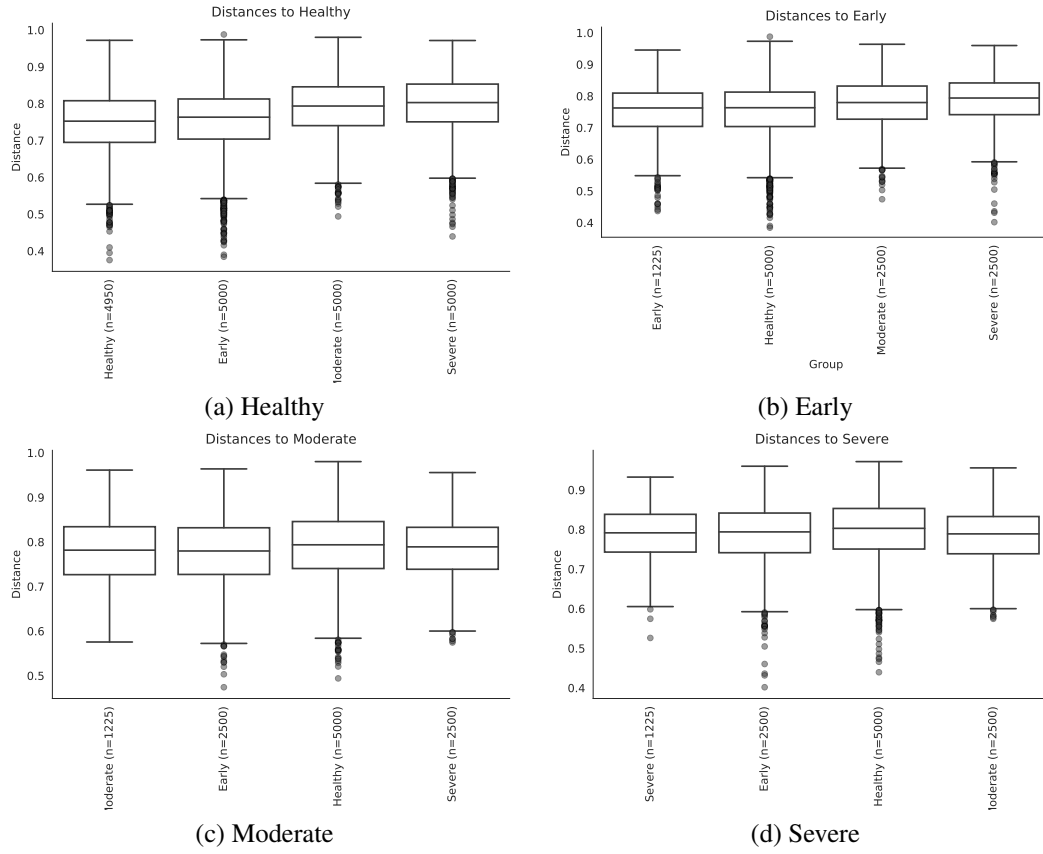


Figure 18: Bray-Curtis Distance Index with DADA2

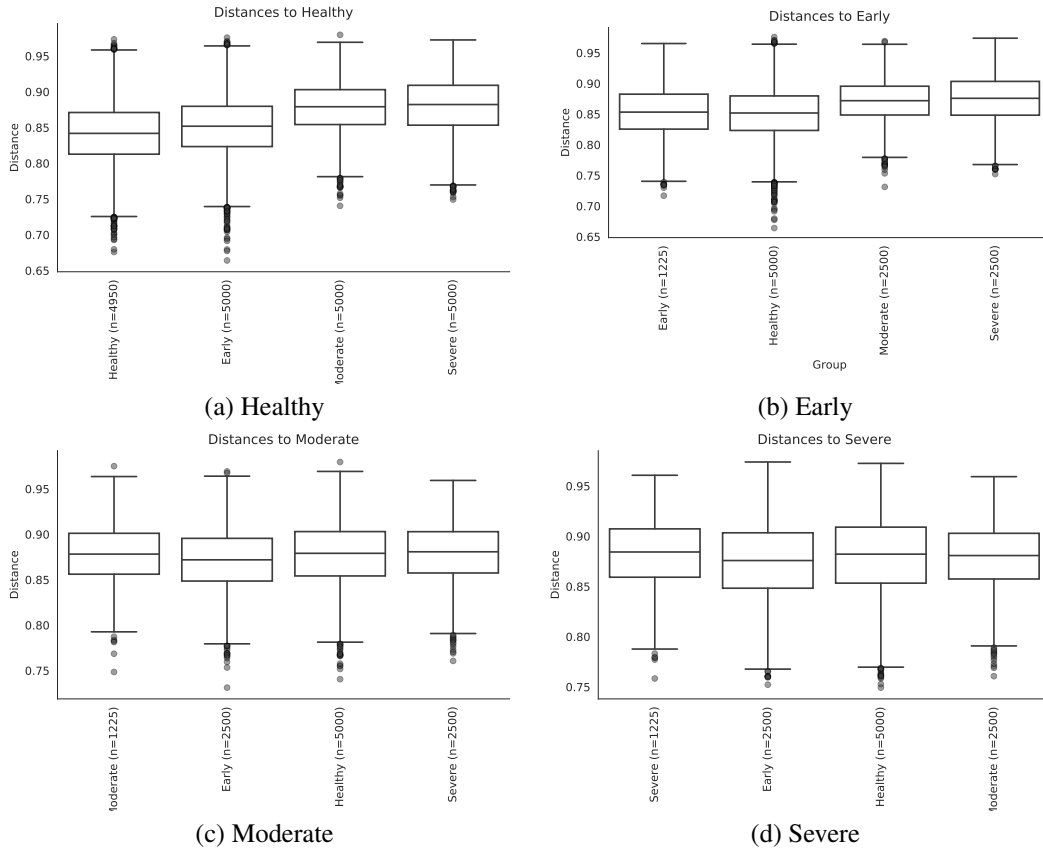


Figure 19: Jaccard Distance Index with DADA2



Figure 20: Unweighted Unifrac Distance Index with DADA2

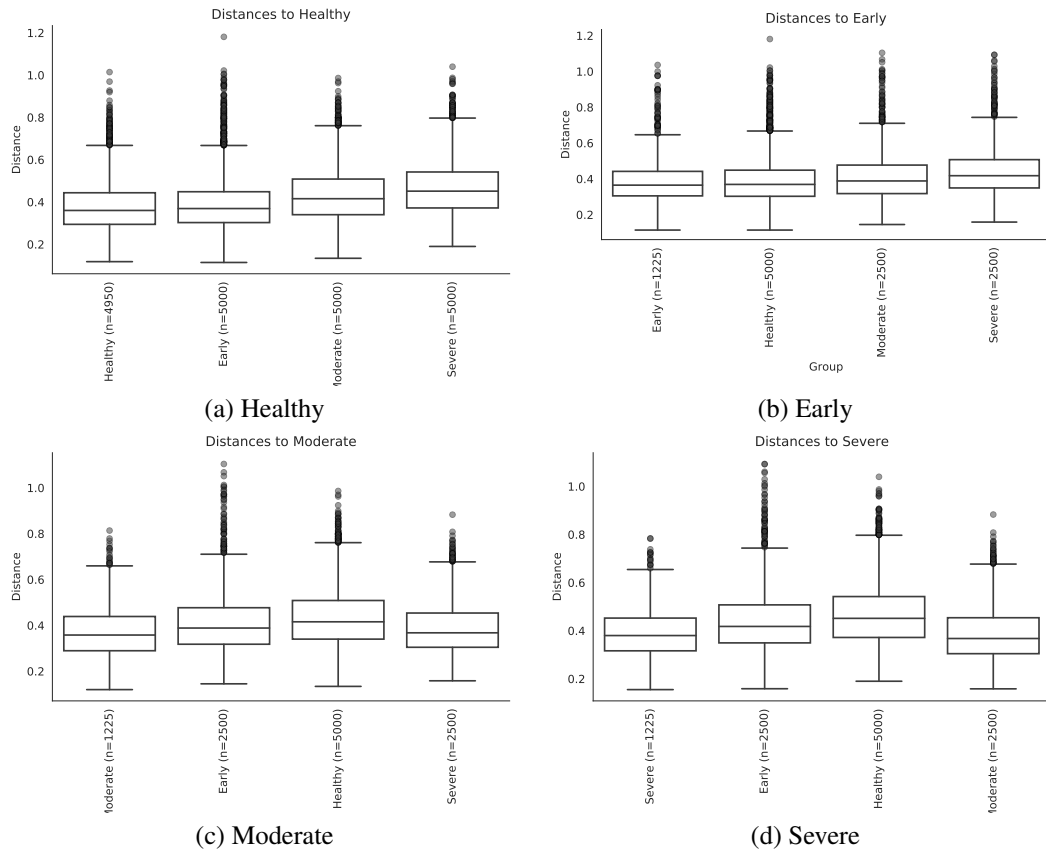


Figure 21: Weighted Unifrac Distance Index with DADA2

Table 17: Unweighted UniFrac Distance Index with Deblur

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|---------|
| Early | Healthy | 150 | 999 | 2.709074154153053 | 0.003 | 0.0036 |
| Early | Moderate | 100 | 999 | 7.547240014264336 | 0.001 | 0.0015 |
| Early | Severe | 100 | 999 | 7.772239667697252 | 0.001 | 0.0015 |
| Healthy | Moderate | 150 | 999 | 19.48285778321118 | 0.001 | 0.0015 |
| Healthy | Severe | 150 | 999 | 20.254907535032658 | 0.001 | 0.0015 |
| Moderate | Severe | 100 | 999 | 1.061788954262309 | 0.34 | 0.34 |

Table 18: Weighted UniFrac Distance Index with Deblur

| Group 1 | Group 2 | Sample size | Permutations | pseudo-F | p-value | q-value |
|----------|----------|-------------|--------------|--------------------|---------|---------|
| Early | Healthy | 150 | 999 | 2.0087857905677193 | 0.088 | 0.088 |
| Early | Moderate | 100 | 999 | 5.981646579135783 | 0.002 | 0.003 |
| Early | Severe | 100 | 999 | 16.572566883582837 | 0.001 | 0.002 |
| Healthy | Moderate | 150 | 999 | 9.494764618252377 | 0.001 | 0.002 |
| Healthy | Severe | 150 | 999 | 20.338834647304648 | 0.001 | 0.002 |
| Moderate | Severe | 100 | 999 | 5.026218407543304 | 0.003 | 0.0036 |

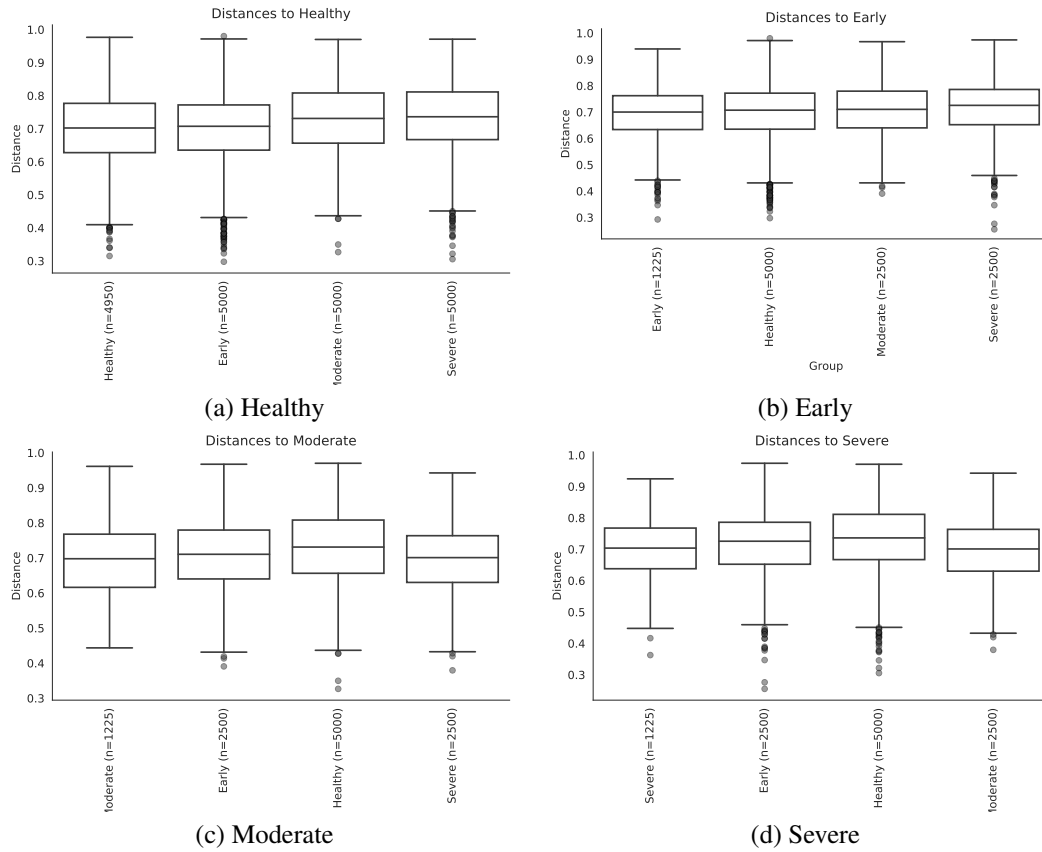


Figure 22: Bray-Curtis Distance Index with Deblur

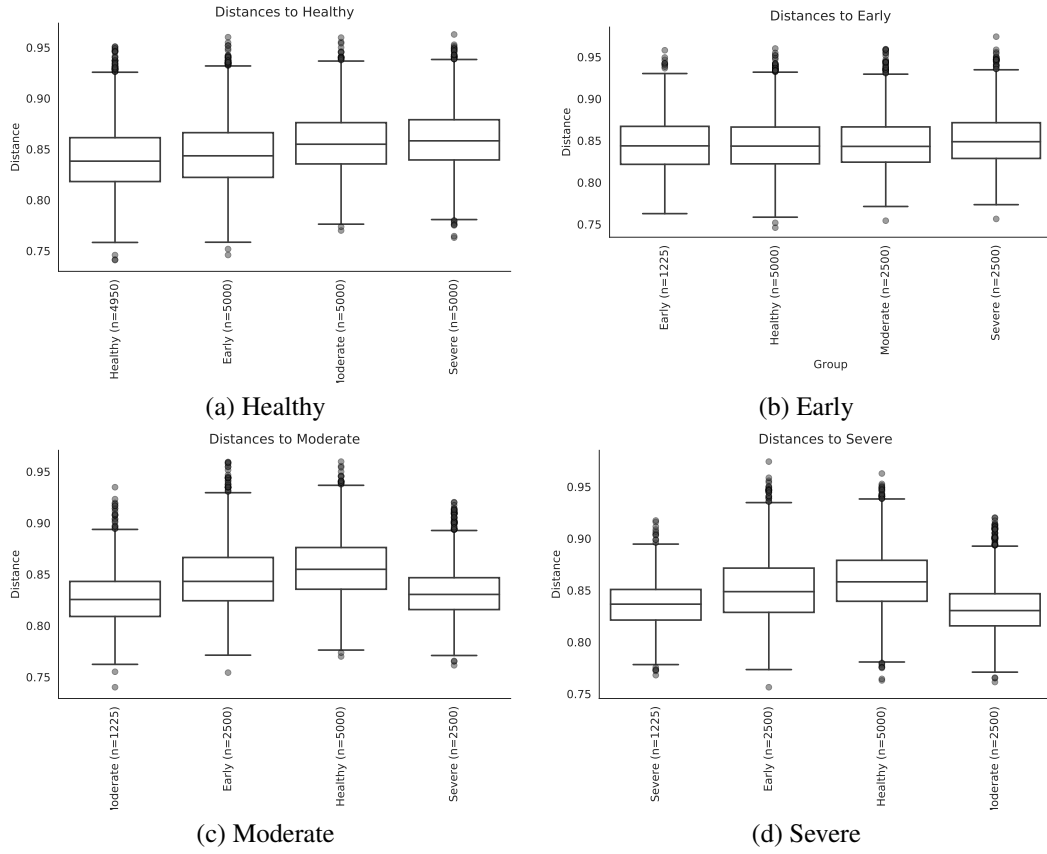


Figure 23: Jaccard Distance Index with Deblur

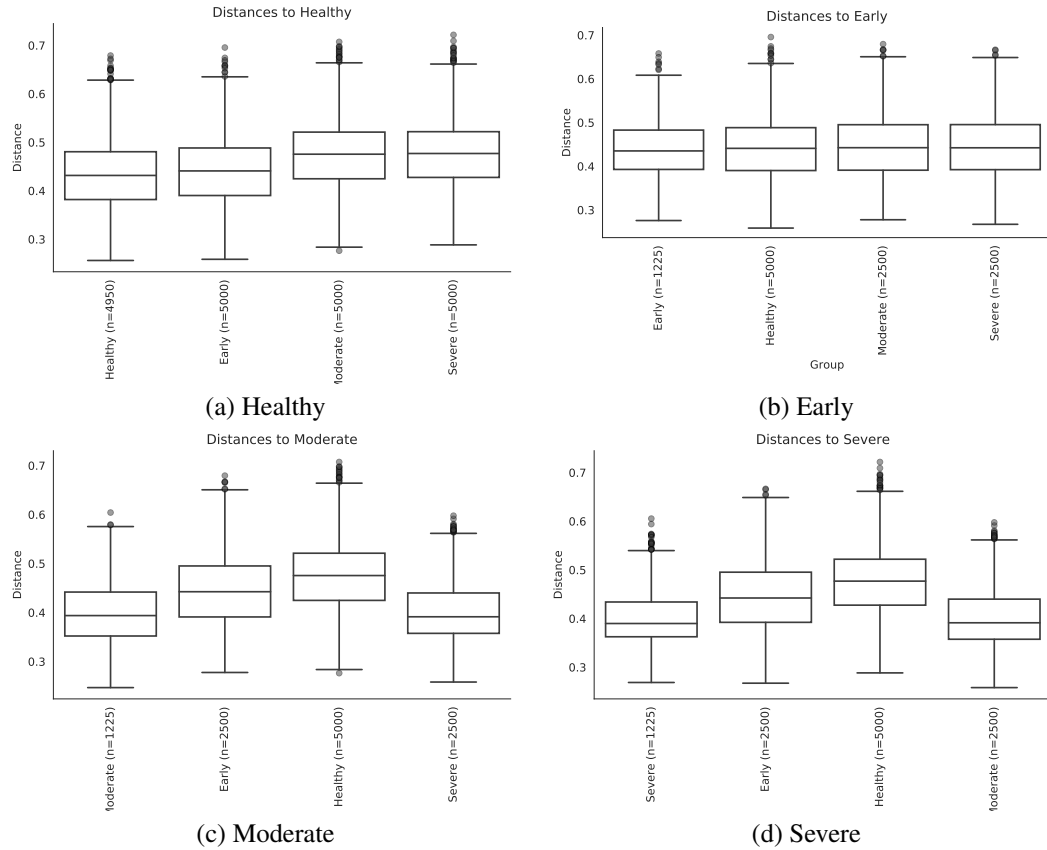


Figure 24: Unweighted Unifrac Distance Index with Deblur

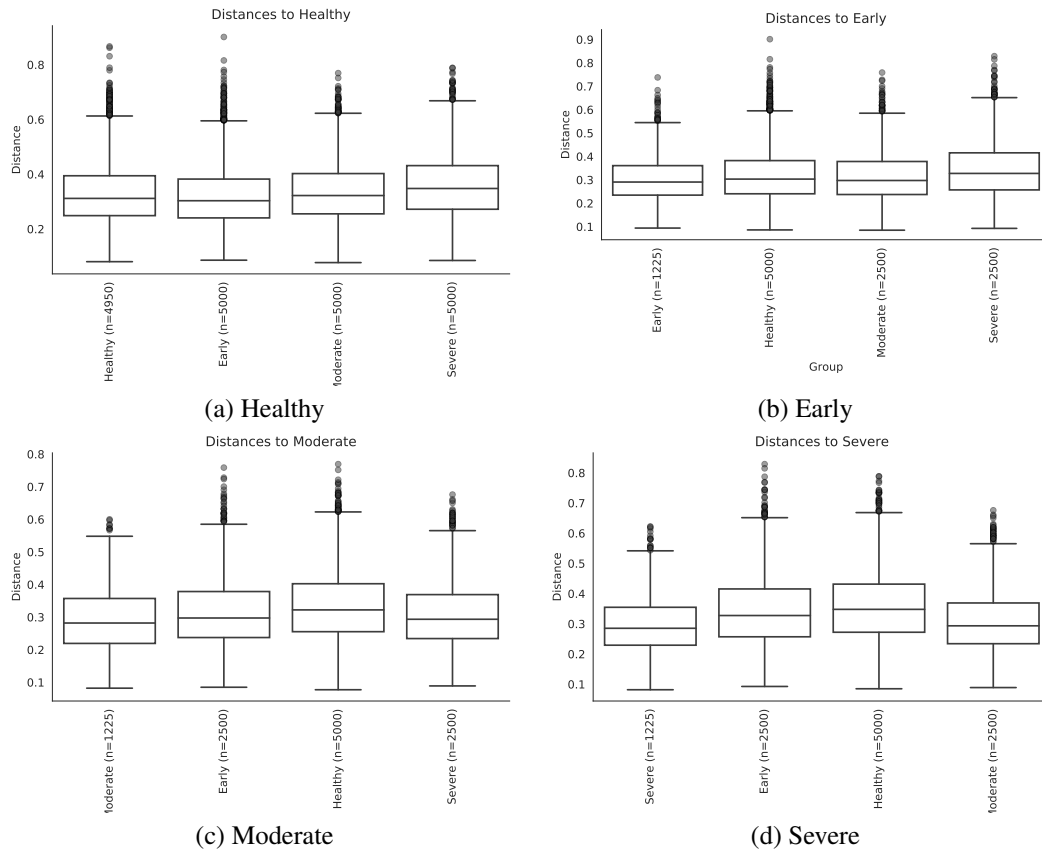


Figure 25: Weighted Unifrac Distance Index with Deblur

Table 19: ANCOM Significant Taxa with DADA2 and Greengenes

| | W | Reject null hypothesis |
|---|-----|------------------------|
| Bacteria Actinobacteria Actinobacteria Actinomycetales Actinomycetaceae Actinomycetes | 326 | True |
| Bacteria Firmicutes Clostridia Clostridiales Peptostreptococcaceae Filifactor | 325 | True |
| Bacteria Spirochaetes Spirochaetes Spirochaetales Spirochaetaceae Treponema | 325 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Prevotellaceae Prevotella intermedia | 323 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas endodontalis | 321 | True |
| Bacteria Spirochaetes Spirochaetes Spirochaetales Spirochaetaceae Treponema amylovorum | 320 | True |
| Bacteria Synergistetes Synergistia Synergistales Dethiosulfovibrionaceae TG5 | 319 | True |
| Bacteria Tenericutes Mollicutes Mycoplasmatales Mycoplasmataceae Mycoplasma | 318 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae Tannerella | 315 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas | 313 | True |
| Bacteria Actinobacteria Actinobacteria Actinomycetales Corynebacteriaceae Corynebacterium durum | 309 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales | 306 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Mogibacteriaceae] | 305 | True |
| Bacteria Proteobacteria Epsilonproteobacteria Campylobacteriales Campylobacteraceae Campylobacter | 305 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Acidaminobacteraceae] | 304 | True |

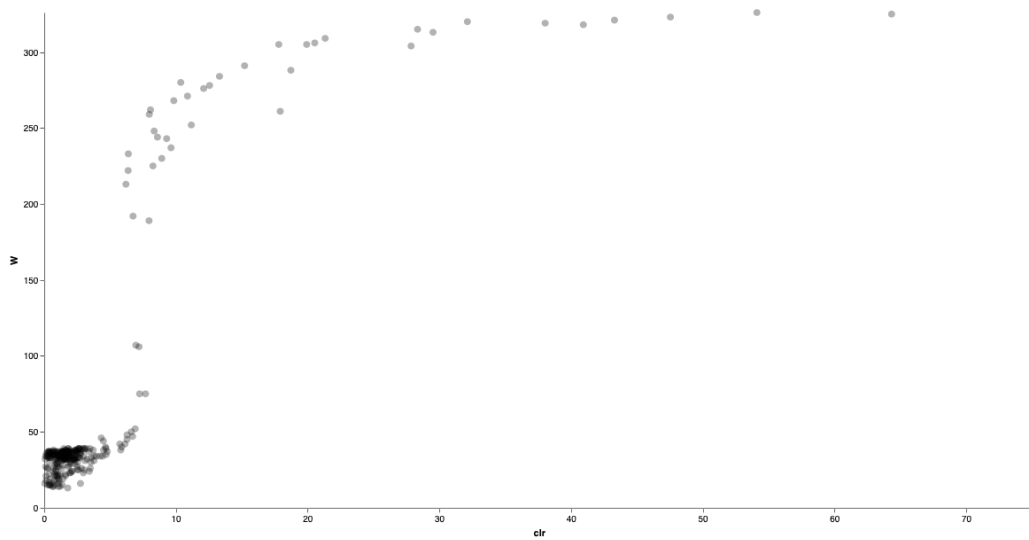


Figure 26: ANCOM Volcano Plot with DADA2 and Greengenes

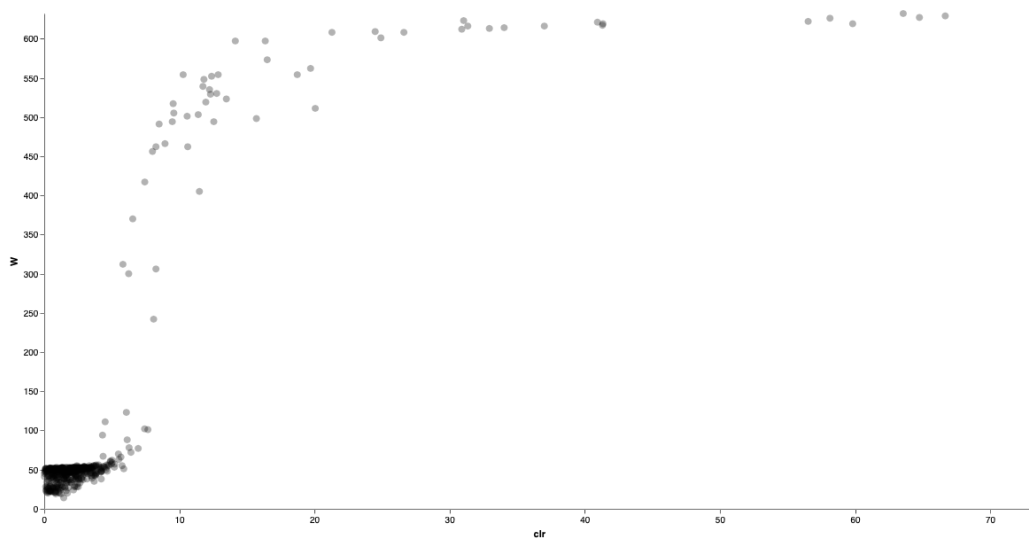


Figure 27: ANCOM Volcano Plot with DADA2 and SILVA

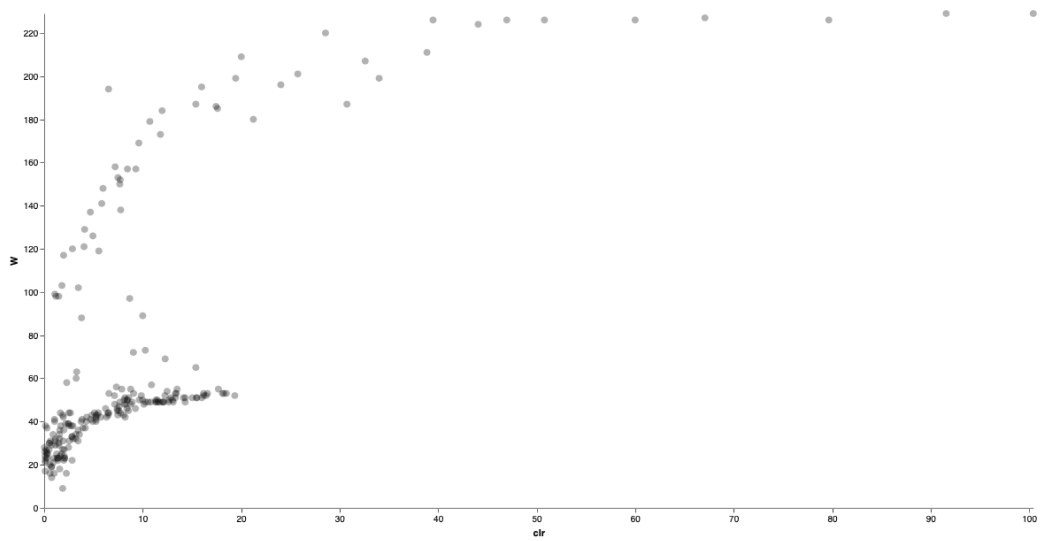


Figure 28: ANCOM Volcano Plot with Deblur and Greengenes

Table 20: ANCOM Significant Taxa with DADA2 and SILVA

| | W | Reject null hypothesis |
|---|-----|------------------------|
| Bacteria Actinobacteriota Actinobacteria Actinomycetales Actinomycetaceae Actinomyces | 632 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas Porphyromonas'gingivalis | 629 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Peptostreptococcaceae Filifactor Filifactor'alocis | 627 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Prevotellaceae Prevotella Prevotella'intermedia | 626 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema Treponema'denticola | 626 | True |
| Bacteria Actinobacteriota Actinobacteria Actinomycetales Actinomycetaceae Actinomyces Schaalia'odontolytica | 623 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Tannerellaceae Tannerella Tannerella'forsythia | 622 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema Treponema'medium | 621 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Anaerovoracaceae [Eubacterium]'nodatum'group [Eubacterium]'nodatum | 619 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema uncultured'bacterium | 619 | True |
| Bacteria Firmicutes Bacilli Mycoplasmatales Mycoplasmataceae Mycoplasma Metamycoplasma'faucium | 617 | True |
| Bacteria Synergistota Synergistia Synergistales Synergistaceae Fretibacterium | 616 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema | 616 | True |
| Bacteria Firmicutes Clostridia Lachnospirales Defluviitaleaceae Defluviitaleaceae'UCG-011 Lachnospiraceae'bacterium | 614 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas | 613 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Anaerovoracaceae [Eubacterium]'brachy'group [Eubacterium]'brachy | 612 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Prevotellaceae | 609 | True |
| Bacteria Actinobacteriota Actinobacteria Corynebacteriales Corynebacteriaceae Corynebacterium Corynebacterium'durum | 608 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Anaerovoracaceae [Eubacterium]'saphenum'group Eubacterium'saphenum | 608 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema Treponema'maltophilum | 601 | True |
| Bacteria Campilobacterota Campylobacteria Campylobacteriales Campylobacteraceae Campylobacter Campylobacter'showae | 597 | True |
| Bacteria Actinobacteriota Actinobacteria Actinomycetales Actinomycetaceae Actinomyces Actinomyces'graevenitzi | 597 | True |
| Bacteria Firmicutes Clostridia Lachnospirales Lachnospiraceae Oribacterium | 573 | True |

Table 21: ANCOM Significant Taxa with Deblur and Greengenes

| | W | Reject null hypothesis |
|---|-----|------------------------|
| Bacteria Firmicutes Clostridia Clostridiales Peptostreptococcaceae Filifactor | 229 | True |
| Bacteria Spirochaetes Spirochaetes Spirochaetales Spirochaetaceae Treponema | 229 | True |
| Bacteria Spirochaetes Spirochaetes Spirochaetales Spirochaetaceae Treponema amylovorum | 227 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Prevotellaceae Prevotella intermedia | 226 | True |
| Bacteria Tenericutes Mollicutes Mycoplasmatales Mycoplasmataceae Mycoplasma | 226 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas endodontalis | 226 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae Tannerella | 226 | True |
| Bacteria Synergistetes Synergistia Synergistales Dethiosulfovibrionaceae TG5 | 226 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas | 224 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales | 220 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Mogibacteriaceae] | 211 | True |
| Bacteria Firmicutes Clostridia Clostridiales Peptostreptococcaceae Peptostreptococcus | 209 | True |
| Bacteria Proteobacteria Deltaproteobacteria Desulfobacterales Desulfobulbaceae Desulfobulbus | 207 | True |
| Bacteria Spirochaetes Spirochaetes Spirochaetales Spirochaetaceae Treponema socranskii | 201 | True |
| Bacteria Proteobacteria Epsilonproteobacteria Campylobacterales Campylobacteraceae Campylobacter | 199 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Acidaminobacteraceae] | 199 | True |
| Bacteria Proteobacteria Gammaproteobacteria Pasteurellales Pasteurellaceae Haemophilus parainfluenzae | 196 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Tissierellaceae] Parvimonas | 195 | True |
| Bacteria Proteobacteria Betaproteobacteria Neisseriales Neisseriaceae Neisseria subflava | 194 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Mogibacteriaceae] Mogibacterium | 187 | True |
| Bacteria Actinobacteria Actinobacteria Actinomycetales Actinomycetaceae Actinomyces | 187 | True |
| Bacteria Firmicutes Clostridia Clostridiales [Tissierellaceae] | 186 | True |
| Bacteria Actinobacteria Actinobacteria Actinomycetales | 185 | True |
| Bacteria Firmicutes Clostridia Clostridiales | 184 | True |
| Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae Oribacterium | 180 | True |
| Bacteria Firmicutes Clostridia Clostridiales Lachnospiraceae | 179 | True |
| Bacteria Bacteroidetes Bacteroidia Bacteroidales Prevotellaceae Prevotella nanceiensis | 173 | True |

Table 22: ANCOM Significant Taxa with DADA2 and SILVA

| | W | Reject null hypothesis |
|---|-----|------------------------|
| Bacteria Actinobacteriota Actinobacteria Actinomycetales Actinomycetaceae Actinomyces | 632 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas Porphyromonas'gingivalis | 629 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Peptostreptococcaceae Filifactor Filifactor'alocis | 627 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Prevotellaceae Prevotella Prevotella'intermedia | 626 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema Treponema'denticola | 626 | True |
| Bacteria Actinobacteriota Actinobacteria Actinomycetales Actinomycetaceae Actinomyces Schaalia'odontolytica | 623 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Tannerellaceae Tannerella Tannerella'forsythia | 622 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema Treponema'medium | 621 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Anaerovoracaceae [Eubacterium]'nodatum'group [Eubacterium]'nodatum | 619 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema uncultured'bacterium | 619 | True |
| Bacteria Firmicutes Bacilli Mycoplasmatales Mycoplasmataceae Mycoplasma Metamycoplasma'faucium | 617 | True |
| Bacteria Synergistota Synergistia Synergistales Synergistaceae Fretibacterium | 616 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema | 616 | True |
| Bacteria Firmicutes Clostridia Lachnospirales Defluviitaleaceae Defluviitaleaceae'UCG-011 Lachnospiraceae'bacterium | 614 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas | 613 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Anaerovoracaceae [Eubacterium]'brachy'group [Eubacterium]'brachy | 612 | True |
| Bacteria Bacteroidota Bacteroidia Bacteroidales Prevotellaceae | 609 | True |
| Bacteria Actinobacteriota Actinobacteria Corynebacteriales Corynebacteriaceae Corynebacterium Corynebacterium'durum | 608 | True |
| Bacteria Firmicutes Clostridia Peptostreptococcales-Tissierellales Anaerovoracaceae [Eubacterium]'saphenum'group Eubacterium'saphenum | 608 | True |
| Bacteria Spirochaetota Spirochaetia Spirochaetales Spirochaetaceae Treponema Treponema'maltophilum | 601 | True |
| Bacteria Campilobacterota Campylobacteria Campylobacteriales Campylobacteraceae Campylobacter Campylobacter'showae | 597 | True |
| Bacteria Actinobacteriota Actinobacteria Actinomycetales Actinomycetaceae Actinomyces Actinomyces'graevenitzi | 597 | True |
| Bacteria Firmicutes Clostridia Lachnospirales Lachnospiraceae Oribacterium | 573 | True |

- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., ... Andersen, G. L. (2006). Greengenes, a chimera-checked 16s rna gene database and workbench compatible with arb. *Applied and environmental microbiology*, 72(7), 5069–5072.
- Flemmig, T. F. (1999). Periodontitis. *Annals of Periodontology*, 4(1), 32–37.
- Gihawi, A., Rallapalli, G., Hurst, R., Cooper, C. S., Leggett, R. M., & Brewer, D. S. (2019). Sepath: benchmarking the search for pathogens in human tissue whole genome sequence data leads to template pipelines. *Genome biology*, 20(1), 1–15.
- Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., ... Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *science*, 312(5778), 1355–1359.
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3), 90–95.
- James, F. C., & Rathbun, S. (1981). Rarefaction, relative abundance, and diversity of avian communities. *The Auk*, 98(4), 785–800.
- Maaten, L. v. d., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov), 2579–2605.
- Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., & Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease*, 26(1), 27663.
- McKinney, W., et al. (2011). pandas: a foundational python library for data analysis and statistics. *Python for High Performance and Scientific Computing*, 14(9).
- Olsen, G. J., & Woese, C. R. (1993). Ribosomal rna: a key to phylogeny. *The FASEB journal*, 7(1), 113–123.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12, 2825–2830.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). Silva: a comprehensive online resource for quality checked and aligned ribosomal rna sequence data compatible with arb. *Nucleic acids research*, 35(21), 7188–7196.
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, 449(7164), 804–810.
- Van Dyke, T. E., & Dave, S. (2005). Risk factors for periodontitis. *Journal of the International Academy of Periodontology*, 7(1), 3.
- Waskom, M., & the seaborn development team. (2020, September). *mwaskom/seaborn*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.592845> doi: 10.5281/zenodo.592845
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., ... others (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1), 27.

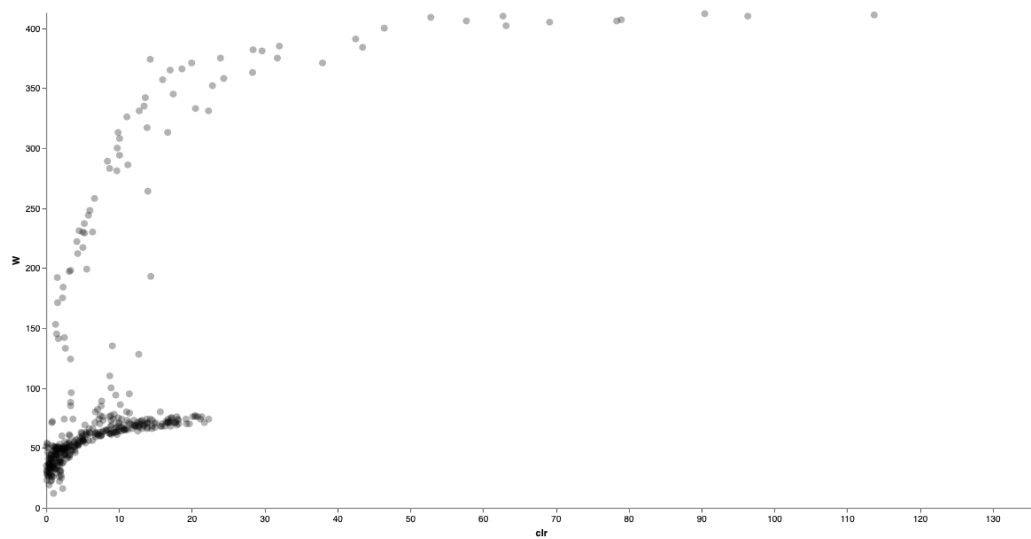


Figure 29: ANCOM Volcano Plot with Deblur and SILVA