**SUPPLEMENTAL MATERIAL**

**The effects of alcohol consumption and increased body mass on the gut microbiota of Parkinson’s Disease patients**

**Jared Dutra, Madeline Fung, Michelle Ling, Rui Lin Zhi**

**SUPPLEMENTAL TABLE CAPTIONS**

**Table S1. Alcohol consumption categories.** Criteria for each category was based on Canadian Centre on Substance Use and Addiction guidelines for daily alcohol consumption in adult men and women (56).

**Table S2.** **BMI categories for PD patients and healthy controls.** Criteria for overweight status was based on BMI score groupings (15).

**Table S3. Pairwise PERMANOVA results based on Weighted UniFrac beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

**Table S4. Pairwise PERMANOVA results based on Bray-Curtis beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

**Table S5. Pairwise PERMANOVA results based on Unweighted UniFrac beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

**Table S6. Pairwise PERMANOVA results based on Jaccard beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

**Table S7. *Coprococcus* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

**Table S8. *Blautia* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

**Table S9. *Roseburia* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

**Table S10. *Lachnospira* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

**Table S11. *Faecalibacterium* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

**Table S12. Pairwise PERMANOVA results based on the Weighted UniFrac beta diversity distances for comparing BMI and PD disease status.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

**Table S13. Two indicator families are associated with overweight PD individuals.** A indicates specificity (RA of a given taxa for each group) and B indicates fidelity (taxa present across the majority of samples from that specific group). Indicator value “stat” is calculated by multiplying A and B. p-values <0.05 = significant.

**SUPPLEMENTAL FIGURE CAPTIONS**

**Figure S1. Alcohol consumption categories do not have an effect on microbial community composition.** The gut microbiome composition of subjects was analyzed with (A) Jaccard, (B) Bray-Curtis, and (C) Unweighted UniFrac beta diversity metrics in QIIME2. Corresponding pairwise PERMANOVA statistics can be found in Tables S4-S6.

**Figure S2. No differences in the composition of gut microbiomes of different alcohol consumption groups were observed across any beta diversity metric.** Beta diversity of gut microbiotas across subjects in the three categories of alcohol consumption based on (A) Jaccard, (B) Bray-Curtis, and (C) Unweighted UniFrac. Corresponding pairwise PERMANOVA statistics can be found in Tables S4-S6.

**Figure S3. No correlation was found between average alcohol consumption (g/day) and UPDRS 1, 2, 3, and 4 scores.** (A)UPDRS 1 and (B) 2 scores are based on non-motor and motor challenges faced by subjects in their day-to-day lives, respectively. (C) UPDRS 3 scores represent the results of a movement analysis examination. (D) UPDRS 4 scores represent motor complications experienced by patients. Higher scores indicate more severe symptoms. Subjects with no alcohol consumption data were excluded from the analysis. Number of patients (n) = 195.

**Figure S4. Five families are differential abundant between overweight and normal PD patients.** The families Veillonellaceae, Lachnospiraceae, Desulfovibrionaceae, and Alcaligenaceae exhibited a positive fold change, indicating a higher abundance in overweight PD patients. The family Victivallaceae exhibited a negative fold change, indicating a higher abundance in normal PD patients.

**SUPPLEMENTAL TABLES**

**Table S1. Alcohol consumption categories.** Criteria for each category was based on Canadian Centre on Substance Use and Addiction guidelines for daily alcohol consumption in adult men and women (56).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Women |  |  | Men |  |  |
| Alcohol consumption | Range (g/day) | n | Alcohol consumption | Range (g/day) | n |
| None | 0 | 23 | **None** | 0 | 38 |
| Low | 0.1 - 13.6 | 90 | **Low** | 0.1 - 27.2 | 107 |
| Moderate | 13.7 - 27.1 | 5 | **Moderate** | 27.3 - 40.7 | 10 |
| High | >27.2 | 9 | **High** | >40.8 | 3 |

**Table S2.** **BMI categories for PD patients and healthy controls.** Criteria for overweight status was based on BMI score groupings (15).

|  |  |  |
| --- | --- | --- |
| Disease status | BMI categories  (normal = 18.5-25; overweight = 25-30) | Number of subjects (n) |
| **Control** | Normal | 41 |
| Overweight | 38 |
| **PD** | Normal | 71 |
| Overweight | 82 |

**Table S3. Pairwise PERMANOVA results based on Weighted UniFrac beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group 1 | Group 2 | Sample size | Permutation | pseudo-F | p-value | q-value |
| High | Low | 165 | 999 | 1.102104 | 0.306 | 0.787 |
| High | Moderate | 21 | 999 | 0.717307 | 0.633 | 0.787 |
| High | None | 54 | 999 | 1.120412 | 0.33 | 0.787 |
| Low | Moderate | 166 | 999 | 0.561177 | 0.746 | 0.787 |
| Low | None | 199 | 999 | 0.51953 | 0.787 | 0.787 |
| Moderate | None | 55 | 999 | 0.558168 | 0.668 | 0.787 |

**Table S4. Pairwise PERMANOVA results based on Bray-Curtis beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group 1 | Group 2 | Sample size | Permutation | pseudo-F | p-value | q-value |
| High | Low | 165 | 999 | 1.767261 | 0.907 | 0.907 |
| High | Moderate | 21 | 999 | 2.773597 | 0.643 | 0.8304 |
| High | None | 54 | 999 | 1.925926 | 0.524 | 0.8304 |
| Low | Moderate | 166 | 999 | 1.574987 | 0.692 | 0.8304 |
| Low | None | 199 | 999 | 0.824537 | 0.512 | 0.8304 |
| Moderate | None | 55 | 999 | 0.717412 | 0.483 | 0.8304 |

**Table S5. Pairwise PERMANOVA results based on Unweighted UniFrac beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group 1 | Group 2 | Sample size | Permutation | pseudo-F | p-value | q-value |
| High | Low | 165 | 999 | 0.913513 | 0.6 | 0.9585 |
| High | Moderate | 21 | 999 | 0.91976 | 0.639 | 0.9585 |
| High | None | 54 | 999 | 1.200759 | 0.172 | 0.909 |
| Low | Moderate | 166 | 999 | 0.614935 | 0.995 | 0.995 |
| Low | None | 199 | 999 | 1.062708 | 0.303 | 0.909 |
| Moderate | None | 55 | 999 | 0.742394 | 0.909 | 0.995 |

**Table S6. Pairwise PERMANOVA results based on Jaccard beta diversity analysis between alcohol consumption groups.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group 1 | Group 2 | Sample size | Permutation | pseudo-F | p-value | q-value |
| High | Low | 165 | 999 | 0.944019 | 0.68 | 0.971 |
| High | Moderate | 21 | 999 | 0.938257 | 0.736 | 0.971 |
| High | None | 54 | 999 | 1.024249 | 0.336 | 0.971 |
| Low | Moderate | 166 | 999 | 0.846103 | 0.971 | 0.971 |
| Low | None | 199 | 999 | 0.998596 | 0.453 | 0.971 |
| Moderate | None | 55 | 999 | 0.878127 | 0.924 | 0.971 |

**Table S7. *Coprococcus* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Disease | Variable | Group 1 | Group 2 | n1 | n2 | p | p.adj | p.adj.signif |
| Control | Abundance | High | Low | 7 | 71 | 0.01 | 0.12 | ns |
| Control | Abundance | High | Moderate | 7 | 4 | 0.066 | 0.792 | ns |
| Control | Abundance | High | None | 7 | 19 | 0.02 | 0.24 | ns |
| Control | Abundance | Low | Moderate | 71 | 4 | 0.96 | 1 | ns |
| Control | Abundance | Low | None | 71 | 19 | 0.663 | 1 | ns |
| Control | Abundance | Moderate | None | 4 | 19 | 0.858 | 1 | ns |
| PD | Abundance | High | Low | 13 | 112 | 0.139 | 1 | ns |
| PD | Abundance | High | Moderate | 13 | 16 | 0.526 | 1 | ns |
| PD | Abundance | High | None | 13 | 43 | 0.414 | 1 | ns |
| PD | Abundance | Low | Moderate | 112 | 16 | 0.519 | 1 | ns |
| PD | Abundance | Low | None | 112 | 43 | 0.622 | 1 | ns |
| PD | Abundance | Moderate | None | 16 | 43 | 0.88 | 1 | ns |

**Table S8. *Blautia* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Disease | Variable | Group 1 | Group 2 | n1 | n2 | p | p.adj | p.adj.signif |
| Control | Abundance | High | Low | 7 | 71 | 0.374 | 1 | ns |
| Control | Abundance | High | Moderate | 7 | 4 | 0.565 | 1 | ns |
| Control | Abundance | High | None | 7 | 19 | 0.336 | 1 | ns |
| Control | Abundance | Low | Moderate | 71 | 4 | 0.238 | 1 | ns |
| Control | Abundance | Low | None | 71 | 19 | 0.413 | 1 | ns |
| Control | Abundance | Moderate | None | 4 | 19 | 0.159 | 1 | ns |
| PD | Abundance | High | Low | 13 | 112 | 0.366 | 1 | ns |
| PD | Abundance | High | Moderate | 13 | 16 | 0.701 | 1 | ns |
| PD | Abundance | High | None | 13 | 43 | 0.262 | 1 | ns |
| PD | Abundance | Low | Moderate | 112 | 16 | 0.48 | 1 | ns |
| PD | Abundance | Low | None | 112 | 43 | 0.578 | 1 | ns |
| PD | Abundance | Moderate | None | 16 | 43 | 0.326 | 1 | ns |

**Table S9. *Roseburia* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Disease | Variable | Group 1 | Group 2 | n1 | n2 | p | p.adj | p.adj.signif |
| Control | Abundance | High | Low | 7 | 71 | 0.712 | 1 | ns |
| Control | Abundance | High | Moderate | 7 | 4 | 0.786 | 1 | ns |
| Control | Abundance | High | None | 7 | 19 | 0.345 | 1 | ns |
| Control | Abundance | Low | Moderate | 71 | 4 | 0.916 | 1 | ns |
| Control | Abundance | Low | None | 71 | 19 | 0.356 | 1 | ns |
| Control | Abundance | Moderate | None | 4 | 19 | 0.802 | 1 | ns |
| PD | Abundance | High | Low | 13 | 112 | 0.162 | 1 | ns |
| PD | Abundance | High | Moderate | 13 | 16 | 0.151 | 1 | ns |
| PD | Abundance | High | None | 13 | 43 | 0.219 | 1 | ns |
| PD | Abundance | Low | Moderate | 112 | 16 | 0.648 | 1 | ns |
| PD | Abundance | Low | None | 112 | 43 | 0.863 | 1 | ns |
| PD | Abundance | Moderate | None | 16 | 43 | 0.604 | 1 | ns |

**Table S10. *Lachnospira* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Disease | Variable | Group 1 | Group 2 | n1 | n2 | p | p.adj | p.adj.signif |
| Control | Abundance | High | Low | 7 | 71 | 0.51 | 1 | ns |
| Control | Abundance | High | Moderate | 7 | 4 | 0.319 | 1 | ns |
| Control | Abundance | High | None | 7 | 19 | 0.23 | 1 | ns |
| Control | Abundance | Low | Moderate | 71 | 4 | 0.493 | 1 | ns |
| Control | Abundance | Low | None | 71 | 19 | 0.345 | 1 | ns |
| Control | Abundance | Moderate | None | 4 | 19 | 0.825 | 1 | ns |
| PD | Abundance | High | Low | 13 | 112 | 0.759 | 1 | ns |
| PD | Abundance | High | Moderate | 13 | 16 | 0.67 | 1 | ns |
| PD | Abundance | High | None | 13 | 43 | 0.857 | 1 | ns |
| PD | Abundance | Low | Moderate | 112 | 16 | 0.265 | 1 | ns |
| PD | Abundance | Low | None | 112 | 43 | 0.366 | 1 | ns |
| PD | Abundance | Moderate | None | 16 | 43 | 0.684 | 1 | ns |

**Table S11. *Faecalibacterium* differential abundance results in subjects sorted by alcohol consumption categories.** Columns n1 and n2 denote sample sizes corresponding to the alcohol consumption groups being compared, Group 1 and Group 2, respectively. Adjusted p-values <0.05 = significant. ns = not significant.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Disease | Variable | Group 1 | Group 2 | n1 | n2 | p | p.adj | p.adj.signif |
| Control | Abundance | High | Low | 7 | 71 | 0.184 | 1 | ns |
| Control | Abundance | High | Moderate | 7 | 4 | 0.403 | 1 | ns |
| Control | Abundance | High | None | 7 | 19 | 0.529 | 1 | ns |
| Control | Abundance | Low | Moderate | 71 | 4 | 0.9 | 1 | ns |
| Control | Abundance | Low | None | 71 | 19 | 0.496 | 1 | ns |
| Control | Abundance | Moderate | None | 4 | 19 | 0.651 | 1 | ns |
| Control | Abundance | No data | None | 2 | 19 | 0.574 | 1 | ns |
| PD | Abundance | High | Low | 13 | 112 | 0.593 | 1 | ns |
| PD | Abundance | High | Moderate | 13 | 16 | 0.703 | 1 | ns |
| PD | Abundance | High | None | 13 | 43 | 0.341 | 1 | ns |
| PD | Abundance | Low | Moderate | 112 | 16 | 0.896 | 1 | ns |
| PD | Abundance | Low | None | 112 | 43 | 0.266 | 1 | ns |
| PD | Abundance | Moderate | None | 16 | 43 | 0.46 | 1 | ns |

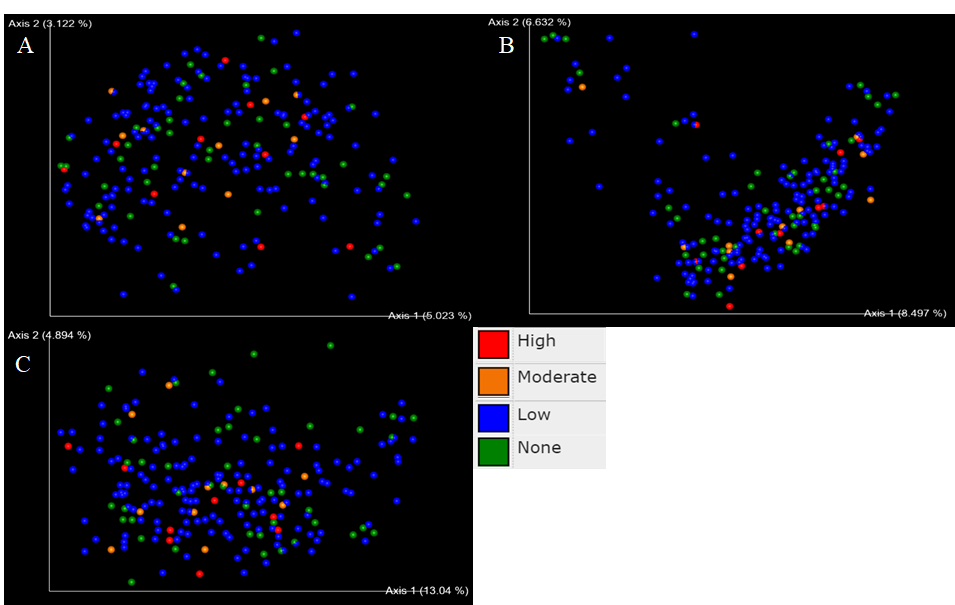
**Table S12. Pairwise PERMANOVA results based on the Weighted UniFrac beta diversity distances for comparing BMI and PD disease status.** Group 1 and 2 denote the two groups being compared. p-value <0.05 = significant. q-value <0.05 = significant.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Group 1 | Group 2 | Sample size | Permutation | pseudo-F | p-value | q-value |
| PD-normal | PD-overweight | 151 | 999 | 1.767261 | 0.110 | 0.220 |
| PD-normal | Control-normal | 110 | 999 | 2.773597 | 0.037 | 0.220 |
| PD-normal | Control-overweight | 108 | 999 | 1.925926 | 0.078 | 0.220 |
| PD-overweight | Control-normal | 121 | 999 | 1.574987 | 0.178 | 0.267 |
| PD-overweight | Control-overweight | 119 | 999 | 0.824537 | 0.436 | 0.482 |
| Control-normal | Control-overweight | 78 | 999 | 0.717412 | 0.482 | 0.482 |

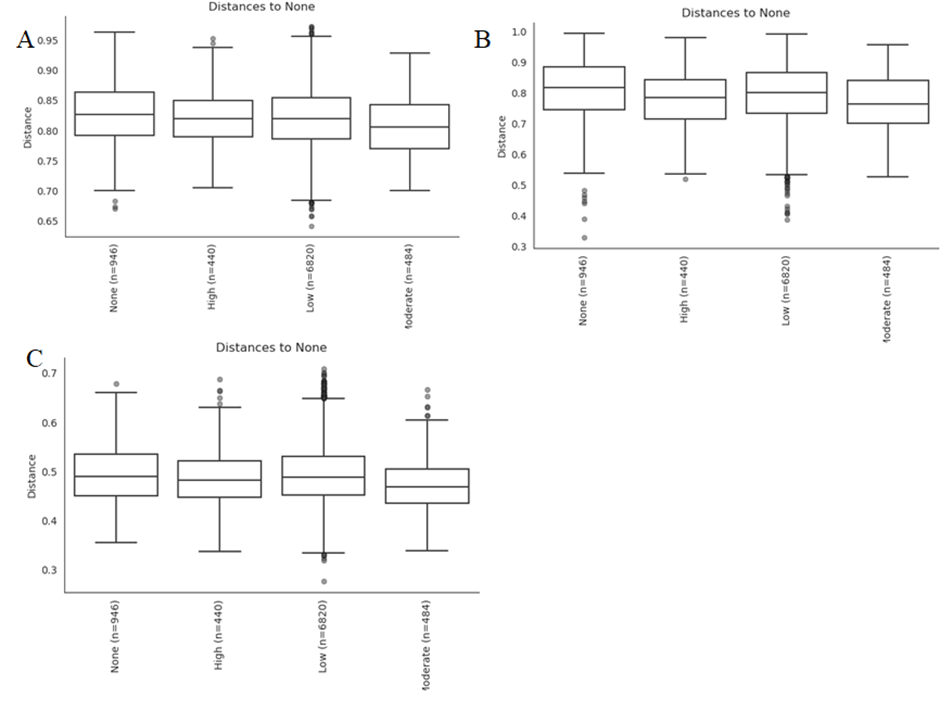
**Table S13. Two indicator families are associated with overweight PD individuals.** A indicates specificity (RA of a given taxa for each group) and B indicates fidelity (taxa present across the majority of samples from that specific group). Indicator value “stat” is calculated by multiplying A and B. p-values <0.05 = significant.

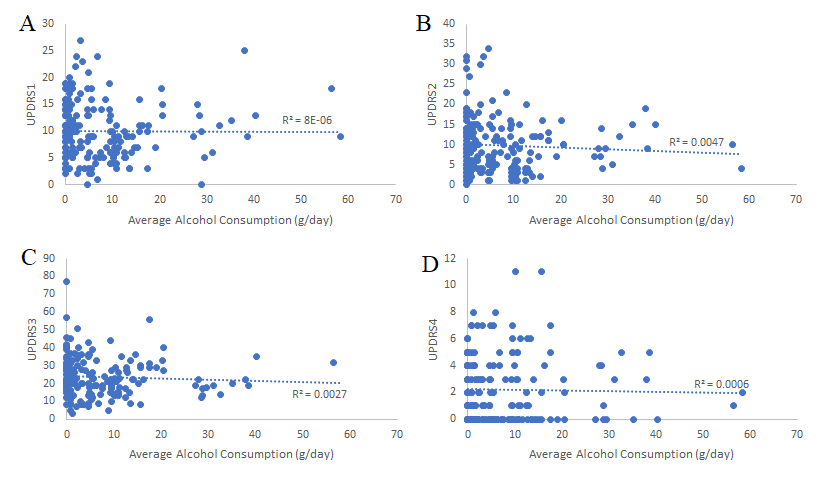
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Indicator taxa (family level) | A | B | stat | p-value |
| k\_\_Bacteria p\_\_Firmicutes c\_\_Clostridia o\_\_Clostridiales f\_\_Veillonellaceae | 0.3633 | 0.9878 | 0.599 | 0.020 |
| k\_\_Bacteria p\_\_Firmicutes c\_\_Clostridia o\_\_Clostridiales f\_\_Peptococcaceae | 0.5643 | 0.2683 | 0.389 | 0.035 |

**SUPPLEMENTAL FIGURES**

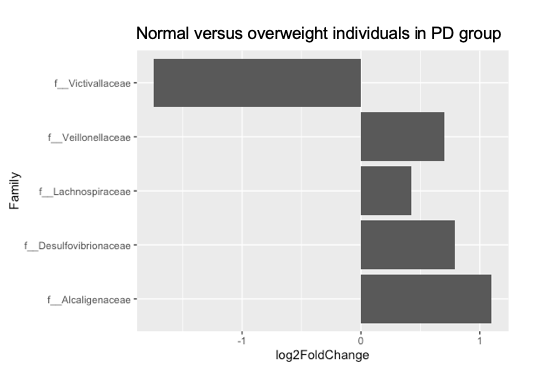


**Figure S1. Alcohol consumption categories do not have an effect on microbial community composition.** The gut microbiome composition of subjects was analyzed with (A) Jaccard, (B) Bray-Curtis, and (C) Unweighted UniFrac beta diversity metrics in QIIME2. Corresponding pairwise PERMANOVA statistics can be found in Tables S4-S6.

**Figure S2. No differences in the composition of gut microbiomes of different alcohol consumption groups were observed across any beta diversity metric.** Beta diversity of gut microbiotas across subjects in the three categories of alcohol consumption based on (A) Jaccard, (B) Bray-Curtis, and (C) Unweighted UniFrac. Corresponding pairwise PERMANOVA statistics can be found in Tables S4-S6.



**Figure S3. No correlation was found between average alcohol consumption (g/day) and UPDRS 1, 2, 3, and 4 scores.** (A)UPDRS 1 and (B) 2 scores are based on non-motor and motor challenges faced by subjects in their day-to-day lives, respectively. (C) UPDRS 3 scores represent the results of a movement analysis examination. (D) UPDRS 4 scores represent motor complications experienced by patients. Higher scores indicate more severe symptoms. Subjects with no alcohol consumption data were excluded from the analysis. Number of patients (n) = 195.

****

**Figure S4. Five families are differential abundant between overweight and normal PD patients.** The families Veillonellaceae, Lachnospiraceae, Desulfovibrionaceae, and Alcaligenaceae exhibited a positive fold change, indicating a higher abundance in overweight PD patients. The family Victivallaceae exhibited a negative fold change, indicating a higher abundance in normal PD patients.

**Script S1. QIIME2 commands.**

#PROCESSING RAW 16S rRNA SEQUENCES

#Importing data and visualizing demultiplexed sequence summary

qiime tools import \

--type 'SampleData[SequencesWithQuality]' \

--input-path /mnt/datasets/project\_2/parkinsons/parkinsons\_manifest.txt \

--output-path demux.qza \

--input-format SingleEndFastqManifestPhred33V2

qiime demux summarize \

--i-data demux.qza \

--o-visualization demux.qzv

#Sequence quality control and feature table construction using DADA2

qiime dada2 denoise-single \

--i-demultiplexed-seqs demux.qza \

--p-trim-left 0 \

--p-trunc-len 251 \

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

--o-table table-dada2.qza \

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

#Generating FeatureTable and FeatureData summaries

qiime feature-table summarize \

--i-table table-dada2.qza \

--o-visualization table.qzv \

--m-sample-metadata-file /mnt/datasets/project\_2/parkinsons/parkinsons\_metadata.txt

qiime feature-table tabulate-seqs \

--i-data rep-seqs-dada2.qza \

--o-visualization rep-seqs.qzv

#Generating tree for phylogenetic diversity analyses

qiime phylogeny align-to-tree-mafft-fasttree \

--i-sequences rep-seqs-dada2.qza \

--o-alignment aligned-rep-seqs.qza \

--o-masked-alignment masked-aligned-rep-seqs.qza \

--o-tree unrooted-tree.qza \

--o-rooted-tree rooted-tree.qza

#ALCOHOL

#Filter metadata to remove subjects who have taken more than 5 doses of antibiotics in the last 5 years (ie. 1 per year average)

qiime feature-table filter-samples \

--i-table /data/table.qza \

--m-metadata-file /mnt/datasets/project\_2/parkinsons/parkinsons\_metadata.txt \

--p-where "'5'<[Abx\_doses\_last\_5\_years]<'25'" \

--o-filtered-table alcohol-abx-filtered-table.qza

qiime feature-table summarize \

--i-table alcohol-abx-filtered-table.qza \

--o-visualization alcohol-abx-filtered-table.qzv \

--m-sample-metadata-file /mnt/datasets/project\_2/parkinsons/parkinsons\_metadata.txt

#Calculate beta diversity metrics using filtered table

qiime diversity core-metrics-phylogenetic \

--i-phylogeny /data/rooted-tree.qza \

--i-table alcohol-abx-filtered-table.qza \

--p-sampling-depth 5105 \

--m-metadata-file /data/alcohol/parkinsons\_metadata.txt \

--output-dir core-metrics-results

#Conduct analysis for all 4 beta diversity metrics (Jaccard, Bray-Curtis, unweighted and weighted Unifrac)

qiime diversity beta-group-significance \

--i-distance-matrix /data/alcohol/core-metrics-results/unweighted\_unifrac\_distance\_matrix.qza \

--m-metadata-file /data/alcohol/parkinsons\_metadata.txt \

--m-metadata-column Alcohol\_Group \

--o-visualization /data/alcohol/core-metrics-results/unweighted-unifrac-significance.qzv \

--p-pairwise

qiime diversity beta-group-significance \

--i-distance-matrix /data/alcohol/core-metrics-results/weighted\_unifrac\_distance\_matrix.qza \

--m-metadata-file /data/alcohol/parkinsons\_metadata.txt \

--m-metadata-column Alcohol\_Group \

--o-visualization /data/alcohol/core-metrics-results/weighted-unifrac-significance.qzv \

--p-pairwise

qiime diversity beta-group-significance \

--i-distance-matrix /data/alcohol/core-metrics-results/jaccard\_distance\_matrix.qza \

--m-metadata-file /data/alcohol/parkinsons\_metadata.txt \

--m-metadata-column Alcohol\_Group \

--o-visualization /data/alcohol/core-metrics-results/jaccard-significance.qzv \

--p-pairwise

qiime diversity beta-group-significance \

--i-distance-matrix /data/alcohol/core-metrics-results/bray\_curtis\_distance\_matrix.qza \

--m-metadata-file /data/alcohol/parkinsons\_metadata.txt \

--m-metadata-column Alcohol\_Group\

--o-visualization /data/alcohol/core-metrics-results/bray\_curtis-significance.qzv \

--p-pairwise

#ASV-based filtering in Qiime2

qiime feature-table filter-features \

--i-table /data/alcohol/alcohol-abx-filtered-table.qza \

--p-min-frequency 142 \

--o-filtered-table /data/alcohol/alcohol-abx-filtered-table-ASVfiltered.qza

qiime taxa filter-table \

--i-table /data/alcohol/alcohol-abx-filtered-table-ASVfiltered.qza \

--i-taxonomy /data/taxonomy.qza \

--p-exclude Archaea,chloroplast,mitochondria \

--o-filtered-table /data/alcohol/alcohol-abx-filtered-table-ASVtaxafiltered.qza

#Exporting files for analysis in R

qiime tools export \

--input-path table.qza \

--output-path exported

qiime tools export \

--input-path taxonomy.qza \

--output-path exported

qiime tools export \

--input-path rooted-tree.qza \

--output-path exported

nano exported/taxonomy.tsv

#Edited column names (`Feature ID` to `#OTUID`, `Taxon` to `taxonomy`, `Confidence` to `confidence`)

biom add-metadata \

-i exported/feature-table.biom \

-o exported/table-with-taxonomy.biom \

--observation-metadata-fp exported/taxonomy.tsv \

--sc-separated taxonomy

#BMI

#Filter metadata to keep only subjects of “control-normal”, “control-overweight”, “PD-normal”, “PD-overweight” Disease\_BMI categories

qiime feature-table filter-samples \

--i-table /data/table.qza \

--m-metadata-file /data/BMI/parkinsons\_metadata.txt \

--p-where "NOT [Disease\_BMI]='null'" \

--o-filtered-table Disease\_BMI-sorted.qza

qiime feature-table summarize \

--i-table Disease\_BMI-sorted.qza \

--o-visualization Disease\_BMI-sorted.qzv \

--m-sample-metadata-file /data/BMI/parkinsons\_metadata.txt

#Calculate beta diversity metrics using filtered table

qiime diversity core-metrics-phylogenetic \

--i-phylogeny /data/rooted-tree.qza \

--i-table Disease\_BMI-sorted.qza \

--p-sampling-depth 5105 \

--m-metadata-file /data/BMI/parkinsons\_metadata.txt \

--output-dir core-metrics-results

#Conduct beta-diversity analysis with weighted Unifrac

qiime diversity beta-group-significance \

--i-distance-matrix /data/BMI/core-metrics-results/weighted\_unifrac\_distance\_matrix.qza \

--m-metadata-file /data/BMI/parkinsons\_metadata.txt \

--m-metadata-column Disease\_BMI \

--o-visualization /data/BMI/core-metrics-results/qzv/weighted-unifrac-Disease\_BMI-significance.qzv \

--p-pairwise

#ASV-based filtering in Qiime 2

qiime feature-table filter-features \

--i-table /data/BMI/Disease\_BMI-sorted.qza \

--p-min-frequency 142 \

--o-filtered-table /data/BMI/Disease\_BMI-sorted-ASVfiltered.qza

qiime taxa filter-table \

--i-table /data/BMI/Disease\_BMI-sorted-ASVfiltered.qza \

--i-taxonomy /data/BMI/taxonomy.qza \

--p-exclude Archaea,chloroplast,mitochondria \

--o-filtered-table /data/BMI/Disease\_BMI-sorted-ASVtaxafiltered.qza

#Exporting files for analysis in R

qiime tools export \

--input-path taxonomy.qza \

--output-path exported\_files\_to\_R

qiime tools export \

--input-path rooted-tree.qza \

--output-path exported\_files\_to\_R

qiime tools export \

--input-path Disease\_BMI-sorted-ASVtaxafiltered.qza \

--output-path exported\_files\_to\_R

nano exported\_files\_to\_R/taxonomy.tsv

# Edit the column names and changes `Feature ID` to `#OTUID`, `Taxon` to `taxonomy`, `Confidence` to `confidence`

biom add-metadata \

-i exported\_files\_to\_R/feature-table.biom \

-o exported\_files\_to\_R/table-with-taxonomy.biom \

--observation-metadata-fp exported\_files\_to\_R/taxonomy.tsv \

--sc-separated taxonomy

**Script S2. R commands.**

#ALCOHOL

# Load CRAN packages

library(tidyverse)

library(vegan)

# Load Bioconductor packages

library(phyloseq)

library(DESeq2)

# Setup helper functions

# Calculate relative abundance

calculate\_RA <- function(x) x/sum(x)

# define this function below, which we'll use in a minute

gm\_mean <- function(x, na.rm = TRUE) {

exp(sum(log(x[x > 0]), na.rm = na.rm) / length(x))

}

# Export biom file and tree from QIIME2 and provide original metadata file

biom\_file <- import\_biom("table-with-taxonomy.biom")

metadata <- import\_qiime\_sample\_data("alcohol\_categorized\_metadata.tsv")

tree <- read\_tree\_greengenes("tree.nwk")

# Combine all information into a single phyloseq object

physeq <- merge\_phyloseq(biom\_file, metadata, tree)

# Convert taxonomic rank from numbers to proper names

colnames(tax\_table(physeq)) <- c("Kingdom", "Phylum", "Class", "Order",

"Family", "Genus", "Species")

# Keep only abundant ASVs

# First determine counts of ASV across all samples

total\_counts <- taxa\_sums(physeq)

# Calculate relative abundance for each ASV

relative\_abundance <- calculate\_RA(total\_counts)

# Determine which ASVs are more abundant than 0.1%

# Change this if you want a different cutoff (0.001 = 0.1%)

abundant <- relative\_abundance > 0.001

abundant\_taxa <- prune\_taxa(abundant, physeq)

# Check the resulting new phyloseq object with much fewer taxa

abundant\_taxa

Genus <- tax\_glom(abundant\_taxa, taxrank = "Genus", NArm = FALSE)

Genus

deseq\_alcohol\_feature

geo\_means <- apply(counts(deseq\_alcohol\_feature), 1, gm\_mean)

deseq\_alcohol\_feature <- estimateSizeFactors(deseq\_alcohol\_feature, geoMeans = geo\_means)

deseq\_alcohol\_feature <- DESeq(deseq\_alcohol\_feature, fitType="local")

feature\_diff\_abund <- results(deseq\_alcohol\_feature)

# Define cutoff for the adjusted p-value

alpha <- 0.05

# Reformat information as data frame including feature as variable

significant\_feature <- as\_tibble(feature\_diff\_abund, rownames = "feature")

# Keep only significant results and sort by adjusted p-value

significant\_feature <- filter(significant\_feature, padj < alpha)

significant\_feature <- arrange(significant\_feature, padj)

# Get the taxonomic information as a data frame

taxa\_df <- as\_tibble(as.data.frame(tax\_table(physeq)), rownames = "feature")

# Combine the significant features with taxonomic classification

significant\_feature <- inner\_join(significant\_feature, taxa\_df)

dim(significant\_feature)

# Plot differential abundance

significant\_feature %>%

ggplot(aes(x = log2FoldChange, y = Genus)) +

geom\_col()

# Convert to relative abundance

physeq\_RA <- transform\_sample\_counts(physeq, calculate\_RA)

genus\_RA <- tax\_glom(physeq\_RA, taxrank = "Genus", NArm = FALSE)

# Graph genera of interest and filter for that genus with subset\_taxa.

faecalibacterium <- subset\_taxa(genus\_RA, Genus == "g\_\_Faecalibacterium")

# The next step enables us to plot the data by changing the table from wide to long format.

blautia\_melt <- psmelt(faecalibacterium)

#define comparisons

library(rstatix)

library(ggpubr)

library(dplyr)

pre\_compare <- subset(blautia\_melt, alcohol\_categories != "No data")

my\_comparison <- list(pre\_compare, c("Control", "PD"))

#Do stat test to compare between groups

stat.test <- pre\_compare %>%

group\_by(Disease) %>%

t\_test(Abundance ~ alcohol\_categories) %>%

adjust\_pvalue(method = "bonferroni") %>%

add\_significance("p.adj")

stat.test

stat.test <- stat.test %>%

add\_xy\_position(x="Disease", dodge = 0.8)

# Plot with ggplot function

# Define what dataset to plot and what variable on x-axis and y-axis

my\_font\_size <- 5

asteriks\_font <- 25

level\_order <- c('None', 'Low', 'Moderate', 'High')

ggplot(filter(blautia\_melt, alcohol\_categories %in% c("High", "Moderate", "Low", "None")), aes(x = as.factor(Disease), y = Abundance)) +

# Visualize data as Boxplot

stat\_compare\_means(comparisons = my\_comparison, label.y = 0.085, tip.length = 0, label = "p.signif")+

stat\_pvalue\_manual(

stat.test, label = "p.adj", tip.length = 0.02,hide.ns = TRUE, label.y = 0.3, size = asteriks\_font)+

geom\_boxplot(aes(fill = fct\_relevel(alcohol\_categories, levels =c("None", "Low", "Moderate", "High")))) +

# Label for x-axis

xlab("Category") +

# Label for y-axis

ylab("Relative Abundance") +

# Plot title

ggtitle("Faecalibacterium") +

theme(plot.title = element\_text(hjust = 0.5))+

# colors for the different groups

scale\_fill\_manual(values=c("royalblue3", "darkgreen", "yellow2", "red4")) +

guides(fill = FALSE) +

# Set a standard theme for plot

theme\_bw(base\_size = 16)

#BMI

# Load packages

library(tidyverse)

library(vegan)

library(phyloseq)

library(DESeq2)

library(ggplot2)

# Load packages for stat test

library(ggpubr)

library(rstatix)

# Setup helper functions

# Calculate relative abundance

calculate\_RA <- function(x) x/sum(x)

# Calculate geometric mean

gm\_mean <- function(x, na.rm = TRUE){exp(sum(log(x[x > 0]), na.rm = na.rm) / length(x))}

# Importing all data files

biom\_file <- import\_biom("table-with-taxonomy.biom")

metadata <- import\_qiime\_sample\_data("parkinsons\_metadata\_BMI.txt")

tree <- read\_tree\_greengenes("tree.nwk")

# Combine all objects into phyloseq object

physeq <- merge\_phyloseq(biom\_file, metadata, tree)

# Set set of random numbers, make sure your analysis is reproducible

set.seed(711)

# Rename column names for taxonomic ranks

colnames(tax\_table(physeq)) <- c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus", "Species")

# Define cutoff for the adjusted p-value

alpha <- 0.05

# Subset samples into PD and control

physeq\_PD = subset\_samples(physeq, Disease == "PD")

physeq\_control = subset\_samples(physeq, Disease == "Control")

# Get data table with taxonomic information of each feature

taxa\_df\_PD <- as\_tibble(as.data.frame(tax\_table(physeq\_PD)), rownames = "feature")

taxa\_df\_ctl <- as\_tibble(as.data.frame(tax\_table(physeq\_control)), rownames = "feature")

#Calculate differential abundance at family level for PD, and run DEseq2 analysis

# Group PD by family rank

family\_PD <- tax\_glom(physeq\_PD, taxrank = "Family", NArm = FALSE)

deseq\_family\_PD <- phyloseq\_to\_deseq2(family\_PD, ~Disease\_BMI)

geo\_means\_family\_PD <- apply(counts(deseq\_family\_PD), 1, gm\_mean)

deseq\_family\_PD <- estimateSizeFactors(deseq\_family\_PD, geoMeans = geo\_means\_family\_PD)

deseq\_family\_PD <- DESeq(deseq\_family\_PD, fitType = "local")

family\_diff\_abund\_PD <- results(deseq\_family\_PD)

significant\_family\_PD <- as\_tibble(family\_diff\_abund\_PD, rownames = "feature")

significant\_family\_PD <- filter(significant\_family\_PD, padj < alpha)

significant\_family\_PD <- inner\_join(significant\_family\_PD, taxa\_df\_PD)

write\_csv(significant\_family, "significant\_family\_PD.csv")

significant\_family\_PD %>%

ggplot(aes(x = log2FoldChange, y = Family)) +

geom\_col()

#Calculate differential abundance at phylum level for PD and controls, and run DEseq2 analysis

# Group PD by phylum rank

phylum\_PD <- tax\_glom(physeq\_PD, taxrank = "Phylum", NArm = FALSE)

deseq\_phylum\_PD <- phyloseq\_to\_deseq2(phylum\_PD, ~Disease\_BMI)

geo\_means\_phylum\_PD <- apply(counts(deseq\_phylum\_PD), 1, gm\_mean)

deseq\_phylum\_PD <- estimateSizeFactors(deseq\_phylum, geoMeans = geo\_means\_phylum\_PD)

deseq\_phylum\_PD<- DESeq(deseq\_phylum\_PD, fitType = "local")

phylum\_diff\_abund\_PD <- results(deseq\_phylum\_PD)

significant\_phylum\_PD <- as\_tibble(phylum\_diff\_abund\_PD, rownames = "feature")

#Dont filter to view see p-values for Firm and Bact

#significant\_phylum <- filter(significant\_phylum, padj < alpha)

significant\_phylum\_PD <- inner\_join(significant\_phylum\_PD, taxa\_df\_PD)

# Group controls by phylum rank

phylum\_ctl <- tax\_glom(physeq\_control, taxrank = "Phylum", NArm = FALSE)

deseq\_phylum\_ctl <- phyloseq\_to\_deseq2(phylum\_ctl, ~Disease\_BMI)

geo\_means\_phylum\_ctl <- apply(counts(deseq\_phylum\_ctl), 1, gm\_mean)

deseq\_phylum\_ctl <- estimateSizeFactors(deseq\_phylum\_ctl, geoMeans = geo\_means\_phylum\_ctl)

deseq\_phylum\_ctl<- DESeq(deseq\_phylum\_ctl, fitType = "local")

phylum\_diff\_abund\_ctl <- results(deseq\_phylum\_ctl)

significant\_phylum\_ctl <- as\_tibble(phylum\_diff\_abund\_ctl, rownames = "feature")

#Dont filter to view p-values for Firm and Bact

#significant\_phylum <- filter(significant\_phylum, padj < alpha)

significant\_phylum\_ctl <- inner\_join(significant\_phylum\_ctl, taxa\_df\_ctl)

# RA for Controls

# Calculate relative abundance

physeq\_ctl\_RA <- transform\_sample\_counts(physeq\_control, calculate\_RA)

phylum\_ctl\_RA <- tax\_glom(physeq\_ctl\_RA, taxrank = "Phylum", NArm = FALSE)

# Graph Bacteroidetes and filter for that phylum with subset\_taxa.

bacteroidetes\_ctl <- subset\_taxa(phylum\_ctl\_RA, Phylum == "p\_\_Bacteroidetes")

# Plot the data by changing the table from wide to long format.

bacteroidetes\_ctl\_melt <- psmelt(bacteroidetes\_ctl)

# Plot with ggplot function

# Manually add stat to the plot

ggplot(bacteroidetes\_ctl\_melt, aes(x = as.factor(Disease\_BMI), y = Abundance)) +

geom\_boxplot(aes(fill = as.factor(Disease\_BMI))) +

xlab("Disease\_BMI") +

ylab("Relative Abundance") +

ggtitle("Bacteroidetes") +

scale\_fill\_manual(values=c("seagreen3", "indianred1")) +

guides(fill = FALSE) +

theme\_bw(base\_size = 16)

# Graph Firmicutes and filter for that phylum with subset\_taxa.

firmicutes\_ctl <- subset\_taxa(phylum\_ctl\_RA, Phylum == "p\_\_Firmicutes")

# Plot the data by changing the table from wide to long format.

firmicutes\_ctl\_melt <- psmelt(firmicutes\_ctl)

# Plot with ggplot function

# Manually add stat to the plot

ggplot(firmicutes\_ctl\_melt, aes(x = as.factor(Disease\_BMI), y = Abundance)) +

geom\_boxplot(aes(fill = as.factor(Disease\_BMI))) +

xlab("Disease\_BMI") +

ylab("Relative Abundance") +

ggtitle("Firmicutes") +

scale\_fill\_manual(values=c("seagreen3", "indianred1")) +

guides(fill = FALSE) +

theme\_bw(base\_size = 16) +

ylim(0,0.8)

#RA for PD

# Calculate relative abundance

physeq\_PD\_RA <- transform\_sample\_counts(physeq\_PD, calculate\_RA)

phylum\_PD\_RA <- tax\_glom(physeq\_PD\_RA, taxrank = "Phylum", NArm = FALSE)

# Graph Bacteroidetes and filter for that phylum with subset\_taxa.

bacteroidetes\_PD <- subset\_taxa(phylum\_PD\_RA, Phylum == "p\_\_Bacteroidetes")

# Plot the data by changing the table from wide to long format.

bacteroidetes\_PD\_melt <- psmelt(bacteroidetes\_PD)

# Plot with ggplot function

# Manually add stat to the plot

ggplot(bacteroidetes\_PD\_melt, aes(x = as.factor(Disease\_BMI), y = Abundance)) +

geom\_boxplot(aes(fill = as.factor(Disease\_BMI))) +

xlab("Disease\_BMI") +

ylab("Relative Abundance") +

ggtitle("Bacteroidetes") +

scale\_fill\_manual(values=c("seagreen3", "indianred1")) +

guides(fill = FALSE) +

theme\_bw(base\_size = 16)

# Graph Firmicutes and filter for that phylum with subset\_taxa.

firmicutes\_PD <- subset\_taxa(phylum\_PD\_RA, Phylum == "p\_\_Firmicutes")

# Plot the data by changing the table from wide to long format.

firmicutes\_PD\_melt <- psmelt(firmicutes\_PD)

# Plot with ggplot function

# Manually add stat to the plot

ggplot(firmicutes\_PD\_melt, aes(x = as.factor(Disease\_BMI), y = Abundance)) +

geom\_boxplot(aes(fill = as.factor(Disease\_BMI))) +

xlab("Disease\_BMI") +

ylab("Relative Abundance") +

ggtitle("Firmicutes") +

scale\_fill\_manual(values=c("seagreen3", "indianred1")) +

guides(fill = FALSE) +

theme\_bw(base\_size = 16) +

ylim(0,0.8)

#Indicator taxa

# Load packages

library(dplyr)

library(phyloseq)

library(indicspecies)

# Function to group asv table by higher order taxonomy, family

group\_by\_taxonomy = function(asv\_table, taxonomy, rank){

asv\_table = as.data.frame(asv\_table)

taxonomy = as.data.frame(taxonomy)

taxonomy$ASV = rownames(taxonomy)

asv\_table$ASV = rownames(asv\_table)

asv\_table = inner\_join(taxonomy,asv\_table,by="ASV")

asv\_table$taxa = apply(asv\_table[,1:rank],1,paste,collapse=" ")

asv\_table = asv\_table[,-(1:8)]

asv\_table = group\_by(data.frame(asv\_table),taxa)

taxa\_table = as.data.frame(summarise\_all(asv\_table,sum))

rownames(taxa\_table) = taxa\_table$taxa

return(taxa\_table[,-1])

}

# Re-load data

biom = import\_biom("table-with-taxonomy.biom")

taxa\_table = otu\_table(biom)

taxonomy = tax\_table(biom)

metadata = read.table("parkinsons\_metadata\_BMI.txt",sep="\t",header=T,row.names = 1, comment.char="")

# Create the taxa table for the rank of taxonomy required, family

# Kingdom=1, Phylum=2, Class=3, Order=4, Family=5, Genus=6, Species=7

taxa\_table = group\_by\_taxonomy(taxa\_table, taxonomy, 5)

# transpose taxa table

t\_table <- t(taxa\_table)

# two tables need same dimensions

# turn t\_table into data frame, make column with row names, make it called sample

new\_df <- rownames\_to\_column(as.data.frame(t\_table), var = "sample")

# make column with row names then select just sample and BMI column

new\_metadata <- rownames\_to\_column(metadata, var = "sample")%>%

select(c(sample, Disease\_BMI))

# combine above two into one table

joined\_df <- right\_join(new\_df, new\_metadata, by = "sample") %>%

drop\_na()

# calculate indicator values

indicator\_multipatt <- multipatt(as.matrix(joined\_df %>% select(-c(Disease\_BMI, sample))),

joined\_df$Disease\_BMI, duleg=TRUE)

indicator\_output = capture.output(summary(indicator\_multipatt,indvalcomp=TRUE))

write.table(indicator\_output,file="indiciator\_values.txt",row.names=F,quote=F)