**Data Analysis Procedure:**

**(Any step number with an \* included outputted results)**

**1)** **split\_libraries.py –m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt –f DoblerR1.fna –q DoblerR1.qual –o 02\_split\_library\_output**

Inputs: mapping file, fna file, qual file

Outputs: Histograms.txt, seqs.fna, and split\_library\_log.txt

**2)** **pick\_otus.py -i 02\_split\_library\_output/seqs.fna -r 00\_fastq\_fasta\_qual\_FILES/97\_otus.fasta -m uclust\_ref**

Inputs: seqs.fna file, 97\_otus reference file, and an otu\_picking\_method (Uclust)

Outputs: seqs\_otus.txt and seqs\_otus.log

**3) pick\_rep\_set.py -i 03\_uclust\_ref\_picked\_otus/seqs\_otus.txt -f 02\_split\_library\_output/seqs.fna -r 00\_fastq\_fasta\_qual\_FILES/97\_otus.fasta -o rep\_set.fna**

Inputs: seqs\_otus.txt, seqs.fna, and 97\_otus.fasta reference file

Outputs: rep\_set.fna file

**4) assign\_taxonomy.py -i 04\_rep/rep\_set.fna -r 00\_fastq\_fasta\_qual\_FILES/97\_otus.fasta -t 01\_check\_id\_map\_output/97\_otu\_taxonomy.txt**

Inputs: rep\_set.fna file, 97\_otus.fasta reference file, 97\_otu\_taxonomy.txt reference file

Outputs: rep\_set\_tax\_assignments.txt file and rep\_set\_tax\_assignments.log file

**5) make\_otu\_table.py -i 03\_uclust\_ref\_picked\_otus/seqs\_otus.txt -t 05\_uclust\_assigned\_taxonomy/rep\_set\_tax\_assignments.txt -o otu\_table.biom**

Inputs: seqs\_otus.txt file and rep\_set\_tax\_assignments.txt file

Outputs: otu\_table.biom file

**6)** **filter\_samples\_from\_otu\_table.py -i 06\_Otus/otu\_table.biom -o filtered\_biom\_table.biom --sample\_id\_fp Discarded8\_ids.txt --negate\_sample\_id\_fp**

Inputs: otu\_table.biom file and Discarded8\_IDS.txt (file includes 8 samples, one on each line, that are removed from table).

Outputs: filtered\_biom\_table.biom

**\*7) filter\_otus\_from\_otu\_table.py -i 06\_otus/filtered\_biom\_table.biom -o 06\_otus/min\_samples\_filtered\_biom\_table.biom --min\_samples 5**

Inputs: filtered\_biom\_table.biom file (--min\_samples 5 = OTU must be observed in minimum of 5 samples for OTU to be retained).

Outputs: min\_samples\_filtered\_biom\_table.biom file

**\*8)** **summarize\_taxa\_through\_plots.py -i min\_samples\_filtered\_biom\_table.biom -o taxa\_summary\_after\_min\_samples\_applied -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt**

Inputs: min\_samples\_filtered\_biom\_table.biom file and mapping file

Outputs: Folder containing taxa summary files (at different levels) and a folder containing taxa summary plots.

**Taxa Summaries by environmental biome groups (Insect, Leaf, Soil, Rhizosphere):**

**summarize\_taxa\_through\_plots.py -i Otu\_Tables/filtered\_biom\_table.biom -o taxa\_summary\_grouped -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -c env\_biome**

Inputs/Outputs: Same as above.

**9) make\_otu\_network.py -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -i min\_samples\_filtered\_biom\_table.biom -o otu\_network/**

Inputs: mapping file and min\_samples\_filtered\_biom\_table.biom file

Outputs: OTU network folder which contains edge and node files to be loaded into cytoscape along with props files labeled by category, which can used for coloring.

**10)** **make\_otu\_heatmap.py -i min\_samples\_filtered\_biom\_table.biom -o heatmap\_sorted.pdf -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -t ghost\_tree\_97\_80clusters.nwk**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and a reference phylogenetic tree extracted from Ghost-tree files.

Outputs: PDF file of a sorted heatmap.

**\*11) beta\_diversity\_through\_plots.py –i 06\_otus/ min\_samples\_filtered\_biom\_table.biom -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -o beta\_ghost\_div\_through\_plot/ –t ghost\_tree\_97\_80clusters.nwk**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and a reference phylogenetic tree extracted from Ghost-tree files.

Outputs: Folder consisting of a distance matrix file, principal coordinates file, and a folder containing results of PCoA plots (weighted and unweighted).

**12) jackknifed\_beta\_diversity.py -i min\_samples\_filtered\_biom\_table.biom -o Jack -e 5 -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -t ghost\_tree\_97\_80clusters.nwk**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and a reference phylogenetic tree extracted from Ghost-tree files.

Outputs: Folder containing several distance matrices (from [beta\_diversity.py](http://qiime.org/scripts/beta_diversity.html)), several rarefied OTU tables (from [multiple\_rarefactions\_even\_depth.py](http://qiime.org/scripts/multiple_rarefactions_even_depth.html)), several UPGMA trees (from [upgma\_cluster.py](http://qiime.org/scripts/upgma_cluster.html)), a supporting file and newick tree with support values (from [tree\_compare.py](http://qiime.org/scripts/tree_compare.html)), and Emperor PCoA plots.

**13) make\_bootstrapped\_tree.py -m Jack/unweighted\_unifrac/upgma\_cmp/master\_tree.tre -s Jack/unweighted\_unifrac/upgma\_cmp/jackknife\_support.txt -o Jack/unweighted\_unifrac/upgma\_cmp/jackknife\_named\_nodes.pdf**

Inputs: From results of jackknifed\_beta\_diversity.py – master\_tree.tre file and jackknife\_support.txt file. This can be done for weighted beta\_diversity results too.

Outputs: jackknife\_named\_nodes.pdf file

**14) make\_emperor.py -i 10\_Beta\_diversity/yes\_insects\_beta\_div/unweighted\_unifrac\_pc.txt -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -t**

**taxa\_summary/min\_samples\_filtered\_biom\_table\_L2.txt --n\_taxa\_to\_keep 5 -o biplots**

Inputs: principal coordinates files from beta\_diversity results, mapping file, and a taxonomy summary file.

Outputs: Folder that contains PCoA biplot.

**\*15) group\_significance.py -i min\_samples\_filtered\_biom\_table.biom -c env\_biome -s kruskal\_wallis -o otu\_kw\_condition.txt -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and test method of Kruskal-Wallis.

Outputs: otu\_kw\_condition.txt file that contains OTU ID, Test statistic, p-value, FDR p-value, Bonferroni p-value, group means, and OTU taxonomy

**\*16) compare\_categories.py --method anosim -i 10\_Beta\_diversity/yes\_insects\_beta\_div/weighted\_unifrac\_dm.txt -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -o anosim/condition\_weighted -c env\_biome**

Inputs: distance matrix file from beta diversity results (weighted or unweighted), mapping file, and test method of anosim.

Outputs: Folder containing a file called anosim\_results.txt for weighted and unweighted beta diversity results.

**\*17) alpha\_diversity.py -i 06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom -m chao1 -o adiv\_chao1.txt**

**GhostTree included:**

**alpha\_diversity.py -i 06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom -m PD\_whole\_tree -o adiv\_pd\_tree.txt -t ghost\_tree\_97\_80clusters.nwk**

Inputs:min\_samples\_filtered\_biom\_table.biom file, method of chao1 and method of PD\_whole\_tree, and a reference phylogenetic tree extracted from Ghost-tree files (newick format).

Outputs: adiv\_chao1.txt for alpha diversity without Ghost-tree and adiv\_pd\_tree.txt for alpha diversity using Ghost-tree.

**\*18) alpha\_rarefaction.py -i 06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom biom -o arare\_max100Abx -t ghost\_tree\_97\_80clusters.nwk -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -e 100**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and a reference phylogenetic tree extracted from Ghost-tree files (newick format).

Outputs: Interactive alpha rarefaction plots in arare\_max100Abx folder.

**19) compare\_distance\_matrices.py –method mantel –i 10\_Beta\_diversity/yes\_insects\_beta\_div/weighted\_unifrac\_dm.txt, 10\_Beta\_diversity/yes\_insects\_beta\_div/unweighted\_unifrac\_dm.txt –o mantel\_out –n 999**

**Mantel Correlogram:**

**Compare\_distance\_matrices.py –method mantel\_corr –i 10\_Beta\_diversity/yes\_insects\_beta\_div/weighted\_unifrac\_dm.txt, 10\_Beta\_diversity/yes\_insects\_beta\_div/unweighted\_unifrac\_dm.txt –o mantel\_correlogram\_out –n 999**

Inputs: Beta diversity distance matrix files (weighted and unweighted) and the method used (mantel and mantel\_corr respectively)

Outputs: One file containing the mantel R statistic and p-value. Two files containing information about the distance classes, their associated mantel statistics and p-values, and an image of the correlogram plot.

**20) PERMANOVA test (Using a program called PRIMER – Dr. Edwards assistance)**

**\*Results outputted from specific commands\***

**7) filter\_otus\_from\_otu\_table.py -i 06\_otus/filtered\_biom\_table.biom -o 06\_otus/min\_samples\_filtered\_biom\_table.biom --min\_samples 5**

Inputs: filtered\_biom\_table.biom file (--min\_samples 5 = OTU must be observed in minimum of 5 samples for OTU to be retained).

Outputs: min\_samples\_filtered\_biom\_table.biom file

-This step provides the filtered OTU table that is used for all the analyses.

\*8) **summarize\_taxa\_through\_plots.py -i min\_samples\_filtered\_biom\_table.biom -o taxa\_summary\_after\_min\_samples\_applied -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt**

Inputs: min\_samples\_filtered\_biom\_table.biom file and mapping file

Outputs: Folder containing taxa summary files (at different levels) and a folder containing taxa summary plots.

**Taxa Summaries by environmental biome groups (Insect, Leaf, Soil, Rhizosphere):**

**summarize\_taxa\_through\_plots.py -i Otu\_Tables/filtered\_biom\_table.biom -o taxa\_summary\_grouped -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -c env\_biome**

**-**This step provides the taxonomic summary composition.

**11) beta\_diversity\_through\_plots.py –i 06\_otus/ min\_samples\_filtered\_biom\_table.biom -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -o beta\_ghost\_div\_through\_plot/ –t ghost\_tree\_97\_80clusters.nwk**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and a reference phylogenetic tree extracted from Ghost-tree files.

Outputs: Folder consisting of a distance matrix file, principal coordinates file, and a folder containing results of PCoA plots (weighted and unweighted).

-This step provides the beta diversity 3D PCoA plots for weighted/unweighted and insects/no insects.

**15) group\_significance.py -i min\_samples\_filtered\_biom\_table.biom -c env\_biome -s kruskal\_wallis -o otu\_kw\_condition.txt -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and test method of Kruskal-Wallis.

Outputs: otu\_kw\_condition.txt file that contains OTU ID, Test statistic, p-value, FDR p-value, Bonferroni p-value, group means, and OTU taxonomy

-This step provides the Kruskal-Wallis group significance results based on the 4 environmental biomes.

**16) compare\_categories.py --method anosim -i 10\_Beta\_diversity/yes\_insects\_beta\_div/weighted\_unifrac\_dm.txt -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -o anosim/condition\_weighted -c env\_biome**

Inputs: distance matrix file from beta diversity results (weighted or unweighted), mapping file, and test method of anosim.

Outputs: Folder containing a file called anosim\_results.txt for weighted and unweighted beta diversity results.

-This step provides the output for the anosim test for both weighted/unweighted.

**17) alpha\_diversity.py -i 06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom -m chao1 -o adiv\_chao1.txt**

**06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom -m observed\_otus -o adiv\_observed\_otus.txt**

**GhostTree included:**

**alpha\_diversity.py -i 06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom -m PD\_whole\_tree -o adiv\_pd\_tree.txt -t ghost\_tree\_97\_80clusters.nwk**

Inputs:min\_samples\_filtered\_biom\_table.biom file, method of chao1, observed\_otus, and method of PD\_whole\_tree, and a reference phylogenetic tree extracted from Ghost-tree files (newick format).

Outputs: adiv\_chao1.txt for alpha diversity without Ghost-tree and adiv\_pd\_tree.txt for alpha diversity using Ghost-tree.

-This step provides the output for alpha diversity statistics using chao1, observed\_otus, and PD\_whole\_tree (using ghost-tree).

**18) alpha\_rarefaction.py -i 06\_otus/biom\_tables\_after\_apply\_min\_samples/min\_samples\_filtered\_biom\_table.biom -o arare\_max100Abx -t ghost\_tree\_97\_80clusters.nwk -m 01\_check\_id\_map\_output/FungiFleabeetleMappingFile\_corrected\_corrected.txt -e 100**

Inputs: min\_samples\_filtered\_biom\_table.biom file, mapping file, and a reference phylogenetic tree extracted from Ghost-tree files (newick format).

Outputs: Interactive alpha rarefaction plots in arare\_max100Abx folder.

-This step provides the alpha rarefaction plots for chao1, observed\_otus, and PD\_whole\_tree.