**Analysis of Fungal Variations among Communities Using ITS Primers**

**By: Jason Dulin**

**Biomedical Informatics**

**Abstract**

Fungi have critical roles in many ecosystems. They can be a threat to the health of humans and cause structural integrity issues in environments. Since most fungal diversity is unknown, there have been new methods to improve phylogenetic and non-phylogenetic analysis. The development of PCR primers for the Internal Transcribed Spacer (ITS), combined with next generation sequencing, has significantly improved the ability to profile fungal diversity of microbes. The high sequence variability within the ITS region allows for more accurate species identification. However, high variability makes multiple sequence alignment and phylogenetic analysis significantly less accurate between distant fungi. To solve this issue, a bioinformatics tool (ghost-tree) was applied. Ghost-tree integrates sequence data from two genetic markers into one phylogenetic tree, which can be used for reliable phylogenetic analysis. The main theme of the project is to determine if there are differences between these four fungal communities. In addition, it is a goal to determine which environments differ most significantly and the species driving these differences among communities. The alpha-diversity test concluded that the soil community had the greatest amount of different fungal species in one sample. Then, a beta-diversity test (UniFrac method) was applied and resulted in a 3D PCoA plot for visualization. A Kruskal-Wallis test was used to see if there were differences of means among the four environments. Then, a PERMANOVA test was conducted to determine environments that differ the most and which species made these environments different. By using pair-wise tests, it was determined that 4 groupings are significantly different from the environment being compared against: Leaf/Rhizosphere (p-value = .003), Leaf/Soil (p-value = .007), Insect/Rhizosphere (p-value = .009), and Insect/Leaf (p-value = .013). Focusing on Leaf/Rhizosphere, few species contributed a significant percent to the differences among the Leaf and Rhizosphere environments. It can be concluded that there are significant differences among the four communities, unknown species contribute significantly to these differences, and soil has the greatest fungal diversity within a sample. By increasing taxonomic resolution, further studies can focus on microbial shifts among fungal communities.

**Committee Members:**

Scott Kelley, Ph.D., Department of Biology, San Diego State University, Committee Chair

Matt Edwards, Ph.D., Department of Biology, San Diego State University, Committee Member

Georg E. Matt, Ph.D., Department of Psychology, San Diego State University, Committee Member

**Background:**

While profiling bacteria has become relatively easy, methods for fungal profiling have been much more difficult to develop. Some important fungal roles in the environment include supporting plant growth in soils, causing serious diseases in humans, and causing damage to the structural integrity of environments. Even though fungi are important, they remain largely unknown with profiling.

The slow progression of methods for fungal profiling is mainly because of the small subunit ribosomal RNA (SSU rRNA) gene. For bacterial communities, the SSU rRNA gene is the most common marker gene used for profiling. However, the accuracy of profiling is much less accurate for fungi. The SSU rRNA gene has evolved slowly in fungi, which means that the reads of fungal SSU rRNA do not differ across taxa enough for accurate taxonomic resolution. As a result, the focus of profiling fungi has turned to the Internal Transcribed Spacer (ITS) region (Schoch, 1). This is an intergenic region that has a much higher mutation rate than the fungal SSU rRNA marker gene. Since the ITS region has a higher mutation rate and greater sequence variability, it is used as the marker gene for fungal profiling.

Even though the ITS region provides greater taxonomic resolution, multiple sequence alignment across distant evolutionary taxa is much more unreliable using this marker gene. Multiple sequence alignment with the ITS region leads to inaccurate phylogenetic trees and diversity calculations. In bacterial communities, methods such as Faiths PD and UniFrac have been extremely useful for taxonomic resolution of community differences (“Faith DP. Conservation evaluation and phylogenetic diversity”). These metrics could be important analytical tools for fungi, however, the high sequence variability has limited their application. To solve this issue, ghost-tree was applied in order to provide greater taxonomic resolution. Ghost-tree is an open-source bioinformatics software tool for creating phylogenetic trees using marker genes (Fouquier). This tool uses sequences from an evolutionary conserved marker gene and can align sequences across distant taxa. Less conserved marker genes can be grafted to the evolutionary conserved marked gene for significantly greater taxonomic resolution.

The application of ghost-tree allows for the creation of accurate phylogenetic trees and phylogenetic diversity calculations. We applied ghost-tree to graft the fungal ITS extension tree to a reference foundation tree (UNITE OTUs v12\_11). This phylogeny can provide insight for the differences in fungal communities through phylogenetic tests and statistical tests. Our UniFrac plots and UniFrac distances explain much more of the variance compared to non-phylogenetic metrics. With this phylogeny, diversity tests such as UniFrac, Anosim, alpha-diversity, and PERMANOVA can be performed to study the differences in fungal communities. In addition to studying the differences between communities, the PERMANOVA can provide insight about which environments differ the most and the specific species driving these differences between communities.

**Test Data**:

The ITS dataset consisted of 43 samples extracted from four different communities: Soil, Leaf, Rhizosphere, and Insect. There were 9 Soil samples, 9 Leaf samples, 9 Rhizosphere samples, and 16 Insect samples. These samples were extracted in three different locations in Germany: Aumuehle, Westerhever, and Schoenberg.

Detailed collection and DNA processing procedures are described in an earlier study (Kelley & Dobler, 2011). The beetles were collected by sweep-nets or aspirators and identified by S. Dobler. All samples were returned live to the lab at the University of Hamburg before being frozen at -80°C. Leg, wings and elytra from beetles were removed and their carcasses washed in 10% hydrogen peroxide for 10 seconds to remove DNA contamination.

DNA was extracted from Beetle pools using PowerSoil DNA Isolation Kits (MoBio Laboratories Inc., Carlsbad) per manufacturer’s direction. Negative extraction controls were performed to identify DNA contamination during sample processing and included in the PCR amplification step. The sample set used the ITS-1F forward primer (5′-CTTGGTCATTTAGAGGAAGTAA-3′) sequence and the ITS2 reverse primer (5′-GCTGCGTTCTTCATCGATGC-3′) sequences to generate ITS1 sequence reads. All analyses referencing Python scripts described below were performed in MacQIIME 1.9.1-20150604.

**Methods:**

The sample set was first demultiplexed before picking Operational Taxonomic Units (OTUs). OTUs are clusters of similar sequences. Each sequence was then grouped into OTUs at the 97% sequence similarity level, using UCLUST-based closed-reference OTU picking (pick\_otus.py) with UNITE OTUs v12\_11 as the reference database. The most abundant sequence in each OTU was chosen as the OTU representative sequence (pick\_rep\_set.py). Along with selecting OTU representative sequences, taxonomy was assigned to each OTU representative sequence (assign\_taxonomy.py).

By using the taxonomic assignments file and the OTU mapping file, an OTU table was created (make\_otu\_table.py). The OTU table was filtered in two different ways. 8 samples were filtered out of the table that did not have at least 100 counts/sample (filter\_samples\_from\_otu\_table.py). The count refers to the minimum total OTUs in a sample for that sample to be retained. Next, OTU’s were then filtered out of the table (filtered\_otus\_from\_otu\_table.py). The OTUs removed were based on a minimum number (7) of samples an OTU must be observed in for that specific OTU to be retained. This step did not remove samples from the table, but removed OTUs within the samples.

The filtered OTU table was used to summarize the taxonomic composition of the samples (summarize\_taxa\_through\_plots.py). The taxonomy summary shows the taxonomic composition of each sample ID. In addition, the summary provides a total composition percentage for each taxonomy identified.

An alpha diversity test was applied in order to determine the species richness levels for the fungal communities. By providing the mapping file, filtered OTU table, and the phylogenetic tree, the alpha diversity test outputs a rarefaction plot. This plot shows the differences among the four fungal community species richness levels. Three different methods can be used for rarefaction plots: PD\_whole\_tree, chao1, and observed\_otus. The output also provides statistics such as average abundance of species richness and error values for each of the three methods. This is done for every sample and grouped by their environmental biome. Then, one-way and two-way interaction tests were applied and shown in Appendix C.

Next, the filtered OTU table was used for computing beta diversity distance matrices and generating 3D PCoA plots (beta\_diversity\_through\_plots.py). In order to use phylogenetic beta diversity metrics using UniFrac, a phylogenetic tree is required as an input. Since there is much still unknown about fungal profiling, there are limited reference trees to analyze fungal samples. Ghost-tree (an open-source bioinformatics software tool) provides a solution to this problem. This software grafted the ITS fungal sequences to the tips of a pre-existing fungal tree (UNITE database).

The phylogenetic tree, mapping file, and filtered OTU table are required as inputs in order to run a beta-diversity test. The distance matrices contain a dissimilarity value for each pairwise comparison. Then, a principal coordinate analysis was used to compare groups of samples based on phylogenetic metrics (“Werner Lab”). Weighted UniFrac distance measures among fungal communities were calculated in order to cluster fungal communities based on their UniFrac distances at an even sampling depth and visualized by principle coordinated analyses (PCoA). PCoA is a technique that helps extract and visualize highly-informative components of variation from complex, multidimensional data. This transformation maps the samples present in the distance matrix to a new set of orthogonal axes. The maximum variation is explained in the first principal coordinate, the second largest amount of variation is explained in the second principal coordinate, and this pattern continues. These principal coordinates were plotted in three dimensions to demonstrate a visualization of differences between fungal samples and communities.

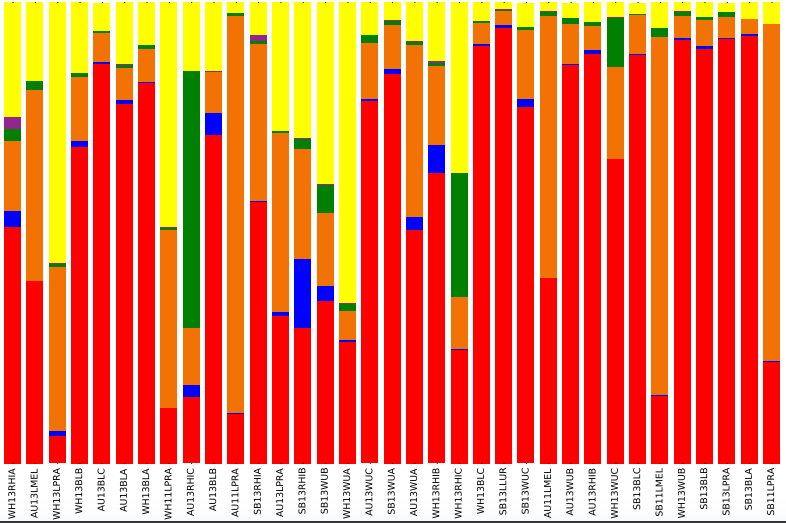
Differences among communities cannot be determined solely on visual representation. Once phylogenetic analyses were completed, statistical analyses were conducted (Caporaso, 3). A Kruskal-Wallis test was used to compare OTU frequencies in sample groups and to distinguish whether or not there are statistically significant differences between the OTU abundances in the different sample groups (group\_significance.py). This test is an expansion of an ANOVA to cases where sample means are unequal and distribution is not normal. The output of this test provides the OTU ID, test-statistic, raw p-value, False Discovery Rate p-value, Bonferroni p-value, and the sample group means. Sample groups were determined by environmental biome. These samples were grouped together based on having the same value in the mapping file under the environmental biome heading. Samples without this value are not included in the comparison. Filtering OTUs, which are found in a low percentage of samples, was significant to prevent 0 variance errors and spurious significance for low abundance OTUs. In addition, the Kruskal-Wallis test also outputs the OTU abundances of the four communities.

Next, an ANOSIM test was conducted (compare\_categories.py). This test allows for the analysis of the strength and statistical significance of sample groupings using the distance matrix (weighted or unweighted) as the main input. The distance matrices were provided from the results of the UniFrac test. ANOSIM tests whether two or more groups of samples are significantly different based on a categorical variable in the mapping file. Since this is a nonparametric test, significance is determined through permutations (999). The output of this test results in a file containing the effect size statistic and p-value.

The final test conducted was a PERMANOVA through a program called PRIMER. PERMANOVA uses an ANOVA experimental design and returns a pseudo-F value and p-value. For the purpose of this project, the analysis focused on the p-value to determine differences between fungal communities by running pair-wise tests. Then, PRIMER was used to analyze which species were driving these differences between the fungal communities. The significant fungal species were selected and profiled through a program called BLAST.

**Results:**

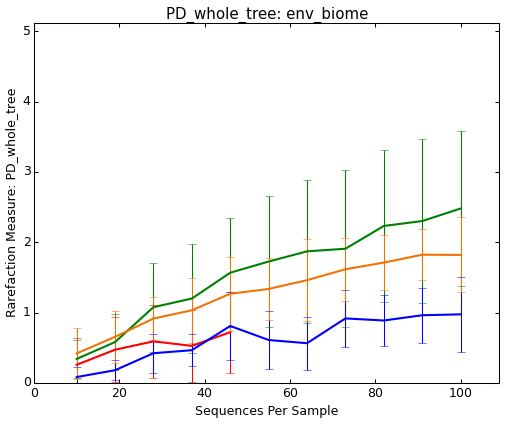
**Taxonomic Composition Summary (Level 2 - Phylum)**

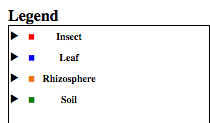




Taxonomic summary of each sample based on phylum.

**Alpha Diversity Rarefaction Plot:**

****

****

This figure shows the within sample species richness levels at different sequence depths for the four fungal communities.

The alpha rarefaction plot shows the soil community as the greatest level of species richness. This means that the soil environment has the greatest number of different fungal species within one sample. The Insect community had the lowest fungal species diversity and the line stopped at approximately 50 sequences per samples. The reason for the cutoff line is because at least one of the samples in that category does not have that many sequences.

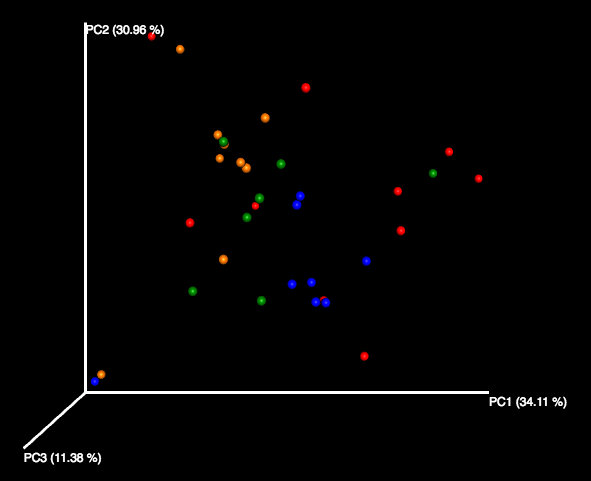
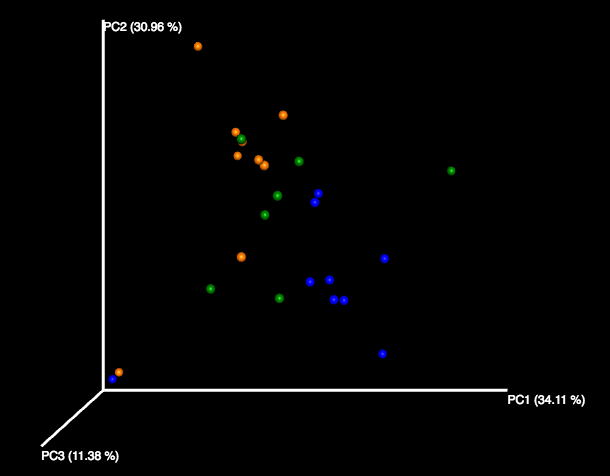
**Table 1. Alpha Diversity Statistics**

****

Table 1 represents alpha-diversity statistics, or within-sample diversity, using an OTU table. The different metrics used to determine within-sample diversity were chao1, observed\_otus, and PD\_whole\_tree. For chao1, the confidence interval (lower and upper bounds) were included for each sample. The metrics demonstrate similar patterns for individual sample diversity, which strengthens the validity of these results.

**PCoA Plots (by environmental biome)**

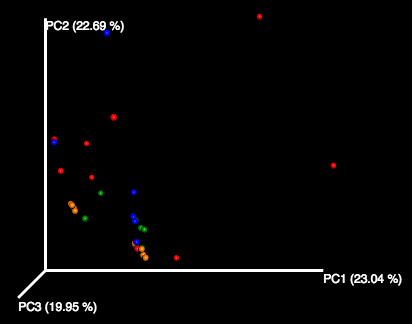
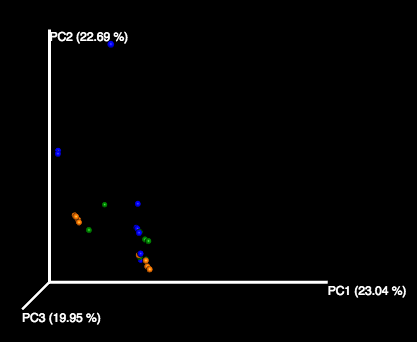
**Weighted – Insects included Weighted – No insects**

** **

3D PCoA Plot showing weighted diversity between samples not including the Insect community.

3D PCoA Plot showing weighted diversity between samples including the Insect community.

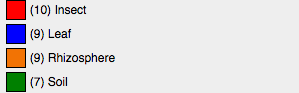
**Unweighted – Insects included Unweighted – No insects**

** **

3D PCoA Plot showing unweighted diversity between samples not including the Insect community.

3D PCoA Plot showing unweighted diversity between samples including the Insect community.

**Key (Weighted and Unweighted)**

****

**Group Significance Results (Test – Kruskal-Wallis):**

By providing the OTU table, the category (environmental biome), the test (Kruskal-Wallis), and the mapping file, an output file was produced to analyze group significance. The output included OTU IDs, p-values for each OTU, corrected p-values by FDR for each OTU, group abundance means for each OTU, and the taxonomy of each OTU. With a p-value cutoff of .05 (5%), there were 7 OTUs that were statistically significant. These significant OTUs were then profiled by inputting the representative sequence of the OTU into BLAST.

**Table 2 – Group Significance Results – Statistically significant OTUs**

|  |  |  |
| --- | --- | --- |
| **OTU ID** | **FDR\_Corrected P-value** | **Taxonomy** |
| **Denovo463** | .00893343 | Unassigned |
| **Denovo887** | .03467254 | Unassigned |
| **AH008235** | .04508437 | K\_fungi; p\_Ascomycota |
| **Denovo1377** | .04508437 | Unassigned |
| **Denovo491** | .04508437 | Unassigned |
| **Denovo847** | .04508437 | K\_fungi; unidentified; unculturedfungus |
| **JN995648** | .04508437 | k\_Fungi; p\_Ascomycota; c\_Leotiomycetes; o\_Helotiales; f\_Incertae\_sedis |

Table 2 shows the statistically significant OTUs including ID names, FDR\_Corrected P-value and OTU taxonomy.

**Table 3 – Group means for statistically significant OTUs**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **OTU ID** | **Soil mean** | **Insect mean** | **Leaf mean** | **Rhizosphere mean** |
| **Denovo463** | 38.42857143 | 81.9 | 696.3333333 | 521.7777778 |
| **Denovo887** | 13.71428571 | 1.4 | 1.666666667 | 9.444444444 |
| **AH008235** | 1.857142857 | 1.1 | 0.777777778 | 33.88888889 |
| **Denovo1377** | 0.142857143 | 0.4 | 0 | 6.888888889 |
| **Denovo491** | 7.857142857 | 0.7 | 0.888888889 | 3.444444444 |
| **Denovo847** | 0.714285714 | 0.4 | 1.111111111 | 4.777777778 |
| **JN995648** | 0.142857143 | 0.1 | 0.222222222 | 4.444444444 |

Table 3 shows the ranked group means for each fungal community. This identifies which OTUs are more abundant in certain environments.

**Profile of OTU representative sequences including sample it was grouped in, using BLAST (“U.S. National Library of Medicine”):**

1) **denovo463** (WH13WUB\_16) - Plantago lanceolata voucher KSUFS728 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence. 100% identity match.

2) **denovo887** (WH13BLA\_161) - Fungal sp. MKOTU44 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1 and 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence. 99% identity match.

3) **AH008235** (accession ID = AH008235) - Dark septate endophyte DS16b 18S ribosomal RNA gene, partial sequence. 100% identity match.

4) **denovo1377** (SB13WUB\_5134) - Uncultured fungus clone Contig81-117-1065\_2913 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, complete sequence; and 5.8S ribosomal RNA gene, partial sequence. 85% identity match.

5) **denovo491** (WH13WUA\_417) – Uncultured Dothideomycetes genomic DNA containing 18S rRNA gene, ITS1, 5.8s rRNA gene, ITS2 and 28S rRNA gene, clone 10S10C22. 100% identity match.

6) **denovo847** (SB13RHIC\_149) - Uncultured Phialea clone d155b\_1\_1 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence. 99% identity match.

7) **JN995648** (accession ID = JN995648) - Cadophora sp. AU\_BD06 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene, complete sequence; and internal transcribed spacer 2, partial sequence. 99% identity match.

**Category Comparison Results (Test – ANOSIM):**

**Table 4. Anosim Results**

|  |  |  |
| --- | --- | --- |
| **Method - Anosim** | Weighted | Unweighted |
| **Test Statistic Name** | R | R |
| **Sample Size** | 35 | 35 |
| **Number of Groups** | 4 | 4 |
| **Test Statistic** | 0.20042495 | 0.13531855516443 |
| **P-value** | 0.001 | .005 |
| **Number of permutations** | 999 | 999 |

Table 4 is a category comparison that provides the size of sample, number of groups tested, the test statistic, the p-value, and the number of permutations used. The R values are closer to 0 meaning that there is no significant dissimilarity between the groups.

**PERMANOVA:**

PERMANOVA, conducted by PRIMER, provided results that demonstrate the differences between two communities. Pair-wise tests were used to show the significance of every combination of communities. Since there were four fungal communities, there are 6 possible combinations of the communities to test. Based on the p-values for each combination, the statistically significant tests were determined. Below, a table is provided that represents the t-values, p-values, and unique permutations for each pair-wise test.

**Table 5. Pair-Wise Tests (Level 6 – Species)**

|  |  |  |  |
| --- | --- | --- | --- |
| **Groups** | **t-value** | **p-value** | **Unique permutations** |
| **Leaf, Rhizosphere** | 1.7984 | 0.003 | 979 |
| **Leaf, Soil** | 1.9481 | 0.007 | 916 |
| **Insect, Rhizosphere** | 1.7418 | 0.009 | 993 |
| **Insect, Leaf** | 1.8425 | 0.013 | 998 |
| **Insect, Soil** | 1.1448 | 0.192 | 961 |
| **Soil, Rhizosphere** | 1.1297 | 0.248 | 906 |

Table 5 shows which communities are significantly different from the community being tested against. As observed in the table, Insect/Soil and Soil/Rhizosphere are the groups not significantly different because of a p-value above the cutoff of 5%. Leaf/Rhizosphere has the greatest statistical significance (p-value=.003).

Based on significance results from the pair-wise tests, a program within PRIMER (Simper) was used to identify the species driving these differences between communities. Tables 5-8 provide the species contribution percentage for the significantly different communities. A 5% contribution to the differences between communities was applied in order to select the most significant species.

**Table 6. Species differing between Leaf and Rhizosphere Communities**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Leaf** | **Rhizosphere** |  |  |  |
| **Species** | **Avg. Abundance** | **Avg. Abundance** | **Avg. Dissimilarity** | **Diss/SD** | **Contribution Percentage** |
| **k\_Fungi; unidentified** | 0.27 | 0.32 | 3.82 | 0.98 | 13.29% |
| **Unassigned; Other** | 0.91 | 0.80 | 2.97 | 1.00 | 10.32% |
| **k\_Fungi; p\_Ascomycota; Other** | 0.04 | 0.13 | 2.16 | 1.18 | 7.49% |
| **k\_Fungi; p\_Ascomycota; c\_Leotiomycetes; o\_Helotiales; f\_Incertae\_sedis; Other** | 0.02 | 0.10 | 1.85 | 0.78 | 6.43% |

Table 6 shows the average abundances of the Leaf/Rhizosphere groups, average dissimilarity, and the contribution percentage to the diversity differences for each species.

**Table 7. Species differing between Leaf and Soil Communities**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Leaf** | **Soil** |  |  |  |
| **Species** | **Avg. Abundance** | **Avg. Abundance** | **Avg. Dissimilarity** | **Diss/SD** | **Contribution Percentage** |
| **Unassigned;Other;Other;Other;Other;Other** | 0.91 | 0.65 | 5.71 | 1.47 | 15.99% |
| **k\_\_Fungi;p\_\_unidentified;c\_\_unidentified;o\_\_unidentified;f\_\_unidentified;g\_\_unidentified** | 0.27 | 0.39 | 3.58 | 1.31 | 10.03% |
| **k\_\_Fungi;p\_\_Basidiomycota;c\_\_Agaricomycetes;o\_\_Agaricales;f\_\_Hygrophoraceae;g\_\_Hygrocybe** | 0.03 | 0.15 | 3.03 | 0.56 | 8.49% |
| **k\_\_Fungi;Other;Other;Other;Other;Other** | 0.09 | 0.15 | 2.44 | 1.19 | 6.83% |
| **k\_\_Fungi;p\_\_Basidiomycota;c\_\_Agaricomycetes;Other;Other;Other** | 0.04 | 0.14 | 2.32 | 0.66 | 6.51% |
| **k\_\_Fungi;p\_\_Ascomycota;c\_\_Sordariomycetes;o\_\_Hypocreales;f\_\_Nectriaceae;g\_\_Fusarium** | 0.05 | 0.15 | 2.23 | 0.98 | 6.25% |

Table 7 shows the average abundances of the Leaf/Soil groups, average dissimilarity, and the contribution percentage to the diversity differences for each species.

**Table 8. Species differing between Insect and Rhizosphere Communities**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Insect** | **Rhizosphere** |  |  |  |
| **Species** | **Avg. Abundance** | **Avg. Abundance** | **Avg. Dissimilarity** | **Diss/SD** | **Contribution Percentage** |
| Unassigned;Other;Other;Other;Other;Other | 0.57 | 0.80 | 6.66 | 1.52 | 14.91% |
| k\_\_Fungi;p\_\_unidentified;c\_\_unidentified;o\_\_unidentified;f\_\_unidentified;g\_\_unidentified | 0.35 | 0.32 | 5.19 | 1.15 | 11.62% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Dothideomycetes;o\_\_Capnodiales;f\_\_Mycosphaerellaceae;g\_\_Cladosporium | 0.33 | 0.11 | 5.01 | 1.04 | 11.21% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Eurotiomycetes;o\_\_Eurotiales;f\_\_Trichocomaceae;g\_\_Penicillium | 0.25 | 0.16 | 2.48 | 0.57 | 5.55% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Sordariomycetes;o\_\_Hypocreales;f\_\_Nectriaceae;Other | 0.09 | 0.03 | 2.28 | 0.44 | 5.10% |

Table 8 shows the average abundances of the Insect/Rhizosphere groups, average dissimilarity, and the contribution percentage to the diversity differences for each species.

**Table 9. Species differing between Insect and Leaf Communities**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Insect** | **Leaf** |  |  |  |
| **Species** | **Avg. Abundance** | **Avg. Abundance** | **Avg. Dissimilarity** | **Diss/SD** | **Contribution Percentage** |
| Unassigned;Other;Other;Other;Other;Other | 0.57 | 0.91 | 8.66 | 1.70 | 21.21% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Dothideomycetes;o\_\_Capnodiales;f\_\_Mycosphaerellaceae;g\_\_Cladosporium | 0.33 | 0.15 | 5.41 | 1.10 | 13.25% |
| k\_\_Fungi;p\_\_unidentified;c\_\_unidentified;o\_\_unidentified;f\_\_unidentified;g\_\_unidentified | 0.35 | 0.27 | 4.83 | 1.29 | 11.82% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Eurotiomycetes;o\_\_Eurotiales;f\_\_Trichocomaceae;g\_\_Penicillium | 0.25 | 0.15 | 2.94 | 0.61 | 7.19% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Sordariomycetes;o\_\_Hypocreales;f\_\_Nectriaceae;Other | 0.09 | 0.03 | 2.63 | 0.46 | 6.43% |
| k\_\_Fungi;p\_\_Ascomycota;c\_\_Eurotiomycetes;o\_\_Chaetothyriales;f\_\_Incertae\_sedis;g\_\_Coniosporium | 0.10 | 0.01 | 2.38 | 0.44 | 5.83% |

Table 9 shows the average abundances of the Insect/Leaf groups, average dissimilarity, and the contribution percentage to the diversity differences for each species.

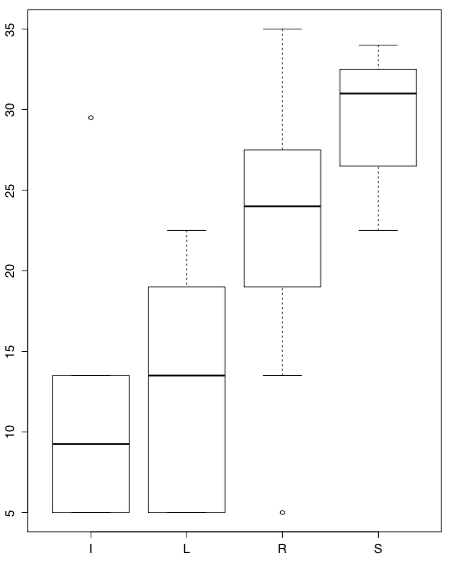
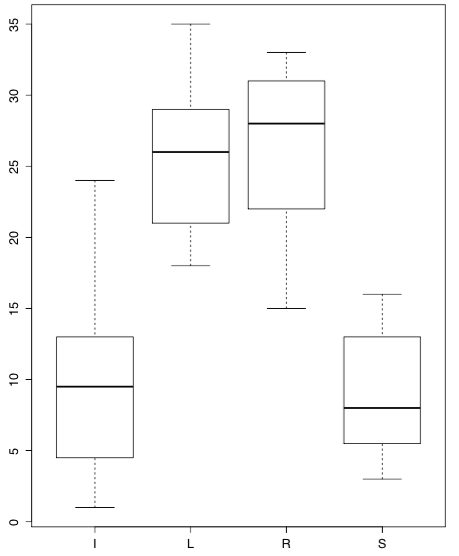
**Conclusion:**

Through phylogenetic diversity tests and statistical significance tests, it can be concluded that there are significant differences between the four communities. Through two-way interaction tests (location and environmental biome), it was concluded that location had no significance to the differences between fungal species diversity. From this conclusion, environmental biome became the sole focus of studying the differences between fungal species. Based on environmental biome, pairwise tests demonstrated which environments had the most significant differences of fungal species. Leaf/Soil and Leaf/Rhizosphere have the greatest variation of fungal species between their environments. Through within-sample diversity calculations, it was also concluded that the soil community contained the greatest number of different species within one sample. From these tests, it was apparent that unknown and uncultured fungal species provided the greatest contribution for making these communities different. By applying phylogeny into diversity metrics, taxonomic resolution was enhanced for detecting differences among these fungal communities. This can lead to a better understanding of the makeup and differences within fungal ecosystems. Further research could focus upon the fungal shifts in microbial communities based on more accurate taxonomies provided by phylogeny.

**Appendix A**

**Kruskal-Wallis Box Plot and Whiskers (based on the 7 statistically significant OTU’s found in Kruskal-Wallis Test):**

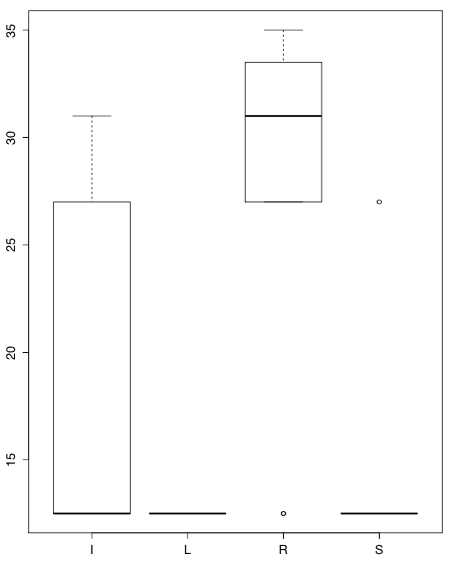
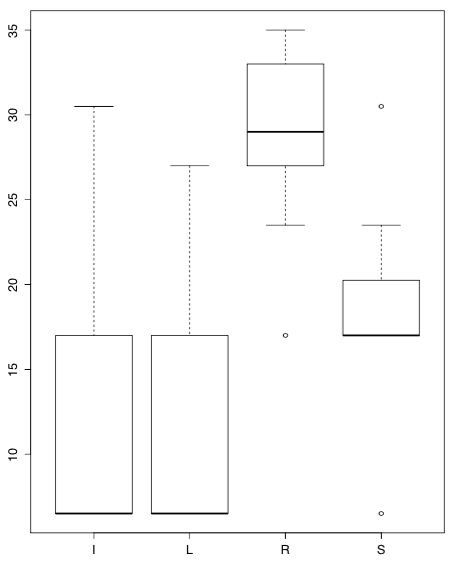
**Denovo463 by environmental biome Denovo887 by environmental biome**

****

Box Plots of OTU abundances, for each sample, grouped by environmental biome:

(I=Insect, L=Leaf, R=Rhizosphere, S=Soil)

**AH008235 by environmental biome Denovo1377 by environmental biome**

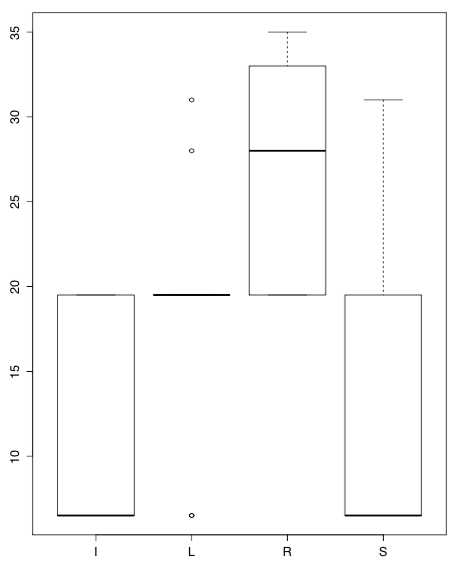
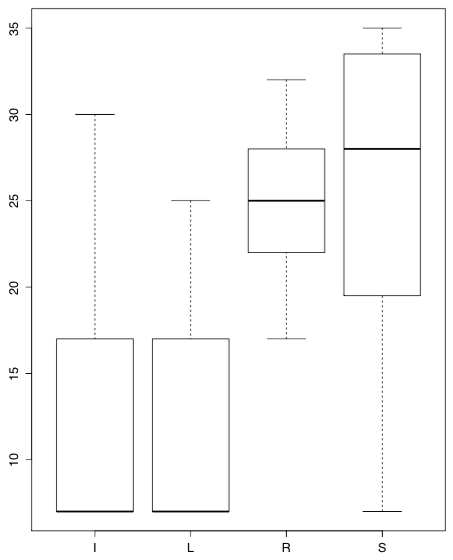
****

Box Plots of OTU abundances, for each sample, grouped by environmental biome:

(I=Insect, L=Leaf, R=Rhizosphere, S=Soil)

**Appendix A Continued**

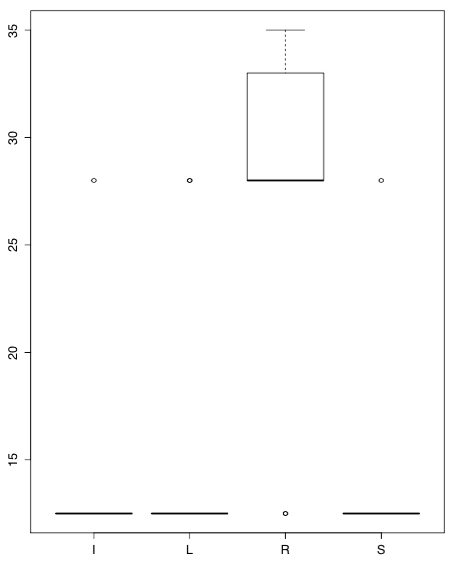
**Denovo491 by environmental biome Denovo847 by environmental biome**

****

Box Plots of OTU abundances, for each sample, grouped by environmental biome:

(I=Insect, L=Leaf, R=Rhizosphere, S=Soil)

**JN995648 by environmental biome**

****

Box Plots of OTU abundances, for each sample, grouped by environmental biome:

(I=Insect, L=Leaf, R=Rhizosphere, S=Soil)

**Appendix B**

**Kruskal-Wallis Pairwise Tests (based on the abundances, for each sample, of the 7 statistically significant OTU’s in the four environmental biomes):**

**Key = S=Soil, L=Leaf, R=Rhizosphere, I=Insect**

**\* Stands for significant difference between environments shown by p-value < .05**

**Denovo463 by environmental biome**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | 15.9 | 7.854849 | 23.945151 | 0.0000431\* |
| **R-I** | 15.5666667 | 7.521516 | 23.611818 | 0.0000594\* |
| **S-I** | -0.9571429 | -9.586016 | 7.67173 | 0.9903301 |
| **R-L** | -0.3333333 | -8.587484 | 7.920817 | 0.9995181 |
| **S-L** | -16.8571429 | -25.681201 | -8.033085 | 0.0000717\* |
| **S-R** | -16.5238095 | -25.347868 | -7.699751 | 0.000096\* |

This table shows the significant differences of group interactions for OTU Denovo463.

**Denovo887 by environmental biome**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | 2.261111 | -6.947078 | 11.46930 | 0.9088828 |
| **R-I** | 11.150000 | 1.941811 | 20.35819 | 0.0127922\* |
| **S-I** | 18.507143 | 8.630846 | 28.38344 | 0.0000950\* |
| **R-L** | 8.888889 | -0.558514 | 18.33629 | 0.0710846 |
| **S-L** | 16.246032 | 6.146333 | 26.34573 | 0.0007212\* |
| **S-R** | 7.357143 | -2.742556 | 17.45684 | 0.2183715 |

This table shows the significant differences of group interactions for OTU Denovo887.

**AH008235 by environmental biome**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | 0.300000 | -9.305673 | 9.9056727 | 0.9997767 |
| **R-I** | 15.911111 | 6.305438 | 25.5167838 | 0.0005022\* |
| **S-I** | 5.657143 | -4.645477 | 15.9597628 | 0.4553051 |
| **R-L** | 15.611111 | 5.755899 | 25.4663233 | 0.0008674\* |
| **S-L** | 5.357143 | -5.178522 | 15.8928079 | 0.5209045 |
| **S-R** | 10.253968 | -20.789633 | 0.2816968 | 0.0587345 |

This table shows the significant differences of group interactions for OTU AH008235.

**Appendix B Continued**

**Denovo1377 by environmental biome**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | -4.750000 | -12.882230 | 3.382230 | 0.4014742 |
| **R-I** | 9.750000 | 1.617770 | 17.882230 | 0.0138727\* |
| **S-I** | -2.678571 | -11.400842 | 6.043699 | 0.8381454 |
| **R-L** | 14.500000 | 6.156508 | 22.843492 | 0.0002701\* |
| **S-L** | 2.071429 | -6.848140 | 10.990997 | 0.9214936 |
| **S-R** | -12.428571 | -21.348140 | -3.509003 | 0.0035445\* |

This table shows the significant differences of group interactions for OTU Denovo1377.

**Denovo491**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | 1.9222222 | -7.949161 | 11.79361 | 0.9515100 |
| **R-I** | 13.3666667 | 3.495284 | 23.23805 | 0.0047030\* |
| **S-I** | 13.8428571 | 3.255248 | 24.43047 | 0.0065493\* |
| **R-L** | 11.4444444 | 1.316619 | 21.57227 | 0.0219658\* |
| **S-L** | 11.9206349 | 1.093534 | 22.74774 | 0.0265301\* |
| **S-R** | 0.4761905 | -10.350910 | 11.30329 | 0.9993781 |

This table shows the significant differences of group interactions for OTU Denovo491.

**Denovo847**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | 7.133333 | -2.504097 | 16.770764 | 0.2066118 |
| **R-I** | 15.800000 | 6.162569 | 25.437431 | 0.0005712\* |
| **S-I** | 2.014286 | -8.322397 | 12.350968 | 0.9514124 |
| **R-L** | 8.666667 | -1.221129 | 18.554462 | 0.1024032 |
| **S-L** | -5.119048 | -15.689545 | 5.451450 | 0.5610222 |
| **S-R** | -13.785714 | -24.356212 | -3.215216 | 0.0067029\* |

This table shows the significant differences of group interactions for OTU Denovo847.

**JN995648**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | 1.8944444 | -6.466016 | 10.254905 | 0.9265260 |
| **R-I** | 12.9500000 | 4.589540 | 21.310460 | 0.0011282\* |
| **S-I** | 0.6642857 | -8.302775 | 9.631346 | 0.9970599 |
| **R-L** | 11.0555556 | 2.477904 | 19.633207 | 0.0074630\* |
| **S-L** | -1.2301587 | -10.400054 | 7.939736 | 0.9831829 |
| **S-R** | -12.2857143 | -21.455609 | -3.115819 | 0.0052082\* |

This table shows the significant differences of group interactions for OTU JN995648**.**

**Appendix C**

**One Way Interaction Tests:**

**One-way analysis of means (not assuming equal variances):**

|  |  |  |
| --- | --- | --- |
| **Alpha Methods** | **F-Value** | **P-Value** |
| **Chao1** | 2.6856 | 0.08004 |
| **Observed\_Otus** | 7.2165 | 0.002569\* |
| **PD\_Whole\_Tree** | 5.812 | 0.006526\* |

Chao1 method shows no significant difference for the species richness level means of the environmental biomes. Observed\_Otus and PD\_Whole\_Tree both show significant differences for the species richness level means of the environmental biomes.

**Pairwise Tests for each alpha diversity method (based on environmental biome):**

**Chao1:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | -18.941630 | -50.045306 | 12.16205 | 0.3651385 |
| **R-I** | 4.077656 | -30.037385 | 38.19270 | 0.9879727 |
| **S-I** | 9.363370 | -22.548328 | 41.27507 | 0.8555398 |
| **R-L** | 23.019286 | -10.341143 | 56.37971 | 0.2602845 |
| **S-L** | 28.305000 | -2.798676 | 59.40868 | 0.0848873 |
| **S-R** | 5.285714 | -28.829327 | 39.40076 | 0.9745510 |

There are no statistically significant interactions for species richness level among the four environmental biomes, using the Chao1 alpha diversity method.

**Observed\_Otus:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | -18.5000000 | -39.886979 | 2.886979 | 0.1090640 |
| **R-I** | 0.2857143 | -23.171888 | 23.743317 | 0.9999867 |
| **S-I** | 14.5555556 | -7.387022 | 36.498133 | 0.2923780 |
| **R-L** | 18.7857143 | -4.153014 | 41.724443 | 0.1394487 |
| **S-L** | 33.0555556 | 11.668576 | 54.442535 | 0.0011570\* |
| **S-R** | 14.2698413 | -9.187761 | 37.727444 | 0.3660773 |

One pairwise test is statistically significant for the Observed\_Otus species richness level. The statistically significant group interaction is Soil and Leaf with a p-value of 0.0011570.

**Appendix C Continued**

**Pairwise Tests for each alpha diversity method continued (based on environmental biome):**

**PD\_Whole\_Tree:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group Interactions** | **diff** | **lwr** | **upr** | **p-adj** |
| **L-I** | -1.1445202 | -2.5102339 | 0.2211934 | 0.1261181 |
| **R-I** | 0.7361235 | -0.7618145 | 2.2340614 | 0.5491210 |
| **S-I** | 0.9998611 | -0.4013315 | 2.4010538 | 0.2339817 |
| **R-L** | 1.8806437 | 0.4158396 | 3.3454478 | 0.0077284\* |
| **S-L** | 2.1443813 | 0.7786677 | 3.5100950 | 0.0009627\* |
| **S-R** | 0.2637376 | -1.2342003 | 1.7616756 | 0.9634198 |

Two pairwise tests are statistically significant for the PD\_Whole\_Tree species richness level. The statistically significant groups are Rhizosphere/Leaf (p-val = 0.0077284) and Soil/Leaf (p-val = 0.0009627).

**Two-Way** **analysis of means (not assuming equal variances) with Location and Environmental Biome:**

**Chao1:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Df** | **Sum Sq.** | **Mean Sq.** | **F-Value** | **Pr(>F)** |
| **Env\_biome** | 3 | 4326 | 1441.9 | 2.084 | 0.130 |
| **Location** | 2 | 1464 | 731.9 | 1.058 | 0.363 |
| **Env\_biome\*Location** | 6 | 1910 | 318.4 | 0.460 | 0.830 |

For Chao1 method, environmental biome, location, and the interaction between them have no significance to differences between species richness level.

**Observed\_Otus:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Df** | **Sum Sq.** | **Mean Sq.** | **F-Value** | **Pr(>F)** |
| **Env\_biome** | 3 | 5253 | 1750.9 | 5.300 | 0.00632\* |
| **Location** | 2 | 744 | 372.1 | 1.126 | 0.34144 |
| **Env\_biome\*Location** | 6 | 776 | 129.4 | 0.392 | 0.87668 |

For Observed\_Otus method, environmental biome is statistically significant while location and the interaction have no significance to the differences between the OTUs.

**PD\_Whole\_Tree:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Df** | **Sum Sq.** | **Mean Sq.** | **F-Value** | **Pr(>F)** |
| **Env\_biome** | 3 | 25.681 | 8.560 | 6.323 | 0.00276\* |
| **Location** | 2 | 2.743 | 1.372 | 1.013 | 0.37870 |
| **Env\_biome\*Location** | 6 | 3.298 | 0.550 | 0.406 | 0.86735 |

For PD\_Whole\_Tree method, environmental biome is statistically significant while location and the interaction between them have no significance to the differences between the calculated branch lengths of the phylogenetic tree applied.

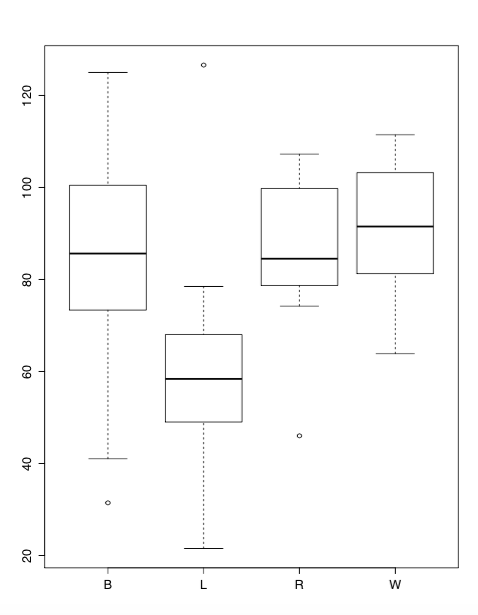
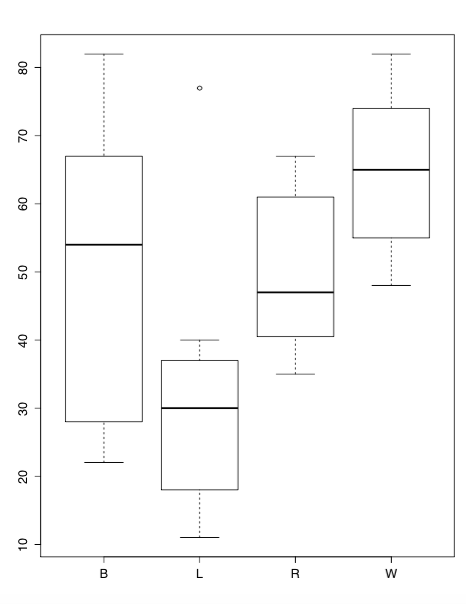
**Conclusion: Location has no contribution to the differences for each method, so the one-way interactions (only environmental-biome) are the focus of analysis.**

**Appendix C Continued**

**Boxplots for each alpha diversity method, based on environmental biome:**

Key: B=Insect, L=Leaf, R=Rhizosphere, W=Soil

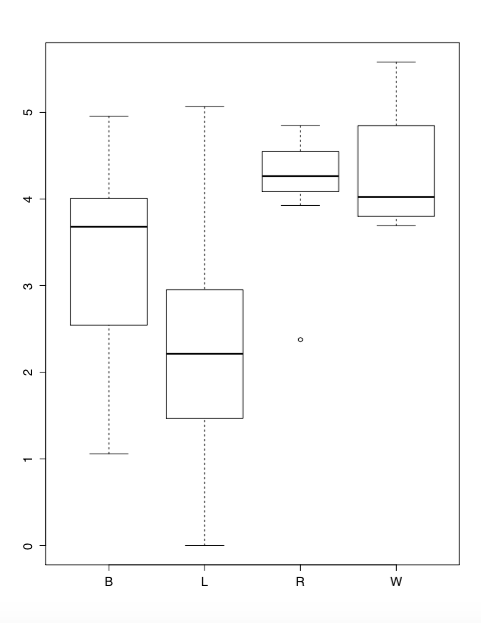
**Chao1: Observed\_Otus:**

** **

Chao1 alpha-diversity method shows overall species richness levels within a sample for each of the four fungal communities.

Observed OTUs alpha-diversity method shows the amount of different OTUs within a sample for each of the four communities.

**PD\_Whole\_Tree:**

****

PD\_Whole\_Tree alpha-diversity method shows the total branch lengths, within a sample, of the phylogenetic tree provided, for each of the four fungal communities.

**Appendix D**

**R Code for Kruskal Wallis Results (Appendix A & B):**

****

**Appendix E**

**R Code for Alpha Diversity Results (Appendix C):**

****

**Appendix F**

**Data File for Kruskal-Wallis Analysis:**

****

|  |
| --- |
| **Key:** **env\_biome:** I=Insect, S=Soil, R=Rhizosphere, L=Leaf  **Values** = Relative abundance of sequences in each significant OTU for each sample. |

**Appendix G**

**Data File for Alpha-Diversity Analysis:**

****

|  |
| --- |
| **Key**: **env\_biome:** B=Leaf, L=Insect, W=Rhizosphere, R=Soil  **Values:** **Chao**: Total species richness within each sample  **Observed\_Otus**: Total number of different OTUs within each sample  **PD\_Whole\_Tree**: Branch length total for phylogenetic tree within each sample. |

**References:**

**1) Jennifer Fouquier, Jai Ram Rideout,, Evan Bolyen, John Chase, Arron Shiffer, Daniel McDonald, Rob Knight, J Gregory Caporaso and Scott T. Kelley. ghost-tree: creating hybrid-gene phylogenetic trees for diversity analyses. 2016.**

**Available from:** [**http://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-016-0153-6**](http://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-016-0153-6)

**2) U.S. National Library of Medicine. National Center for Biotechnology Information.**

**Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi**

**3) Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61:1–10.**

**Available from:** [**http://www.sciencedirect.com/science/article/pii/0006320792912013**](http://www.sciencedirect.com/science/article/pii/0006320792912013)

**4) Werner Lab. QIIME Overview.**

**Available from:** [**http://www.wernerlab.org/teaching/qiime/overview/**](http://www.wernerlab.org/teaching/qiime/overview/)

**Access date: 04/02/16**

**5) Schoch, Seifert, Huhndorf, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a Universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. USA. 2012, 109:6241–6.**

**Available from: http://www.pnas.org/content/109/16/6241.full**

**6) Caporaso G, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6.**

**Available from: http://www.nature.com/nmeth/journal/v7/n5/full/nmeth.f.303.html**