Winter 2012

Effects of nitrogen fertilization on fungal community structure in a temperate hardwood forest: Implicit links between structure, function and resilience

Eric W. Morrison
University of New Hampshire, Durham

EFFECTS OF NITROGEN FERTILIZATION ON FUNGAL COMMUNITY STRUCTURE IN A TEMPERATE HARDWOOD FOREST: IMPLICIT LINKS BETWEEN STRUCTURE, FUNCTION AND RESILIENCE

BY

ERIC W. MORRISON

B.S., University of New Hampshire, 2009

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

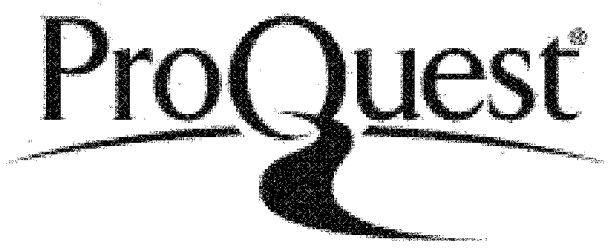
the Requirement for the Degree of

Master of Science

in

Microbiology

December, 2012



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This thesis has been examined and approved.

Thesis Director, Dr. Serita D. Frey
Professor of Soil Microbial Ecology

Dr. W. Kelley Thomas
Professor of Genomics

Dr. Louis Tisa
Professor of Microbiology and Genetics

Sept 24 2012
Date

ACKNOWLEDGEMENTS

This work was supported by the NSF and the NSF LTER program for funding the Harvard Forest LTER site. The Hubbard Center for Genome Studies at UNH provided use of equipment and lab space for which I am grateful. Many people contributed to this project including Jobriah Anderson, Krys Morris, Holly Bik, Linda van Diepen and Mel Knorr. I am grateful for your knowledge and willingness to share it. Thanks you to all of the members of the Frey lab for your friendship throughout this process and for teaching me the skills necessary to perform this work and to be successful as a scientist. I am grateful to my committee: Dr. Tisa, for your guidance through the Masters program and for providing an example of critical thought and dedication to science. Dr. Thomas, thank you for providing thoughtful comments, technical knowledge, and guidance throughout every stage of this research. Dr. Frey, thank you for giving me the opportunity to work on this project and for your confidence in my abilities to this point and beyond. You shared insight and perspective throughout this research and have given me every opportunity to succeed and grow as a scientist, thank you.

Finally, thank you to my friends and family. You were critical in guiding me towards a career in science, and your friendship throughout this process was crucial to my success.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

ABSTRACT

EFFECTS OF NITROGEN FERTILIZATION ON FUNGAL COMMUNITY STRUCTURE IN A TEMPERATE HARDWOOD FOREST: IMPLICIT LINKS BETWEEN STRUCTURE, FUNCTION AND RESILIENCE

by

Eric W. Morrison

University of New Hampshire, December 2012

This study documents soil fungal communities at the Harvard Forest Chronic Nitrogen Addition experiment, which was established in 1989 to test the effects of long-term nitrogen fertilization on ecosystem processes. Researchers at this site have observed an accumulation of soil carbon in the nitrogen fertilized plots and a decrease in fungal biomass, ligninolytic enzyme activity, and rates of litter decay. We hypothesized that decreased decomposition rates in nitrogen-fertilized plots were due to changes in the structure of the fungal community, especially Basidiomycetes, the primary decomposers of lignin in this ecosystem. We performed a marker gene study of fungal communities in the organic soil horizon using 454 high-throughput sequencing of three separate loci. The dominant OTU increased significantly in relative abundance in the highest N treatment. Additionally, Basidiomycete community composition was altered by N additions. These results suggest that changes in fungal community structure may contribute to decreased decomposition rates.

INTRODUCTION

Nitrogen (N) is a limiting nutrient in many ecosystems, and changes in levels of available N have the potential to disrupt essential ecosystem services by changing the microbial and plant communities that provide them. Global rates of N deposition more than doubled from approximately 34 Tg N yr^{-1} in 1860 to 100 Tg N yr^{-1} in 1995, and are predicted to double again by the year 2050 (Galloway *et al.*, 2008). Rates of N deposition in the northeast U:S. are predicted to reach 10-20 kg N ha^{-1} yr^{-1} by the year 2050, approximately doubling current levels (Galloway *et al.*, 2008). Human induced N deposition has been shown to reduce soil organic matter (SOM) decomposition rates in temperate ecosystems by altering the function of soil communities involved in SOM decomposition, though it is unclear whether this is due to changes in community composition or direct effects of N on the activity of soil organisms (Kellner *et al.*, 2010; Frey *et al.*, 2004; Zak *et al.*, 2011).

Soil microorganisms, especially fungi, play a fundamental role in the decomposition of SOM, the largest reservoir of C in the terrestrial biosphere (Lynd *et al.*, 2002, Herman *et al.*, 2008). Soil organic matter contains twice as much C as the atmosphere, and changes in SOM decomposition rates due to human activity will be important in modeling global change stressors such as climate warming (Zak *et al.*, 2011). Allison and Martiny (2008) describe a

paradigm in which microbial communities may be resistant to disturbance (community composition is unaffected), resilient to disturbance (community composition returns to a base level after disturbance), may be functionally redundant (community composition is altered by disturbance, but the new community is functionally similar to the previous community), or may be altered in function and composition by a disturbance. Here we describe compositional changes in soil fungal communities due to long-term N additions and hypothesize as to resiliency and changes in functionality of the fungal community.

Soil fungi fulfill key functional roles in the decomposition process. The most recalcitrant compound in leaf litter, lignin, is primarily broken down by the Basidiomycota. Some Ascomycetes have lignocellulolytic capabilities, though they are generally less efficient than Basidiomycetes (Cullen and Kersten, 2004; Kellner *et al.*, 2010; Osono, 2007). Lignin encases the more labile components of leaf litter, such as cellulose and hemicellulose, preventing access to these C sources until lignin is broken down (Herman *et al.* 2008; Melillo *et al.* 1982). Disturbance of the primary lignolytic community may cause a compositional shift in communities of downstream decomposers, as access to lignin-protected substrates is lost. Alternatively, Zak *et al.*, (2011) have suggested that fitness reduction in the saprobic Basidiomycete community due to reduction in lignolytic activity and decreased availability of cellulose and hemicellulose may result in compositional shifts resulting from less-efficient lignocellulolytic organisms filling unoccupied niche space.

A high N content in leaf litter, as well as high levels of available N in soils are thought to inhibit the activity of lignolytic fungi (Frey *et al.*, 2004; Fog, 1988; Zak *et al.,* 2011). *In vitro* studies of lignolytic gene expression have shown that expression of phenol oxidase, lignin peroxidase, and Mn-peroxidase are down regulated by excess N (Boominathan *et al.*, 1990; Vanderwoude *et al.* 1993; Worrall *et al.*, 1997). Peroxidase enzymes are responsible for the initial attack on the complex lignin molecule, breaking it down into its constituent subunits, which can then be attacked by phenol oxidases and other enzymes (Bugg *et al.*, 2011). In contrast to *in vitro* studies, Kellner *et al.* (2010) found no difference in the presence or absence of lignolytic transcripts in an N fertilized forest soil. However, lignolytic enzymes may be post-transcriptionally regulated or differentially regulated by N. It is unclear whether direct N inhibition of lignolytic gene expression is responsible for decreased decomposition rates observed in N fertilized environments, or whether decreased decomposition rates may be due to changes in community composition through effects of N on inter-species competition, for example (Baar and Stanton, 2000; Kellner *et al.* 2010; Zak *et al.*, 2011).

The Chronic Nitrogen Addition (CNA) Study at the Harvard Forest Long-Term Ecological Research (LTER) site in central Massachusetts has been fertilized with ammonium nitrate since 1989 to form an N deposition gradient with ambient N deposition (control), an intermediate level of N addition representing levels expected in some regions of the world by the year 2050 (low N), and high levels of N addition representing chronic N saturation of the ecosystem (high N)

(Aber and Magill 2004; Aber *et al.* 1989; Galloway *et al.*, 2008). Researchers at Harvard Forest have shown that the decomposition of leaf litter is slowed by N fertilization (Magill and Aber 1998; Micks *et al.*, 2004) and that SOM levels and the relative abundance of lignin have significantly increased in the N fertilized plots (Frey *et al.,* in prep). Fungal biomass, the fungal:bacterial biomass ratio, and the activity of phenol oxidase, have also decreased with increasing N fertilization (Frey *et al.*, 2004). These observations suggest that the microbial decomposer community in the N fertilized plots, especially the fungal component of the community, has been adversely impacted by long-term N fertilization. Our objectives in this study were to test the effects of N fertilization on fungal community composition through changes in relative abundance of specific taxa, and through phylogenetic and OTU based community diversity approaches.

MATERIALS AND METHODS

Study Site, Sample Collection, and Processing

Soil samples were collected from the Chronic Nitrogen Addition Study (CNA) at the Harvard Forest Long-Term Ecological Research (LTER) site in Petersham, Massachusetts in November 2009. Treatment plots were established in 1988 and consist of a mixed hardwood stand composed primarily of black and red oak (*Quercus velutina* Lam.; *Q. borealis* Michx. f.) with Typic Dystrudepts soils of the Gloucester series (Peterjohn et al. 1994). Soils are treated with nitrogen additions in the form of an aqueous ammonium-nitrate solution applied monthly during the growing season (Aber *et al.*, 1989). Three 30 x 30 m plots receive one of three N treatments: ambient N deposition (control), 50 kg N ha^{-1} yr^{-1} (low N), or 150 kg N ha^{-1} yr^{-1} (high N) (Aber *et al.*, 1993). Each treatment plot is divided into thirty-six 5 m^2 subplots with the outer subplots excluded from sampling to account for edge effects.

Four soil samples (2.5 cm diameter) were collected from the O-horizon (O_a and O_e layers) from each of three randomly chosen subplots within each N treatment plot and then combined for one composite sample per subplot (resulting in three replicates per treatment megaplot). Samples were transported to the University of New Hampshire on ice, passed through a 2 mm sieve sterilized with ethanol, and fine roots were removed. Samples were flash frozen in liquid N and stored at -80°C within 6 hours of sampling.

Loci and Primers

The rDNA internal transcribed spacer region (ITS) is the most widely used taxonomic marker for fungi and has been proposed as the fungal marker gene for the Barcode of Life project (Seifert, 2009). The ITS segments of rDNA, ITS1 and ITS2, are approximately 300 base pairs and 450 base pairs long, respectively, allowing one entire region, but not both at once, to be sequenced using Roche 454 Titanium sequencing technology due to limitations in read length at the time of this writing. These introns are highly variable, allowing for genus or species level identification of fungi (Bruns and Shefferson, 2004, Horton and Bruns, 2001). However, the amount of intraspecific ITS variation varies across the kingdom Fungi causing uncertainty in taxonomic identification and community structure analyses (e.g. operational taxonomic unit (OTU) based α and β diversity estimates) using standard barcoding techniques (Nilsson *et al.*, 2008). Use of both of these regions as statistical replicates for taxonomic identification and community structure analyses may offer the power to resolve uncertainty that arises when using one of these regions alone.

Use of phylogenetically informative tests of community structure may allow inference of function given the assumption that clade specific functional traits can be inferred from phylogenetic distance (Faith *et al.*, 2009). Additionally, Fierer *et al.* (2012) showed that there was high correlation between phylogenetic estimates of community composition, estimates of potential community function from metagenomic assays, and estimates of community function from catabolic profiling of soils under long-term N enrichment. Communities performing

essential ecosystem services, such as decomposition, that have phylogenetically distant members may therefore have functional differences that are significant on an ecosystem scale. The variability of ITS sequences, along with the range of variability among different taxa, does not allow for well-supported phylogenetic inference (Amend *et al.,* 2010; Koljalg *et al.* 2005). Large sub-unit (LSU) rDNA sequences are more conserved than ITS sequences and the variable D1-D3 regions of the LSU allows reasonable phylogentic inference across the fungi (Amend *et al.,* 2010; Edwards and Zak, 2010; Koljalg *et al.* 2005). We chose to use the LSU D2-D3 region to test the phylogenetic diversity of the potential lignin-decomposing fungal community (Basidiomycota) as a proxy for potential differences in community functionality.

DNA Extraction and Sequencing

DNA extraction was performed at the UNH Hubbard Center for Genome Studies. Extractions were performed on 0.75 g of soil with the MoBio RNA PowerSoil® Total RNA Isolation Kit and MoBio RNA PowerSoil® DNA Elution Accessory Kit. An additional purification step was performed with the MoBio PowerClean® DNA Clean-Up Kit. PCR primers targeting the ITS1 and ITS2 region of fungal ribosomal DNA (rDNA) and the large-subunit (LSU) D2-D3 region of Basidiomycete rDNA were used for amplification of fungal taxonomic marker genes. Primers ITS1f (Gardes and Bruns, 1993) and 5.8s (Vilgalys and Hester, 1990) were used to target the ITS1 region. Primers ITS4 (White *et al.,* 1990) and 5.8sr (Vilgalys and Hester, 1990) were used to target the ITS2 region.

Primers LR21r (Hopple and Vilgalys, 1993) and LR5f (Tedersoo *et al.*, 2008) were used to target the LSU D2-D3 region of Basidiomycetes. Primer sets were constructed for each locus by combining locus specific primers with nucleotide multiplex identifiers (MID) and the Roche 454 Titanium FLX A or B primers. The PCR reaction was optimized for annealing temperature and DMSO concentration using an amplifiable soil DNA extract as a positive control. All other reagents were used according to manufacturer's recommendations. PCR was performed on 1 µL of soil DNA extract, or on 1 µL water as a negative control, using final concentrations or amounts of the following reagents in a 25 µL reaction: 1 mM $MgCl_2$, 200 µM of each deoxynucleotide triphosphate, 2% DMSO, 2.5 µl 10x Phusion buffer, 0.5 units Phusion New England Biolabs polymerase and 0.25 nM each of the forward and reverse primers. The PCR reaction conditions were as follows: a thirty second initial denaturation step at 98°C, 27 cycles of denaturation at 98°C for ten seconds, annealing at 62°C for 10 seconds, and extension at 72°C for eight seconds, then a final extension step at 72°C for five minutes. Electrophoresis was performed on a 1.5% agarose gel to confirm fragment size. PCR products were analyzed on a Thermo Scientific NanoDrop 1000 spectrophotometer for absorbance spectra and DNA concentration. PCR products were purified using the solid-phase reversible immobilization purification method to remove fragments below 300 base pairs in length (DeAngelis *et al.*, 1995). Concentrations were measured by spectrophotometry and samples pooled at equimolar concentration. A final purification of the sample library was performed with a QIAGEN Genomic-tip purification column. Massively parallel

sequencing was performed at the University of Illinois Roy J. Carver Biotechnology Center on the pooled amplicons using Roche 454 GS FLX Titanium sequencing technology (Margulies *et al.*, 2005).

Sequence Processing

Sequences were filtered for sequence quality and assigned to samples by MID tags using AmpliconNoise (Quince *et al.*, 2011) as implemented in the QIIME analysis pipeline. AmpliconNoise filters sequences for sequencing noise and PCR errors based on an expectation-maximization algorithm (e.g. 'denoising') and filters sequences based on standard quality measures for 454 sequencing data (e.g. sequence length, maximum homopolymer length, primer mismatches). The ITS1 and ITS2 sequences were clustered at 95% sequence similarity and LSU sequences were clustered at 100% sequence similarity with UCLUST as implemented in QIIME (Edgar, 2010; Caporaso *et al.*, 2010). Chimera removal was performed using the USEARCH program with the UCHIME *de novo* method and in-house perl scripts (Edgar 2010).

Taxonomic Assignment and Community Structure

The following analyses were performed in QIIME using default parameters unless otherwise noted. Taxonomy was assigned to ITS1 and ITS2 OTUs by local BLAST against the NCBI non-redundant nucleotide database with BLAST results parsed with portions of the OCTUPUS software package (Sung W, in prep) and in-house perl scripts. ITS1 and ITS2 OTU tables of relative

abundances were rarefied to equal sequencing depths within each locus then combined. OTU relative abundance was summed by family for analysis of changes in relative abundance by treatment. In-house perl scripts and R 2.14.1 were used to calculate ANOVA and Tukey's HSD for changes in taxon relative abundance by N treatment. Frey *et al.*, (in prep) found that fungal biomass decreased in the low N and high N treatments. We normalized our relative abundance estimates using these biomass data to provide an estimate of "absolute" abundance and reran statistical analyses for significant changes in abundance.

Representative sequences from 100% sequence similarity clustering of LSU sequences were aligned and a neighbor-joining tree constructed in clustalw 2.0.3 (Larkin *et al.*, 2007). All OTU tables were rarefied for α diversity and β diversity estimates. Rarefaction curves, ACE richness estimates, Shannon diversity and evenness were computed. The PD whole tree phylogenetic α diversity metric was computed for the LSU dataset (Faith and Baker, 2006). Abundance based Bray-Curtis and presence-absence Sørensen-Dice similarity matrices were calculated using ITS1 and ITS2 OTU tables. Unweighted and weighted Unifrac similarity matrices were calculated using LSU 100% similarity OTU tables. Relative abundance for all OTUs was calculated and averaged across samples. OTUs were divided into OTUs that made up less than 2% of relative abundance on average, Ascomycete OTUs, Basidiomycete OTUs, Basidiomycete OTUs that made up greater than 2% of relative abundance and those that made up less than 2% of relative abundance on average for analysis

of the rare, Ascomycete, Basidiomycete, dominant Basidiomycete and rare Basidiomycete communities. Beta and α diversity estimates were recalculated for all communities. Principal coordinates analysis of various distance matrices was used to visualize distance between samples. Significant differences in community composition across N treatments were determined by one thousand Monte-Carlo simulations of distance matrices and ANOSIM with significant differences less than $P = 0.05$. Correlation between all ITS1 and ITS2 β diversity distance matrices was determined by a Mantel test with 1000 permutations.

The dominant OTU in both ITS datasets was identified as an "environmental sequence." In order to more accurately/informatively identify these dominant OTUs we extracted ITS1 and ITS2 regions from representative sequences from putative *Russula* OTUs from ITS1 and ITS2 datasets and a representative *R. atropurea* sequence using the Fungal ITS1 Extractor (Nilsson *et al.,* 2010). ITS1 and ITS2 sequences were aligned separately with MUSCLE (Edgar 2004). A maximum parsimony tree was built from each alignment in MEGA 5.05 using the max-mini branch bound search method with ten initial trees (Tamura *et al.,* 2011).

RESULTS

Amplicon Sequencing

We obtained 15,728 sequences from the ITS1 region, 8,060 sequences from the ITS2 region, and 11,681 sequences from the LSU region after denoising, quality control, and removal of chimeras and likely non-fungal sequences as determined by BLAST against NCBI. ITS1 sequences were rarified to 800 sequences per sample, and ITS2 sequences were rarefied to 530 sequences per sample to account for effects of unequal sampling depth. Rarefaction curves for ITS1 and ITS2 samples are presented in Figures 1 and 2. This resulted in one sample from each of the low and high N treatments being excluded from β diversity analyses of ITS2 sequences due to poor sequence coverage in those samples. Though we obtained as many as 3000 LSU sequences for one sample from the high N treatment, the three replicate control samples had 274 to 672 LSU sequences. Due to poor sequencing depth across control samples, LSU sequences were rarefied to 270 sequences per sample. However, rarefaction curves indicate that this level of sampling captured the majority of OTU richness (Figure 3).

Taxonomy and Diversity

Members of the Basidiomycota dominated the communities in all N treatments ranging from nearly 54% of relative abundance in the control plots to 66-67% of relative abundance in the low and high N communities

(Figure 4). The Ascomycota ranged from 14% to 22% of relative abundance across the treatments (Figure 4). Complete lists of all families of fungi with their relative abundances are given in Appendix B. All communities were dominated by the *Russula* genus, with one OTU being dominant (9.71% in control, 29.84% in low N, 37.91% in high N; Figure 5). This OTU (ITS1 OTU 164; ITS2 OTU 301) significantly increased in relative abundance by 28.2% from control to high N (Figure 5a). All other *Russula* OTUs combined decreased in relative abundance from 29.03% in control communities to less than 15.4% in low N and high N communities though this trend was not significant. A maximum parsimony tree of ITS1 *Russula* OTUs showed high similarity between the dominant OTU (OTU 164), which was identified as an "environmental sequence" by our taxonomic identification method, and a representative *R. atropurpurea* sequence, with all other *Russula* OTUs forming separate clades (Figure 5b).

We assessed broad changes in taxon relative abundance at the family level because genus level sequence identification may be inaccurate (Vilgalys *et al.,* 2003). Though the *Russula* genus was dominant in all communities, there were no significant changes in the relative abundance of the Russulaceae family across N treatments. Twelve families of fungi underwent significant changes in relative abundance ($P < 0.05$) (Table 1). Interestingly, all of these families increased in relative abundance in either low or high N communities. However, few of these families made up more than 2% of relative abundance in any community. Notably, the Hypocreaceae increased from less than 0.7% relative abundance in the control and low N treatments to 1.85% in the high N plots

(Table 1). The Mortierellaceae increased from control (1.03%) to high N (2.75%) (Table 1). The Sclerodermataceae were not detected in the control community and increased from 0.17% in the low N community to 2.60% in the high N community. Additionally, members of the Sebacinaceae and Bolecetaceae underwent marginally significant increases in relative abundance in low N (P = 0.1). The Sebacinaceae represented less than 0.3% of relative abundance in control and high N plots and 1.52% of relative abundance in the low N community. The Bolecetaceae represented less than 0.08% of relative abundance in control and high N communities and 0.96% of relative abundance in the low N community.

Baldrian *et al.* (2012) demonstrated that rare members of the soil fungal DNA pool may be highly active and contribute significantly to community function. We therefore separated our ITS communities into dominant members (>2% relative abundance on average), which were composed of ectomycorrhizal Basidiomycetes, and rare members, representing <2% total relative abundance on average. Additionally, because Basidiomycetes are thought to be more prominently involved in lignin degradation than Ascomycetes, we separated our communities into Ascomycete, and rare Basidiomycete communities.

The ACE richness estimator performed on ITS1 OTUs indicated higher species richness in the high N community than either control or low N communities though this trend was only marginally significant (P = 0.073), while ACE richness estimates from ITS2 OTUs indicated significantly higher richness in the high N community (P = 0.032). Additionally, rarefaction curves of singleton

OTUs from both ITS1 and ITS2 indicated a greater number of singletons in the high N community than the control community (ITS1 $P = 0.04$; ITS2 $P = 0.002$; Figure 6). There were no differences in OTU based α diversity estimates for the Basidiomycete community; however, the ITS1 locus indicated a marginally significant increase in singleton Ascomycete OTUs in the high N community as compared to the control community ($P = 0.067$), and all metrics trended towards increased Ascomycete diversity in N enriched communities. The ITS2 locus showed a similar trend in increased Ascomycete richness in N enriched communities. Significant differences between control and high N communities were observed in ITS2 Shannon diversity ($P = 0.014$), and the high N community had significantly higher numbers of observed OTUs and singleton OTUs than the control and low N communities ($P < 0.04$). Additionally, a phylogenetic estimate of α diversity (PD whole tree) performed on the LSU region indicated increased phylogentic diversity in the low N community (PD = 1.54) and decreased diversity in the high N community (PD = 1.24) compared to the control community (PD = 1.44), though this trend was not significant.

Community composition

Mantel tests showed high correlation between ITS1 and ITS2 β diversity estimates for all distance matrices accept for Ascomycete member abundance (Table 3). ITS1 results are reported here. Community composition of the total community under high N additions as determined by community membership and member abundance significantly differed from that in control and low N plots

based on OTU based diversity metrics (Figure 7a). Taxa biplots against both β-diversity estimates indicated the Pleuteaceae were associated with the control treatment, the Thelephoraceae with control and low N communities, Tricholomataceae with the low N treatment, and the Sclerodermataceae and Mortierellaceae were associated with the high N treatment (Table 2). ANOSIM showed significant dissimilarity in both membership and abundance (Table 3).

The Ascomycete community under high N additions tended to be different from that in control and low N communities, and as a whole, was moderately structured by N (Figure 7b, Table 3). Ascomycete families that were associated with the high N community included saprobes and fungal or plant parasites (Table 2). Families associated with the control and low N communities included ectomycorrhizal fungi and saprobes, notably the Myxotrichaceae which contains many cellulolytic members.

Phylogentic and OTU based analysis of Basidiomycete community membership showed significant differences between communities by N treatment (Figure 7c and 7d). The dominant Basidiomycete community which was composed primarily of the ectomycorrhizal Russulaceae and Cortinariaceae was more strongly structured by N than the rare Basidiomycete community (Table 3). Nitrogen additions tended to have more of an effect on Basidiomycetes than Ascomycetes (Table 3). Though the LSU D2-D3 region is not often used for taxonomic identification of fungi, there was broad agreement between taxa biplots against phylogentic community membership and OTU based community membership. The control and low N communities tended to share many

ectomycorrhizal/saprobic families, excluding the Pleutiaceae, which were primarily associated with the control community. Additionally, several ectomycorrhizal taxa were associated exclusively with the low N community while the Exidiaceae, which are saprobic or parasitic, and Sclerodermataceae, which are primarily ectomycorrhizal, were associated with the high N community (Table 3). In all cases, N fertilization resulted in a gradient effect, where high levels of N had the greatest effect on community composition, and low levels of N had an intermediate effect or no significant effect. Control and low N communities tended to cluster together except for in analyses of the Basidiomycete community where there were significant differences between all N amendment regimes (Figure 7).

DISCUSSION

Overview

Simulated N deposition in the form of applications of N fertilizer to soils have been shown to decrease SOM decomposition rates and promote lignin accumulation (Frey *et al.*, 2004; Magill and Aber, 1998), as well as to reduce phenol oxidase (DeForest *et al.*, 2004; Frey *et al.*, 2004) and peroxidase enzyme activity (DeForest *et al.*, 2004), and reduce fungal biomass and fungal:bacterial biomass ratios (Frey *et al.*, 2004; Frey *et al.*, in prep) in temperate forest ecosystems. Additionally, N has been shown to affect the community structure of bacteria in soils, favoring copiotrophic bacteria under N fertilized conditions and oligotrophic bacteria under ambient N deposition conditions (Fierer *et al.*, 2012). It has been hypothesized that changes in the decomposition process are due to changes in the structure of the fungal decomposer community in soils or due to changes in the activity of this community (e.g. community function) (Frey *et al.*, 2004; Zak *et al.*, 2011). As Fierer *et al.* (2012) note it is not possible to determine from long-term environmental manipulations whether community compositional changes due to a treatment effect are driving community functional changes, or whether direct effects of a treatment on community function determine changes in community composition. However, assessment of structural changes in the community can provide insight into how changes in community composition contribute to changes in community function and changes in ecosystem processes such as decomposition (e.g. are microbial communities resilient to

disturbance, functionally redundant, etc.). Our primary objectives were 1) to test the effects of N additions on fungal community diversity, 2) to test whether the relative abundance of specific abundant taxa was affected by N treatments, and to determine whether any of these taxa might be implicated in contributing to decreased decomposition rates, and 3) to test whether fungal community composition, specifically the Basidiomycete community, was affected by N additions.

The most dominant OTU in the forest floor at CNA, putatively identified as a close relative of *Russula atropurpurea*, increased in relative abundance from the control to high N treatment. When relative abundance was normalized by total fungal biomass in the Oa and Oe soil horizons at CNA this OTU was unaffected by N additions while all other members of the Russulaceae decreased in the low N and high N treatments. We found an increase in OTU richness in the high N treatment at CNA, which was primarily driven by an increase in Ascomycete OTU richness. Ascomycete OTUs associated with the high N treatment tended to be from families primarily composed of fungal or plant parasites and saprobes. We observed a gradient response of both OTU based and phylogenetic Basidiomycete community composition to low and high levels of N addition, with all treatments having different Basidiomycete communities and the control treatment being more variable than low N and high N treatments. The dominant taxa, all ectomycorrhizal Basidiomycetes (*Russula* and *Cortinarius* spp.) tended to be phylogentically structured by high levels of N addition, with no significant differences between the control and low N treatments due to high

variability in phylogentic community composition in the low N treatment. Both low and high levels of N additions changed the phylogenetic community composition of the rare Basidiomycete community. The Ascomycete community was structured by high levels of N, in agreement with the increase in OTU richness observed in the high N treatment.

Composition of *Russula* Species

The response of the dominant *Russula* OTU suggests that this organism is either nitrophillic or unaffected by N additions. One of the primary roles of EMF is to transport N to hosts in exchange for plant C, and EMF are often excluded from symbiosis under conditions of high N availability (Hobbie and Hobbie, 2008). The increase in *Russula atropurpurea* relative abundance suggests that this species is either able to survive with lower levels of host C inputs, is transferring other benefits to host plants and is therefore maintained in symbiosis, or is better able to forage for its own C resources. The decline in absolute abundance of other *Russula* species suggests that these EMF are unable to withstand the effects of N additions on host plant symbioses. There is a growing consensus that EMF act as saprobes in some situations (Courty *et al.*, 2010; Cullings and Courty, 2009; Talbot *et al.*, 2009), though there is much debate as to the magnitude of their role in decomposition (Baldrian 2009). Reduced abundance of the EMF community may be contributing to decreased decomposition rates under N additions at CNA. Additionally, EMF interactions with the saprobic community may contribute to saprobic community activity,

primarily through competition for resources (Baar and Stanton, 2000; Osono 2007). Interactions between lignolytic fungi have been shown to increase production of lignolytic enzymes (White and Boddy, 1992; Savoie *et al.*, 1998; Tsujiyama and Minami, 2005). Saprobic fungi-EMF interactions may have a similar effect, however more research is needed to determine interspecies or inter-functional group interactions in the field.

Increased OTU Richness Under High N

The increased taxonomic richness observed in the high N community suggests that the resource niche left available by changes in the Basidiomycete lignin decomposing community has allowed several less specialized taxa to take advantage of this niche. Alternative hypotheses are that decreased abundance of dominant taxa has allowed detection of the more rare community, or that N is acting as a mutagen, such that ITS sequences are more variable under high levels of N, arbitrarily inflating OTU richness at our sequence similarity clustering threshold. The mutagen hypothesis seems unlikely given that the trend for increased richness is Ascomycete specific in ITS sequences, and does not translate to the Basidiomycete LSU, although we acknowledge that there would be greater selection against mutations in the LSU region. The hypothesis that the increased richness we observed is an effect of sequencing depth also seems unlikely given that the most dominant OTU in our dataset increases in relative abundance in the high N treatment, and that fungal biomass is significantly decreased in both the low N and high N treatments (Frey *et al.*, 2004, Frey *et al.*,

in prep), suggesting that if the trend in richness was an effect of sequencing depth versus fungal biomass, we would have observed increased richness in both the low N and high N communities.

One explanation for the increase in Ascomycete richness is that Ascomycetous saprobes such as Dermataceae species, which can partially decompose lignin and lignin by-products (Kellner *et al.*, 2010; Osono, 2007), are taking advantage of niche space left available by the altered lignolytic Basidiomycete community, but that no taxa are able to gain a competitive advantage allowing many taxa to coexist within the same niche space. Zak *et al.* (2011) suggested that N limitation of the ability of saprobic Basidiomycetes to decompose lignin may limit access to more labile substrates encased in lignin, allowing other lignocellulolytic organisms to gain a competitive advantage. Our data support this hypothesis at least under high levels of N addition. Alternatively, interactions with the dominant EMF community, which is altered in the high N treatment but less so in the low N treatment, may be driving the changes observed in the high N Ascomycete community. Finally, many of the families of Ascomycetes associated with the high N treatment are potential parasites of other fungi. Pathogenic Ascomycetes may be taking advantage of reduced fitness of EMF and saprobic Basidiomycete communities resulting from lack of resource (C) availability.

Community Composition

Lozupone *et al.* (2007) suggest that differences in community membership are likely driven by a strong selective pressure from the environment. Our data showed that N acts as a strong selective pressure on membership of soil fungal communities, and Basidiomycete communities in particular. One of the major changes we observed with N additions was a shift in the phylogenetic community composition of the dominant EMF. *Russula* and *Cortinarius* species were dominant in the control and low N treatments, but phylogenetically different *Russula* OTUs were present in the high N community as compared to the control and low N communities, and dominant *Cortinarius* OTUs were absent in the high N community.

Phylogenetic community membership of rare Basidiomycetes was structured by both low and high levels of N. This change in community composition correlates with decreased decomposition rates observed in these communities (Frey *et al.*, in prep). One hypothesis for the commonly observed decrease in lignin decomposition under N enrichment is that N causes a down-regulation of lignolytic transcript production (Edwards *et al.*, 2011; Zak *et al.*, 2011). Kellner *et al.* (2010) suggested that fungal genes involved in decomposition are not completely turned on or off by N addition, and therefore decreased decomposition rates observed under N additions must be due to changes in community structure, or may be a result of down-regulation of transcripts or post-transcriptional down-regulation of decomposition enzymes. Lignolytic transcript down-regulation may cause fewer resources to be available

to downstream decomposers, as lignin encases more labile energy sources such as cellulose and hemi-cellulose and lack of lignin breakdown would reduce availability of these substrates. Reduced resource availability may cause the change in community composition we observed. Alternatively, increased N availability may be directly favoring a subset of the community in competition for more labile resources, thereby altering community composition and causing exclusion of the lignolytic Basidiomycete community. It is unclear which of these hypotheses is the case at this time.

Our data showed that Ascomycete community membership was structured by N. Kellner *et al.* (2010) suggested that Ascomycetes contribute more to lignin break-down than is traditionally believed; however, while Ascomycetes have been shown to produce laccase and aromatic oxidizing enzymes that are involved in intermediate stages of lignin decay, they have not been shown to produce the peroxidase enzymes that are essential to the initial breakdown of lignin, and they are generally less efficient lignin degraders than lignolytic Basidiomycetes (Kellner *et al.*, 2010; Osono, 2007). It is possible that Ascomycetous saprobes have taken advantage of new niche breadth available in the high N treatment, and to a lesser extent the low N plots, and may be taking advantage of secondary products of lignin breakdown that are available from a less active, less competitive Basidiomycete community. Nitrogen additions had a greater effect on Basidiomycete community composition than Ascomycete community composition, suggesting that the Basidiomycete community may be under direct selective pressure, and the Ascomycete community may be

responding to changes in the Basidiomycete community. The Ascomycete saprobes and fungal/plant parasites that drove the observed changes in Ascomycete community composition under high levels of N addition, together with changes in the Basidiomycete community (e.g. both dominant EMF and rare community) suggest that the high N community represents a functionally different community under long-term N saturation that may not be resilient to N disturbance. The low N community represents a community where the dominant members functionality has not changed, but where N seems to be structuring the Basidiomycete community to the point where there is an effect on the decomposition process.

Conclusion

Levels of N addition comparable to those expected by the year 2050 have caused a change in the phylogenetic structure of the rare Basidiomycete community whose members drive the decomposition process. These compositional changes are concomitant with decreased decomposition rates and lignin accumulation observed under low levels of N addition. Faith *et al.* (2009) suggest that phylogentic diversity may be a reasonable proxy for the functional capacity of a community, and Fierer *et al.* (2012) confirm that phylogentic community structure, functional capacity, and catabolic profiles of soil bacterial communities are highly correlated. This work confirms that the phylogenetic community structure of the lignolytic Basidiomycete community may be a predictor for changes in community functionality in the decomposition process in

soils, though it is unclear whether community structure causes these changes in functionality, or whether changes in community structure are due to of the effects of N on resource availability. Additionally, this work supports the idea that N additions have caused a shift to a community that is not functionally similar to the community under ambient N deposition (e.g. communities are not functionally redundant). High levels of N addition, representative of long-term N saturation, have resulted in a restructuring of the dominant community, composed of ectomycorrhizal Basidiomycetes, and the rare community, composed of saprobic and ectomycorrhizal Basidiomycetes, Ascomycetes and unidentified taxa. Saprobic Ascomycetes may be taking advantage of niche space left available by changes in the saprobic Basidiomycete community. Long-term N saturation represented by the high N community is likely to result in permanent alteration of community composition and function, as the dominant community members have shifted and new functionally different community members have populated niche space.

REFERENCES

Aber JD, Magill AH. (2004). Chronic nitrogen additions at the Harvard Forest (USA): the first 15 years of a nitrogen saturation experiment. *For Ecol Manage* **196:** 1–5.

Aber JD, Magill A, Boone R, Melillo JM, Steudler P. (1993). Plant and soil responses to chronic nitrogen additions at the Harvard Forest, Massachusetts. *Ecol Appl* **3:** 156-166.

Aber JD, Nadelhoffer KJ, Steudler P, Melillo JM. (1989). Nitrogen saturation in northern forest ecosystems. *Bioscience* **39:** 378-386.

Allison SD, Martiny JBH. (2008). Resistance, resilience, and redundancy in microbial communities. *Proc Natl Acad Sci U S A* **105:** 11512-11519.

Amend AS, Seifert KA, Samson R, Bruns TD. (2010). Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proc Natl Acad Sci U S A* **107:** 13748–13753.

Baar J, Stanton NL. (2000). Ectomycorrhizal fungi challenged by saprotrophic basidiomycetes and soil microfungi under different ammonium regimes in vitro. *Mycol Res* **104:** 691–697.

Baldrian P. (2009). Ectomycorrhizal fungi and their enzymes in soils: is there enough evidence for their role as facultative soil saprotrophs? *Oecologia* **161:** 657–660.

Baldrian P, Kolarik M, Stursova M, Kopecky J, Valaskova V, Vetrovsky T, *et al.* (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J* **6:** 248-258.

Boominathan K, Dass SB, Randall TA, Reddy CA, Lansing E. (1990). Nitrogen-deregulated mutants of *Phanerochaete chrysosporium* – a lignin-degrading basidiomycete. *Arch Microbiol* **153:** 521–527.

Bruns TD, Shefferson RP. (2004). Evolutionary studies of ectomycorrhizal fungi: recent advances and future directions. *Can J Bot* **82:** 1122-1132.

Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. (2011). Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep* **28:** 1883-1896.

Cannon PF, Kirk PM. (2007). Fungal Families of the World. CAB International: Cambridge, MA.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* doi:10.1038/nmeth.f.303.

Cullen D, Kersten PJ. (2004). Enzymology and molecular biology of lignin degradation. In: Brambl R, Marzluf GA, (eds). The Mycota III: Biochemistry and Molecular Biology, 2nd edition. Springer-Verlag: Berlin-Heidelberg, pp 249-273.

Cullings K, Courty P. (2009). Saprotrophic capabilities as functional traits to study functional diversity and resilience of ectomycorrhizal community. *Oecologia* **161:** 661–664.

Courty P, Buée M, Diedhiou AG, Frey-Klett P, Tacon FL, Rineau F, *et al.* (2010). The role of ectomycorrhizal communities in forest ecosystem processes: new perspectives and emerging concepts. *Soil Biol Biochem* **42:** 679-698.

DeAngelis MM, Wang DG, Hawkins TL. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Res* **23:** 4742-4743.

Deforest JL, Zak DR, Pregitzer KS, Burton AJ. (2004). Atmospheric nitrate deposition, microbial community composition, and enzyme activity in northern hardwood forests. *Soil Sci Soc Am J* **68:** 132–138.

Edgar RC, Drive RM, Valley M. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Edwards IP, Zak DR. (2010). Phylogenetic similarity and structure of Agaricomycotina communities across a forested landscape. *Mol Ecol* **19:** 1469-1482.

Edwards IP, Zak DR, Kellner H, Eisenlord SD, Pregitzer KS. (2011). Simulated atmospheric N deposition alters fungal community composition and suppresses ligninolytic gene expression in a northern hardwood forest. *PloS One* **6:** e20421.

Faith DP, Baker AM. (2006). Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol Bioinf Online* **2:** 121–128.

Faith DP, Lozupone CA, Nipperess D, Knight R. (2009). The cladistic basis for the phylogenetic diversity (PD) measure links evolutionary features to

environmental gradients and supports broad applications of microbial ecology's "phylogenetic beta diversity" framework. *Int J Mol Sci* **10:** 4723–4741.

Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R. (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* **6:** 1007–17.

Fog K. (1988). The effect of added nitrogen on the rate of decomposition of organic matter. *Biological Reviews* **63:** 433-462.

Frey SD, Knorr M, Parrent JL, Simpson RT. (2004). Chronic nitrogen enrichment affects the structure and function of the soil microbial community in temperate hardwood and pine forests. *For Ecol Manage* **196:** 159–171.

Galloway JN, Townsend AR, Erisman JW, Bekunda M, Cai Z, Freney JR, *et al.* (2008). Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320:** 889–892.

Gardes M, Bruns TD. (1993). ITS primers with enhanced specificity for basidiomycetes. *Mol Ecol* **2:** 113-118.

Herman J, Moorhead D, Berg B. (2008). The relationship between rates of lignin and cellulose decay in aboveground forest litter. *Soil Biol Biochem* **40:** 2620–2626.

Hobbie EA, Hobbie JE. (2008). Natural abundance of ^{15}N in nitrogen-limited forests and tundra can estimate nitrogen cycling through mycorrhizal fungi: a review. *Ecosytems* **11:** 815-830.

Hopple JS, Vilgalys R. (1994). Phylogenetic relationships among coprinoid taxa and allies based on data from restriction site mapping of nuclear rDNA. *Mycologia* **86:** 96-107.

Horton TR, and Bruns TD. (2001). The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. *Mol Ecol* **10:** 1855-1871.

Kellner H, Vandenbol M, Zak DR. (2010). Fungi Unearthed : Transcripts encoding lignocellulolytic and chitinolytic enzymes in forest soil. *PLoS One* **5:** e10971.

Kõljalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, *et al.* (2005). UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytol* **166:** 1063–1068.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947–2948.

LoBuglio KF, Berbee ML, Taylor JW. (1996). Phylogenetic origins of the asexual mycorrhizal symbionts *Cenococcum geophilum* Fr. And other mycorrhizal fungi among the Ascomycetes. *Mol Phylogenet Evol* **6:** 287-294.
Lozupone CA, Hamady M, Kelley ST, Knight R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* **73:** 1576–85.

Lynd LR, Weimer PJ, Zyl WHV, Pretorius IS. (2002). Microbial Cellulose Utilization : Fundamentals and Biotechnology. *Microbiol Mol Biol Rev* **66:** 506–577.

Magill AH, Aber JD. (1998). Long-term effects of experimental nitrogen additions on foliar litter decay and humus formation in forest ecosystems. *Plant and Soil* **203:** 301–311.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–80.

Melillo JM, Aber JD, Muratore JF. (1982). Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. *Ecology* **63:** 621-626.

Micks P, Downs MR, Magill AH, Nadelhoffer KJ, Aber JD. (2004). Decomposing litter as a sink for -enriched additions to an oak forest and a red pine plantation. *For Ecol Manage* **196:** 71–87.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N. (2008). Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evol Bioinf Online* **4:** 193–201.

Nilsson RH, Veldre V, Hartmann M, Unterseher M, Amend A, Bergsten J, *et al.* (2010). An open source software package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-throughput community assays and molecular ecology. *Fungal Ecol* **3:** 284-287.

Osono T. (2007). Ecology of ligninolytic fungi associated with leaf litter decomposition. *Ecol Res* **22:** 955–974.

Peterjohn WT, Melillo, JM, Steudler PA, Newkirk KM (1994). Responses of trace gas fluxes and N availability to experimentally elevated soil temperatures. *Ecol Appl* **4:** 617-625.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12:** 38.

Savoie JM, Mata G, Billette C. (1998). Extracellular laccase production during hyphal interactions between *Trichoderma* sp. and Shiitake, *Lentinula edodes*. *Appl Microbiol Biotechnol* **49:** 589-593.

Seifert KA. (2009). Progress towards DNA barcoding of fungi. *Mol Ecol Resour* **9:** 83–89.

Talbot JM, Allison SD, Treseder KK. (2008). Decomposers in disguise: mycorrhizal fungi as regulators of soil C dynamics in ecosystems under global change. *Funct Ecol* **22:** 955–963.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28:** 2731-2739.

Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I. (2008). Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. *New Phytol* **180:** 479-490.

Tsujiyama S, Minami M. (2005). Production of phenol-oxidizing enzymes in the interaction between white-rot fungi. *Mycoscience* **46:** 268–271.

Van der Woude MW, Boominathan K, Adinarayana Reddy C. (1993). Nitrogen regulation of lignin peroxidase and manganese-dependent peroxidase production is independent of carbon and manganese regulation in *Phanerochaete chrysosporium*. *Arch Microbiol* **160:** 1-4.

Vilgalys R. (2003). Taxonomic misidentification in public DNA databases. *New Phytol* **160:** 4-5.

Vilgalys R, Hester M. (1990). Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *J Bacteriol* **172:** 4238-4246.

White NA, Boddy L. (1992). Extracellular enzyme localization during interspecific fungal interactions. *FEMS Microbiol Lett* **98:** 75-79.

White TJ, Bruns T, Lee S, Taylor J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ (eds). PCR protocols: a Guide to Methods and Applications. Academic Press: San Diego, CA, pp 315-322.

Worrall JJ, Anagnost SE, Zabel RA. (1997). Comparison of wood decay among diverse lignicolous fungi. *Mycologia* **89:** 199-219.

Zak DR, Pregitzer KS, Burton AJ, Edwards IP, Kellner H. (2011). Microbial responses to a changing environment: implications for the future functioning of terrestrial ecosystems. *Fungal Ecol* **4:** 386-395.

Table 1. Families of fungi with significant changes in relative abundance calculated by ANOVA and Tukey's HSD ($p < 0.05$). Letters indicate significant differences.

		Relative abundance (%)		
		Control	Low N	High N
Ascomycota	Botryosphaeriaceae	0.02^{a}	0.00^{a}	0.20^{b}
	Herpotrichiellaceae	0.00^{a}	0.00^{a}	0.21^{b}
	Dermateaceae	0.04^{a}	0.12^{ab}	0.23^{b}
	Leotiaceae	0.00^{a}	0.10^{b}	0.00^{a}
	Venturiaceae	0.04^{a}	0.11^{ab}	0.51^{b}
	Clavicipitaceae	0.00^{a}	0.02^{ab}	0.20^{b}
	Hypocreaceae	0.68^{a}	0.48^{a}	1.85^{b}
Basidiomycota	Pleurotaceae	0.00^{a}	0.67^{b}	0.19^{ab}
	Sclerodermataceae	0.00^{a}	0.17^{a}	2.60^{b}
	Sporidiobolaceae	0.00^{a}	0.02^{ab}	0.10^{b}
	Tremellaceae	0.03^{a}	0.00^{a}	0.26^{b}
Mucoromycotina	Mortierellaceae	1.03^{a}	0.90^{a}	2.75^{b}

Table 2. Ecology of families or groups that are associated with one or more communities through either beta-diversity measurements or significant changes in relative abundance. All ecological data is from Cannon and Kirk (2007) except for data on Cenococcum (LoBuglio *et al.*, 1996). Primary nutritional modes are derived from ecological data and are designated as ectomycorrhizal (E), saprobic (S), parasitic/pathogenic (P) or lichenized (L). Plot associations are designated as C (control), L (low N) or H (high N).

Phylum	Family/group	Nutrition	Plot	Ecology Summary
Ascomycetes	Pyronemataceae	S	C, L	saprobic on soil, dung or rotten wood
	Myxotrichaceae	S, E	C, L	many species cellulolytic, other mycorrhizal
	Cenococcum	E	C, L	ectomycorrhizal
	Leotiaceae	S	L	saprobic especially on ferns
	Helotiaceae	S	L, H	usually saprobic on herbaceous and woody material
	Umbilicariaceae	L	L, H	lichenized with green algae
	Botryosphaeriaceae	S	H	saprobic on woody tissue
	Clavicipitaceae	P	H	parasitic on Insecta, fungi, or *Gramineae*
	Dermateaceae	S, P	H	saprobic or parasitic on herbaceous and woody material
	Herpotrichiellaceae	S	H	mostly saprobic on plants or fungi
	Hypocreaceae	S, P	H	saprobic on rotting wood or parasitic on other fungi
	Venturiaceae	S, P	H	usually growing on woody substrata, often parasitic on other fungi
Basisdiomycetes	Amanitaceae	E, S	C	ectomycorrhizal or saprobic on rotten wood, plant remains or humus
	Atheliaceae	E, S	C, L	ectomycorrhizal or saprobes/active decomposers
	Cortinariaceae	E	C, L	ectomycorrhizal with woody plants
	Thelephoraceae	E, S	C, L	on the ground, soil, a few on wood, some species are ectomycorrhizal
	Boletaceae	E, S	L	mostly ectomycorrhizal, some saprobes on decaying wood and leaf litter
	Inocybaceae	E	L	ectomycorrhizal, terrestrial or on very rotten wood
	Pleurotaceae	S	L	saprobic on woody substrata, or rarely root associated
	Sebacinaceae	E	L	mycorrhizal or ectomycorrhizal with a broad-range of plants
	Tricholomataceae	E, S	L	ectomycorrhizal, occurring in coniferous and broadleaved forests.
	Corticiaceae	S, P	L, H	saprobes or pathogens on a wide range of mainly woody substrata
	Exidiaceae	S, P	H	saprobic or possibly weakly parasitic on wood and bark
	Sclerodermataceae	E	H	often associated with rotten wood, mostly ectomycorrhizal
	Sporidiobolaceae	S	H	saprobic, found in a wide variety of habitats
	Tremellaceae	P	H	usually growing on woody substrata, often parasitic on other fungi
Mucoromycotina	Mortierellaceae	S	H	saprobic in soil

Table 3. Pearson's correlation coefficicent from ANOSIM on the effects of N treatment on community composition from relative abundance weighted (Bray-Curtis or weighted Unifrac), and presence absence based (Sorensen-Dice or unweighted Unifrac) β diversity metrics. ANOSIM was calculated for the total community and the rare community (all OTUs less than 2% of relative abundance averaged across samples), as well as for Ascomycetes, Basidiomycetes, the dominant Basidiomycete community (all OTUs greater than 2% of relative abundance averaged across samples), and the rare Basidiomycete community separately. Significant values are in bold ($P = 0.05$), highly significant values are in bold and italics ($P = 0.01$). Asterisks indicate where Mantel tests indicate higher correlation between ITS1 and ITS2 distance matrices when ANOSIM trends are not in agreement.

		Total	Rare	Basidiomycetes	dominant Basidiomycetes	rare Basidomycetes	Ascomycetes
ITS1	Bray-Curtis	***0.486****	***0.819***	**0.486**	***0.481***	***0.646***	0.206
	Sorensen-Dice	***0.782***	***0.597***	***0.642***	***0.844***	**0.465***	**0.556***
ITS2	Bray-Curtis	**0.663***	***0.588***	***0.461***	0.55	***0.46***	***0.535***
	Sorensen-Dice	**0.625**	**0.498**	***0.786***	0.35	***0.609****	***0.494****
LSU weighted Unifrac		na	na	0.235	**0.226**	-0.078	na
	unweighted Unifrac	na	na	***0.44***	***0.63***	***0.523***	na

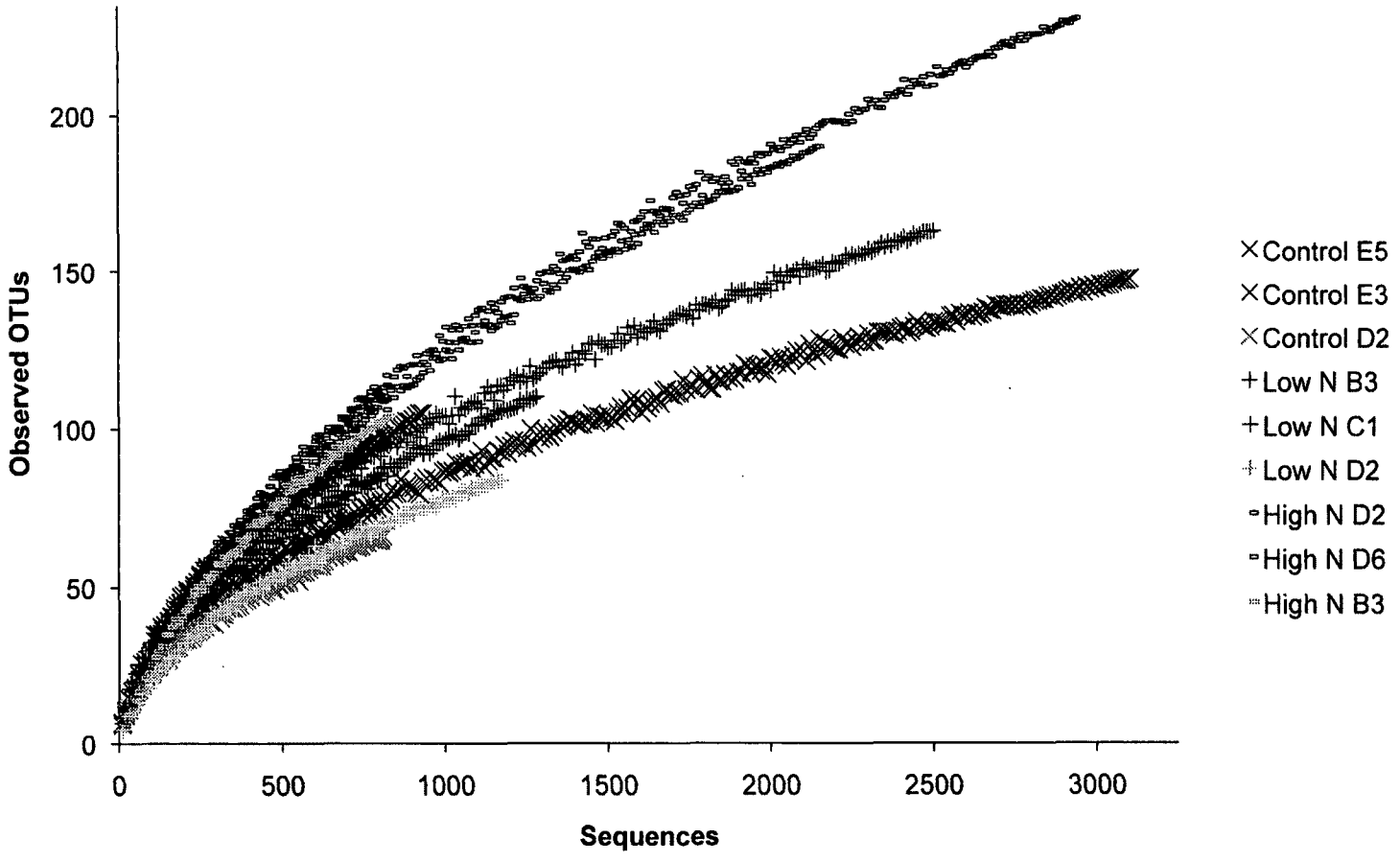


Figure 1. Species accumulation curves for all samples from ITS1 sequences clustered at 95% sequence similarity.

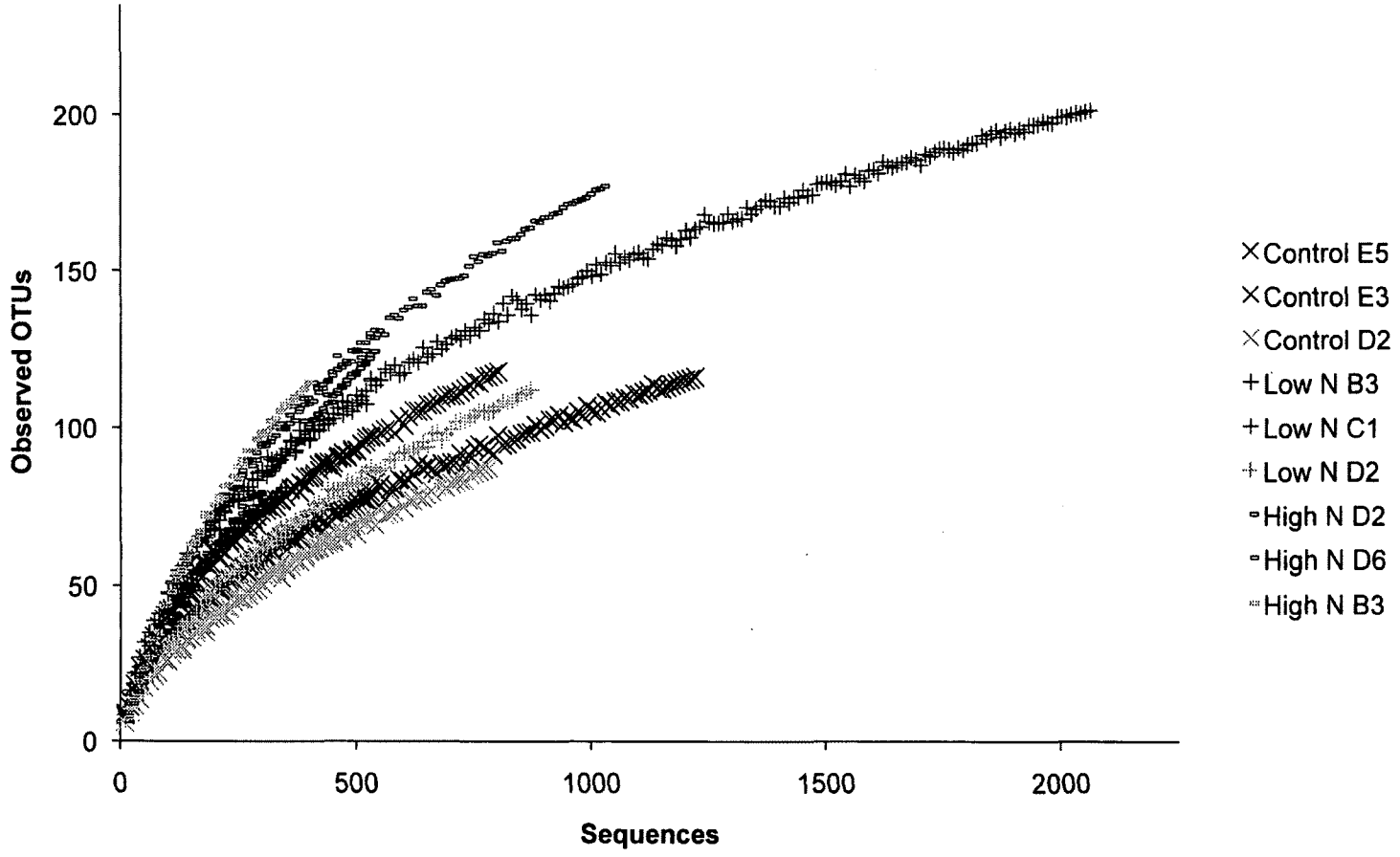


Figure 2. Species accumulation curves for all samples from ITS2 sequences clustered at 95% sequence similarity.

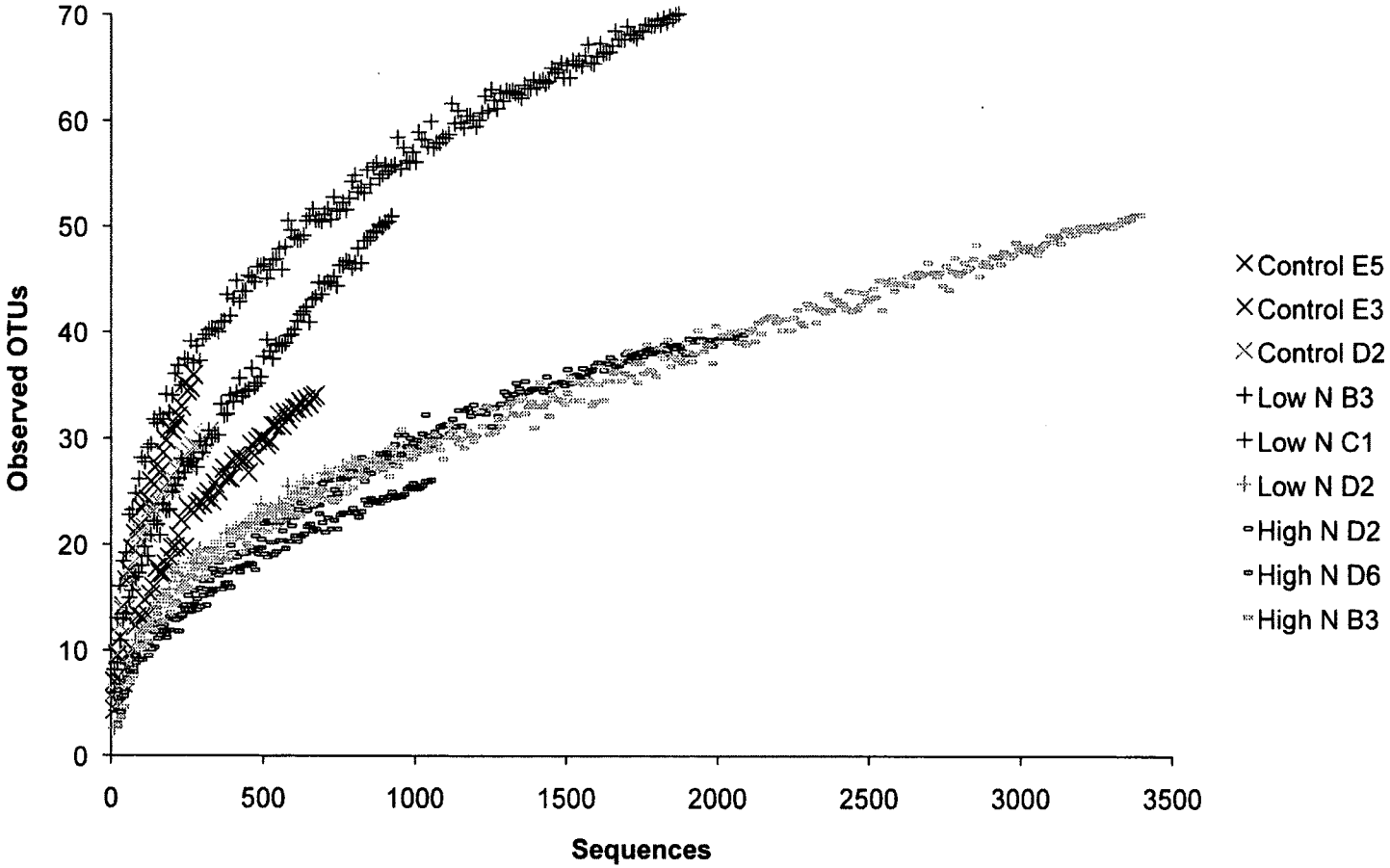


Figure 3. Species accumulation curves for all samples from LSU sequences clustered at 100% sequence similarity.

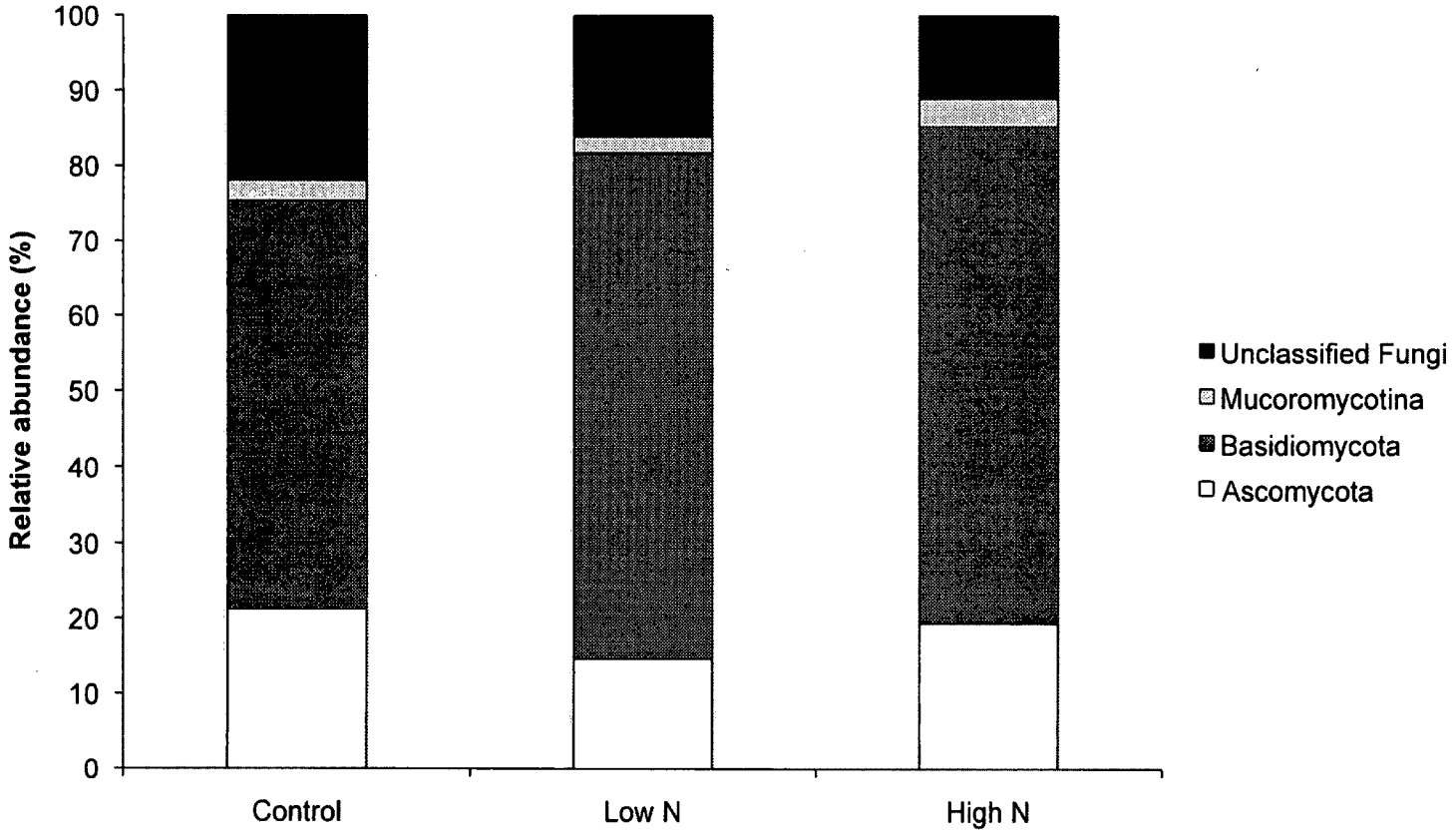


Figure 4. Relative abundance of fungal phyla and subphyla that made up greater than 0.1% of relative abundance in Control, Low N or High N treatments.

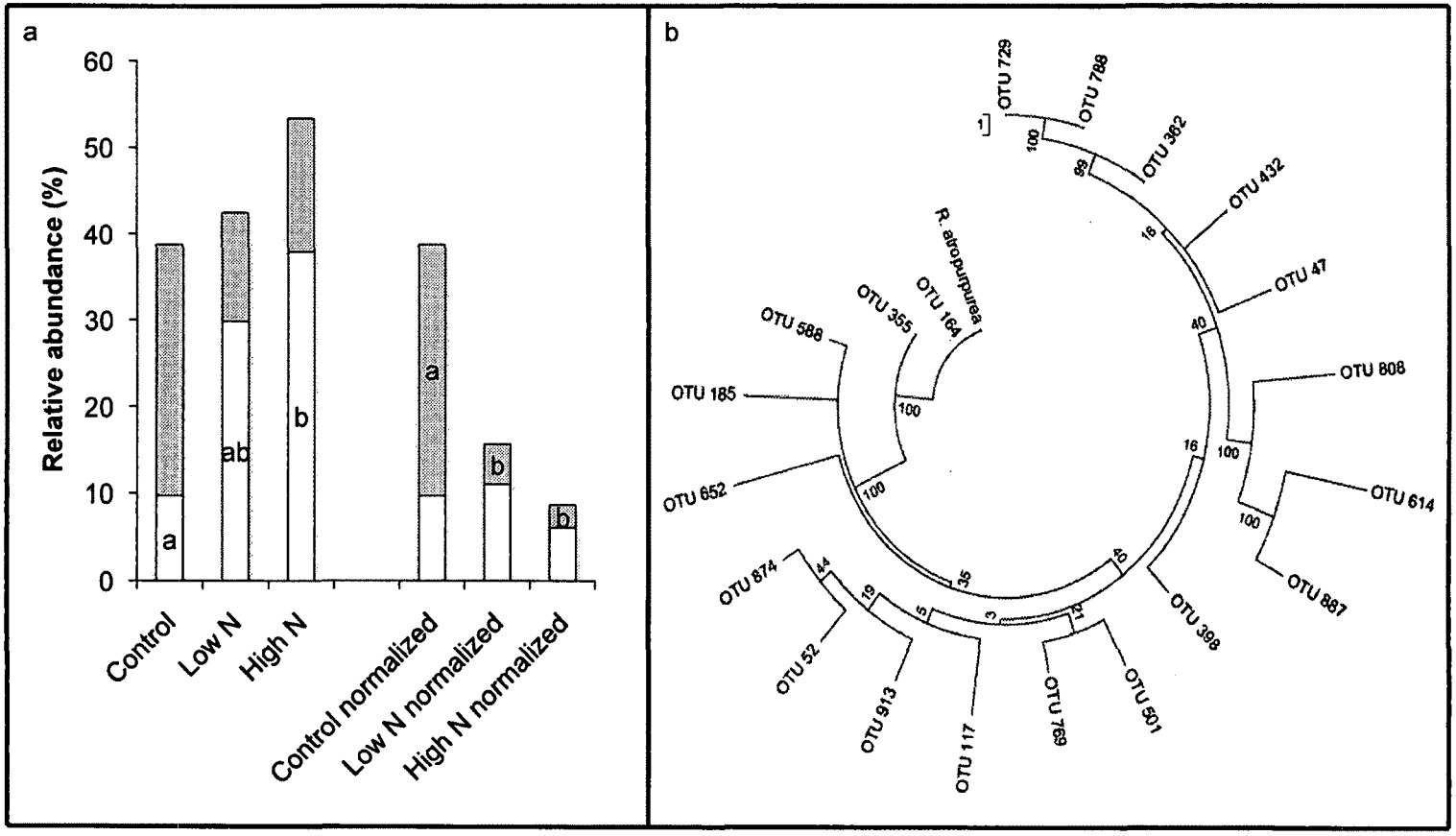


Figure 5. a. Relative abundance of *Russula atropurpurea* (white bars) and other *Russula spp.* (gray bars) in control, low N and high N treatments, b. Maximum parsimony tree of OTUs identified as *Russula spp.* and a *Russula atropupurea* representative sequence from NCBI. OTU 164 is putatively identified as *Russula atropurpurea*.

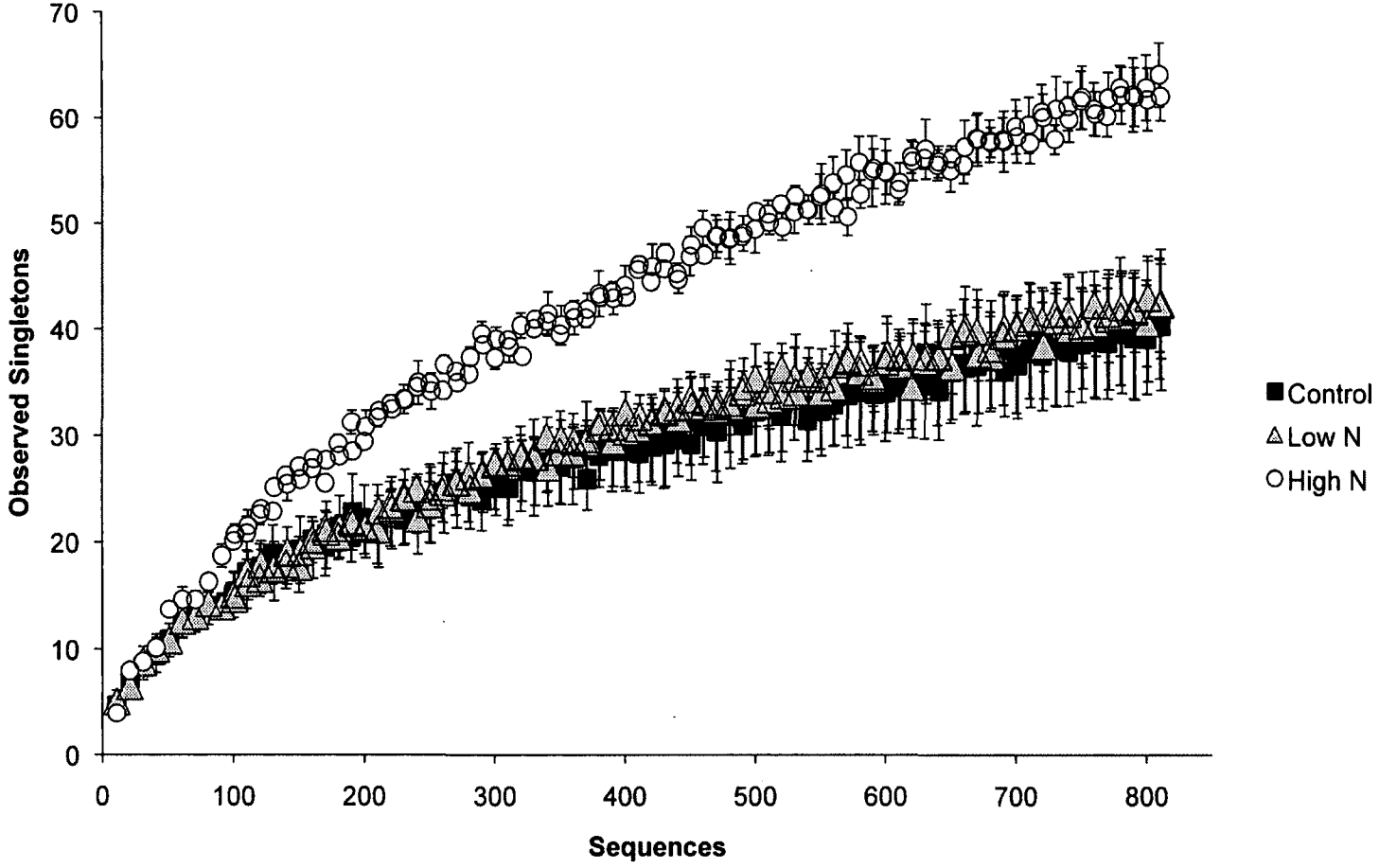


Figure 6. Number of singletons observed by sequencing depth from ITS1 sequences clustered at 95% sequence similarity in Control, Low N and High N soils.

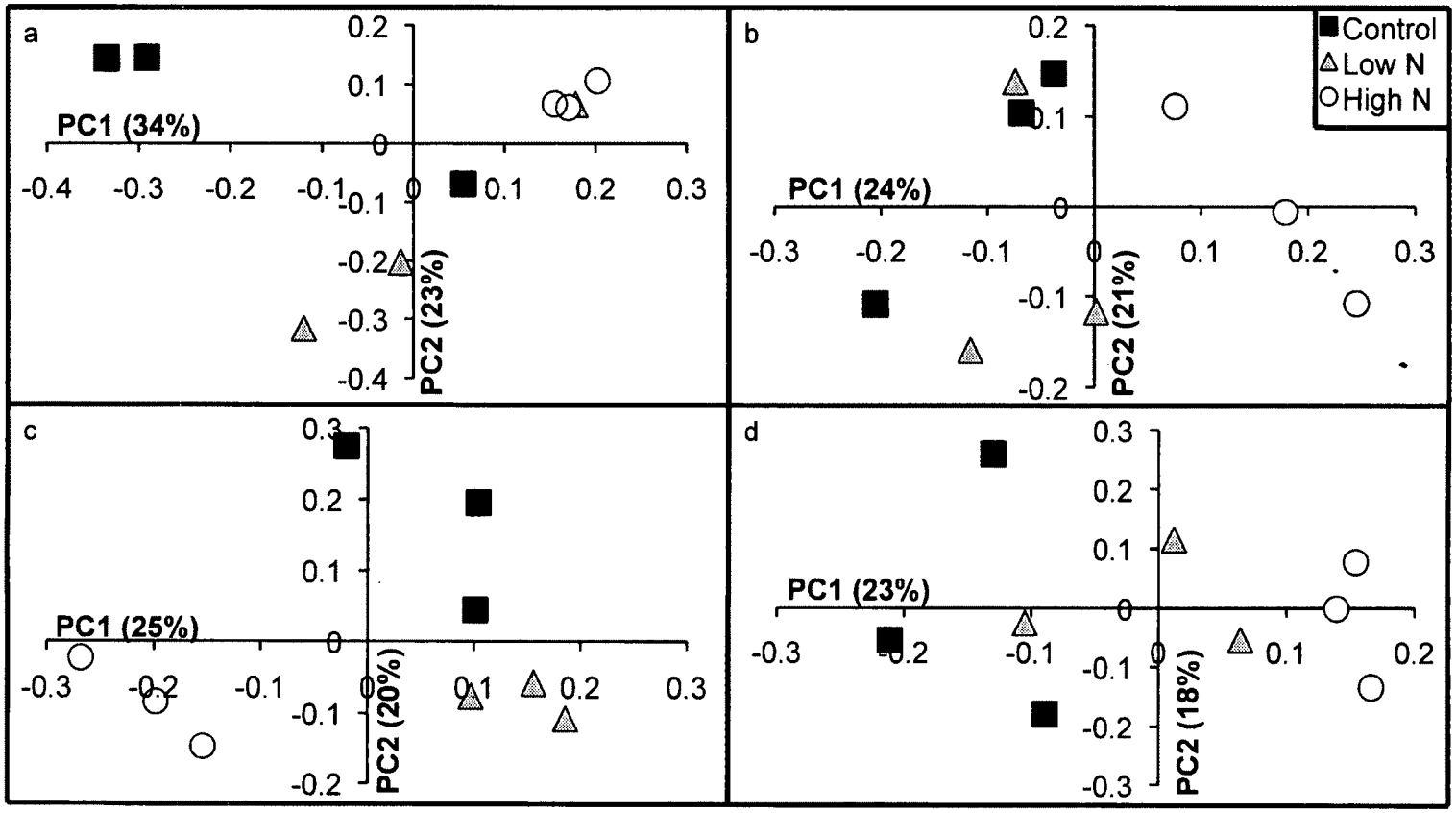


Figure 7. Principal coordinates analysis of a) total community member abundance, b) Ascomycete community membership, c) Basidiomycete community membership, and d) phylogentic Basidiomycete community membership.

APPENDICES

APPENDIX A. Sample DNA concentration (ng/μl) and absorption spectra (260/280) of initial DNA extraction.

Sample ID	ng/μl	260/280
Control D2	1002.99	1.56
Control E3	1619.82	1.54
Control E5	812.05	1.53
Low N C1	1077.41	1.48
Low N D2	574.95	1.76
Low N B3	683.23	1.63
High N B3	549.13	1.52
High N D6	831.93	1.50
High N D2	353.59	1.55

APPENDIX B. Sample DNA concentration (ng/μl) and absorption spectra (260/280) of PCR amplicons and sample normalization for sequencing library.

Locus	Sample ID	ng/μl	260/280	Volume to add for normalization (μl)	Volumes (μl) multiplied by 9 for 90 ul total sample
LSU	Control D2	35.31	1.76	0.50	4.50
LSU	Control E3	51.70	0.78	0.34	3.07
LSU	Control E5	67.28	0.87	0.26	2.36
LSU	Low N B3	55.20	1.64	0.32	2.88
LSU	Low N C1	56.08	1.49	0.31	2.83
LSU	Low N D2	86.84	1.42	0.20	1.83
LSU	High N D2	59.66	1.55	0.30	2.66
LSU	High N B3	56.22	1.64	0.31	2.83
LSU	High N D6	56.02	0.82	0.32	2.84
LSU	Control 3	37.69	0.67	0.47	4.22
LSU	Low N 3	17.66	1.52	1.00	9.00
LSU	High N 1	70.32	0.9	0.25	2.26
ITS1	Control D2	153.72	1.15	0.11	1.03
ITS1	Control E3	50.17	0.78	0.35	3.17
ITS1	Control E5	160.00	1.16	0.11	0.99
ITS1	Low N B3	147.66	1.13	0.12	1.08
ITS1	Low N C1	74.11	0.9	0.24	2.14
ITS1	Low N D2	121.24	1.09	0.15	1.31
ITS1	High N D2	49.76	1.76	0.35	3.19
ITS1	High N B3	148.85	1.13	0.12	1.07
ITS1	High N D6	65.18	0.91	0.27	2.44
ITS1	Control 3	21.26	2.65	0.83	7.48
ITS1	Low N 3	156.25	1.19	0.11	1.02
ITS1	High N 1	48.69	2.04	0.36	3.26
ITS2	Control D2	79.34	0.91	0.22	2.00
ITS2	Control E3	119.61	1.33	0.15	1.33
ITS2	Control E5	108.00	1.04	0.16	1.47
ITS2	Low N B3	127.51	1.1	0.14	1.25
ITS2	Low N C1	67.01	1.21	0.26	2.37
ITS2	Low N D2	58.16	1.17	0.30	2.73
ITS2	High N D2	190.19	1.22	0.09	0.84
ITS2	High N B3	137.61	1.16	0.13	1.16
ITS2	High N D6	88.29	1.02	0.20	1.80
ITS2	Control 3	78.80	0.96	0.22	2.02
ITS2	Low N 3	78.89	0.97	0.22	2.01
ITS2	High N 1	140.33	1.16	0.13	1.13

APPENDIX C. Relative abundance of all taxonomic groups to the family level from combined ITS1 and ITS2 data.

Taxonomy				Control	Low N	High N
Ascomycota	Acrodontium incertae sedis	Acrodontium incertae sedis	Acrodontium incertae sedis	0.00	0.00	0.04
Ascomycota	Dothideomycetes	Botryosphaeriales	Botryosphaeriaceae	0.02	0.00	0.20
Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	0.00	0.00	0.02
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	0.11	0.07	0.32
Ascomycota	Dothideomycetes	Capnodiales	Teratosphaeriaceae	0.00	0.00	0.07
Ascomycota	Dothideomycetes	Cenococcum incertae sedis	Cenococcum incertae sedis	6.21	3.25	0.07
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	0.00	0.00	0.10
Ascomycota	Dothideomycetes	Pleosporales	Coniothyrium incertae sedis	0.00	0.04	0.04
Ascomycota	Dothideomycetes	Pleosporales	Didymellaceae	0.03	0.00	0.02
Ascomycota	Dothideomycetes	Pleosporales	Phaeosphaeriaceae	0.00	0.04	0.12
Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	0.00	0.07	0.00
Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	0.00	0.00	0.11
Ascomycota	Dothideomycetes	Venturiales	Venturiaceae	0.06	0.04	0.11
Ascomycota	Dothideomycetes	Mycorrhizal samples	Mycorrhizal samples	0.16	0.07	0.04
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	0.00	0.00	0.21
Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	1.06	0.46	0.77
Ascomycota	Eurotiomycetes	Onygenales	Paracoccidioides incertae sedis	0.00	0.02	0.02
Ascomycota	Lecanoromycetes	Lecanorales	Physciaceae	0.21	0.04	0.04
Ascomycota	Lecanoromycetes	Lecanorales	Environmental samples	0.12	0.00	0.00
Ascomycota	Lecanoromycetes	Umbilicariales	Umbilicariaceae	0.11	0.10	0.20
Ascomycota	Lecanoromycetes	Xanthoria incertae sedis	Teloschistaceae	0.06	0.00	0.00

Taxonomy				Control	Low N	High N
Ascomycota	Leotiomycetes	Antarctomyces incertae sedis	Thelebolaceae	0.00	0.04	0.00
Ascomycota	Leotiomycetes	Colpoma incertae sedis	Rhytismataceae	0.12	0.02	0.00
Ascomycota	Leotiomycetes	Geomyces incertae sedis	Myxotrichaceae	0.06	0.15	0.10
Ascomycota	Leotiomycetes	Helotiales	Cadophora incertae sedis	0.03	0.00	0.00
Ascomycota	Leotiomycetes	Helotiales	Chaetomella incertae sedis	0.00	0.00	0.02
Ascomycota	Leotiomycetes	Helotiales	Dermateaceae	0.04	0.12	0.23
Ascomycota	Leotiomycetes	Helotiales	Gloeotinia incertae sedis	0.05	0.06	0.20
Ascomycota	Leotiomycetes	Helotiales	Helotiaceae	0.53	1.15	2.09
Ascomycota	Leotiomycetes	Helotiales	Hyaloscyphaceae	0.47	0.24	0.38
Ascomycota	Leotiomycetes	Helotiales	Leotiaceae	0.00	0.10	0.00
Ascomycota	Leotiomycetes	Helotiales	Phialocephala incertae sedis	0.00	0.04	0.02
Ascomycota	Leotiomycetes	Helotiales	Sclerotiniaceae	0.00	0.00	0.04
Ascomycota	Leotiomycetes	Helotiales	Environmental samples	0.05	0.56	0.32
Ascomycota	Leotiomycetes	Helotiales	unclassified Helotiales	0.05	0.11	0.28
Ascomycota	Leotiomycetes	Leohumicola incertae sedis	Leohumicola incertae sedis	0.84	0.44	0.59
Ascomycota	Leotiomycetes	Lophodermium incertae sedis	Rhytismataceae	0.04	0.02	0.00
Ascomycota	Leotiomycetes	Microglossum incertae sedis	Microglossum incertae sedis	0.00	0.00	0.04
Ascomycota	Leotiomycetes	Oidiodendron incertae sedis	Myxotrichaceae	2.75	1.63	1.72
Ascomycota	Orbiliomycetes	Orbiliales	Orbiliaceae	0.08	0.00	0.06
Ascomycota	Pezizomycetes	Pezizales	Morchellaceae	0.00	0.04	0.15
Ascomycota	Pezizomycetes	Pezizales	Pezizaceae	0.02	0.02	0.05
Ascomycota	Pezizomycetes	Pezizales	Pyronemataceae	0.00	0.59	0.20
Ascomycota	Pezizomycetes	Pezizales	Tuberaceae	0.06	0.00	0.04
Ascomycota	Pezizomycetes	Pezizales	Unclassified Pezizales	0.00	0.10	0.04
Ascomycota	Rhizosphaera incertae sedis	Rhizosphaera incertae sedis	Rhizosphaera incertae sedis	0.04	0.11	0.51
Ascomycota	Saccharomycetes	Saccharomycetales	Candida incertae sedis	0.56	0.11	0.41
Ascomycota	Saccharomycetes	Saccharomycetales	Dipodascaceae	0.00	0.00	0.02

Taxonomy				Control	Low N	High N
Ascomycota	Saccharomycetes	Saccharomycetales	Lipomycetaceae	0.00	0.11	0.00
Ascomycota	Saccharomycetes	Saccharomycetales	Trichomonascaceae	1.39	0.26	0.56
Ascomycota	Schizosaccharomycetes	Schizosaccharomycetales	Schizosaccharomycetaceae	0.00	0.02	0.02
Ascomycota	Sordariomycetes	Diaporthales	Valsaceae	0.03	0.07	0.11
Ascomycota	Sordariomycetes	Glomerellales	Glomerellaceae	0.19	0.48	0.10
Ascomycota	Sordariomycetes	Hypocreales	Clavicipitaceae	0.00	0.02	0.20
Ascomycota	Sordariomycetes	Hypocreales	Cordycipitaceae	1.15	0.30	0.67
Ascomycota	Sordariomycetes	Hypocreales	Fusarium incertae sedis	0.00	0.00	0.04
Ascomycota	Sordariomycetes	Hypocreales	Hypocreaceae	0.68	0.48	1.85
Ascomycota	Sordariomycetes	Hypocreales	Myrothecium incertae sedis	0.06	0.00	0.05
Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	0.05	0.02	0.31
Ascomycota	Sordariomycetes	Hypocreales	Ophiocordycipitaceae	0.03	0.30	0.02
Ascomycota	Sordariomycetes	Hypocreales	Verticillium incertae sedis	0.00	0.04	0.17
Ascomycota	Sordariomycetes	Magnaporthales	Magnaporthaceae	0.09	0.06	0.09
Ascomycota	Sordariomycetes	Ophiostomatales	Ophiostomataceae	0.00	0.00	0.07
Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	0.12	0.00	0.11
Ascomycota	Sordariomycetes	Sordariales	Sordariaceae	0.00	0.00	0.02
Ascomycota	Sordariomycetes	Xylariales	Amphisphaeriaceae	0.00	0.00	0.11
Ascomycota	Sordariomycetes	Xylariales	Diatrypaceae	0.00	0.00	0.07
Ascomycota	Sordariomycetes	Xylariales	Xylariaceae	0.00	0.07	0.04
Ascomycota	Environmental samples	Environmental samples	Environmental samples	0.82	0.02	0.46
Ascomycota	Mycorrhizal samples	Mycorrhizal samples	Mycorrhizal samples	0.02	0.00	0.04
Ascomycota	Unclassified Ascomycota	Unclassified Ascomycota	Unclassified Ascomycota	2.66	2.52	4.25
Basidiomycota	Agaricomycetes	Agaricales	Agaricaceae	0.03	0.07	0.15
Basidiomycota	Agaricomycetes	Agaricales	Amanitaceae	2.81	1.28	0.41
Basidiomycota	Agaricomycetes	Agaricales	Cortinariaceae	5.98	2.98	2.33

Taxonomy				Control	Low N	High N
Basidiomycota	Agaricomycetes	Agaricales	Inocybe incertae sedis	0.00	0.59	0.02
Basidiomycota	Agaricomycetes	Agaricales	Lycoperdaceae	0.00	0.00	0.05
Basidiomycota	Agaricomycetes	Agaricales	Lyophyllaceae	0.00	0.00	0.07
Basidiomycota	Agaricomycetes	Agaricales	Marasmiaceae	0.04	0.00	0.00
Basidiomycota	Agaricomycetes	Agaricales	Pleurotaceae	0.00	0.67	0.19
Basidiomycota	Agaricomycetes	Agaricales	Psathyrellaceae	0.00	0.22	0.04
Basidiomycota	Agaricomycetes	Agaricales	Strophariaceae	0.00	0.00	0.02
Basidiomycota	Agaricomycetes	Agaricales	Tricholomataceae	0.66	6.36	0.17
Basidiomycota	Agaricomycetes	Atheliales	Atheliaceae	0.52	0.26	0.10
Basidiomycota	Agaricomycetes	Boletales	Boletaceae	0.02	0.96	0.07
Basidiomycota	Agaricomycetes	Boletales	Sclerodermataceae	0.00	0.17	2.60
Basidiomycota	Agaricomycetes	Cantharellales	Botryobasidiaceae	0.00	0.06	0.19
Basidiomycota	Agaricomycetes	Cantharellales	Ceratobasidiaceae	0.00	0.00	0.04
Basidiomycota	Agaricomycetes	Cantharellales	Clavulinaceae	0.02	0.07	0.11
Basidiomycota	Agaricomycetes	Cantharellales	Hydnaceae	0.00	0.07	0.00
Basidiomycota	Agaricomycetes	Corticiales	Corticiaceae	0.00	0.00	0.09
Basidiomycota	Agaricomycetes	Hymenochaetales	Hymenochaetaceae	0.03	0.00	0.07
Basidiomycota	Agaricomycetes	Polyporales	Auriscalpiaceae	0.00	0.00	0.04
Basidiomycota	Agaricomycetes	Polyporales	Coriolaceae	0.05	0.02	0.11
Basidiomycota	Agaricomycetes	Polyporales	Ganodermataceae	0.00	0.06	0.06
Basidiomycota	Agaricomycetes	Polyporales	Polyporaceae	0.09	0.17	0.38
Basidiomycota	Agaricomycetes	Russulales	Russulaceae	39.97	45.08	54.26
Basidiomycota	Agaricomycetes	Sebacinales	Sebacinaceae	0.31	1.52	0.11
Basidiomycota	Agaricomycetes	Thelephorales	Thelephoraceae	1.26	3.24	1.05
Basidiomycota	Agaricomycetes	Unclassified Agaricomycetes	Unclassified Agaricomycetes	0.00	0.04	0.07
Basidiomycota	Agaricostilbomycetes	Sporobolomyces incertae sedis	Sporobolomyces incertae sedis	0.03	0.00	0.00
Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	0.28	0.12	0.05

Taxonomy				Control	Low N	High N
Basidiomycota	Microbotryomycetes	Rhodotorula incertae sedis	Rhodotorula incertae sedis	0.03	0.19	0.31
Basidiomycota	Microbotryomycetes	Sporidiobolales	Rhodosporidium incertae sedis	0.08	0.06	0.07
Basidiomycota	Microbotryomycetes	Sporidiobolales	Rhodotorula incertae sedis	0.00	0.00	0.04
Basidiomycota	Microbotryomycetes	Sporidiobolales	Sporobolomyces incertae sedis	0.03	0.07	0.15
Basidiomycota	Microbotryomycetes	Sporobolomyces incertae sedis	Sporobolomyces incertae sedis	0.00	0.02	0.10
Basidiomycota	Rhizoctonia incertae sedis	Rhizoctonia incertae sedis	Rhizoctonia incertae sedis	0.14	0.05	0.10
Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	0.00	0.00	0.02
Basidiomycota	Tremellomycetes	Filobasidiales	Cryptococcus incertae sedis	0.48	0.10	0.53
Basidiomycota	Tremellomycetes	Tremellales	Bullera incertae sedis	0.03	0.00	0.26
Basidiomycota	Tremellomycetes	Tremellales	Cryptococcus incertae sedis	0.54	0.74	0.76
Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	0.16	0.10	0.05
Basidiomycota	Environmental samples	Environmental samples	Environmental samples	0.36	0.02	0.07
Basidiomycota	Unclassified Basidiomycota	Unclassified Basidiomycota	Unclassified Basidiomycota	0.07	1.56	0.43
Chytridiomycota	Chytridiomycetes	Rhizophydiales	Terramycetaceae	0.02	0.02	0.07
Glomeromycota	Glomeromycetes	Diversisporales	Scutellosporaceae	0.04	0.05	0.05
Glomeromycota	Glomeromycetes	Glomerales	Glomeraceae	0.00	0.02	0.00
Mortierellomycotina	Mortierella incertae sedis	Mortierellales	Mortierellaceae	1.03	0.90	2.75
Mucoromycotina	Umbelopsis incertae sedis	Mucorales	Umbelopsis incertae sedis	1.69	1.30	0.93
Neocallimastigomycota	Neocallimastigomycetes	Neocallimastigales	Neocallimastigaceae	0.00	0.00	0.05
Environmental samples	Environmental samples	Environmental samples	Environmental samples	18.29	10.51	5.17
Unclassified Fungi	Unclassified Fungi	Unclassified Fungi	Unclassified Fungi	2.23	4.37	4.79
Other	Other	Other	Other	1.39	1.16	0.97
Incertae sedis	Incertae sedis	Incertae sedis	Incertae sedis	0.00	0.02	0.05

APPENDIX D. Scripts for taxonomic identification of representative sequences. Sequences were first sorted by OTU number (sortRepFasta.pl), and then input to the OCTUPUS pipeline at the blast step through the getTaxonomy.pl step (Sung W, in prep). Data was then run through scripts that were either modified from OCTUPUS or created for this dataset.

sortRepFasta.pl

```
#!/usr/bin/perl
#Eric Morrison
# 6/15/2011
#sortRepFasta.pl


use strict;
use warnings;

my $fasta = $ARGV[0];

open (FASTA, "$fasta") || die "Cant open fasta file. \n";
open (DEST, ">$fasta.sort") || die "Cant open dest.\n";

chomp (my @seqs = <FASTA>);
my %fasta = @seqs;
my %seqs;
foreach my $key (keys %fasta)
        {
        my @num = split(" ", $key);
        $num[0] =~ s/>//;
        print $num[0], "\n";
        my $num = $num[0];
        $seqs{$num} = $key."\n".$fasta{$key};
        }
foreach my $sort (sort {$a<=>$b} keys %seqs)
        {
        print DEST $seqs{$sort}, "\n";
        }
```

fakeDiveQiime.pl

```
#!/usr/bin/perl
#
# fakeDivQiime.pl
# Eric Morrison
# 6/15/2011
#
# This script takes a Qiime rep_set.fna file as input and writes
# a OTCUPUS type "div" file with arbitrary nucleotide diversity
# numbers. This file can be used for input to blastFinal.
#
# Usage: fakeDivQiime.pl [input]

use strict;
use warnings;

my $input = $ARGV[0];

open (IN, "$input") || die "Cant open input.\n";
open (OUT, ">$input.fakeDiv") || die "Cant open output.\n";

chomp(my @cons = <IN>);
my $cons = join ("\n", @cons);
@cons = split(">", $cons);
shift @cons;
my %cons;

foreach (@cons)
        {
        my @seqs = split("\n", $_);
        my $head = shift @seqs;
        my @head = split(" ", $head);
        $head = $head[0];
        my $seq = join ("", @seqs);
        $cons{$head} = $seq;
        }

print OUT "octu\tsequences\toctuLength\tnucDiversity\n";

foreach my $oc (sort {$a <=> $b} keys %cons)
        {
        print OUT $oc, "\t1\t1\t0.0", "\n";
        }
```

blastFinalforQiimeTaxa.pl (modified from OCTUPUS)

```
#!/usr/bin/perl
#Way Sung 6/2007
# modifications made by Eric Morrison 6/15/2011

#   This file is part of OCTUPUS.

#   OCTUPUS is free software: you can redistribute it and/or modify
#   it under the terms of the GNU General Public License as published by
#   the Free Software Foundation, either version 3 of the License, or
#   (at your option) any later version.

#   OCTUPUS is distributed in the hope that it will be useful,
#   but WITHOUT ANY WARRANTY; without even the implied warranty of
#   MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See
the
#   GNU General Public License for more details.

#   You should have received a copy of the GNU General Public License
#   along with OCTUPUS.  If not, see <http://www.gnu.org/licenses/>.

sub usage()
{
  print STDERR q
  (
        Usage: blastFinal.pl <blastFilterFile> <div> <taxonomyFile> <outfile>

        This program combines processed files into a tab-delimited file
        that can be opened in excel.

        Parameters
        blastFilterFile - filtered blast file from blastFilter.pl
        div - diversity file from octu.pl
        taxonomyFile - taxonomy file from getTaxonomy.pl
        destination file
  );
  exit;
}

if ( ( @ARGV == 0 ) || $ARGV[0] eq "-h" )  #no arguments or help option
{
    &usage();
}
```

```
#Initialization of command line variables
my $blastFilter = $ARGV[0];
my $blastCount = $ARGV[1];
my $blastTaxonomy = $ARGV[2];
my $outfile = $ARGV[3];


#Initialization of source and destination file, printing headers in destination
open (FILTER, "$blastFilter") || die "cant load file";
open (COUNT, "$blastCount") || die "cant load file";
open (TAXA, "$blastTaxonomy") || die "cant load file";
open (DEST, ">$outfile") || die "cant load file";

print DEST join ("\t", "OCTU", "Sequences", "SeqLen", "BitScore", "E-Value",
"MatchBases", "TotalBases", "RefLength", "PercentID", "DBmatch", "GI",
"Definition", "Life", "Domain", "Kingdom", "Phylum", "Class", "Order", "Family",
"Genus", "Species");
print DEST "\n";

%count = ();
while (<COUNT>)
{
        chomp;
        if ($_ =~ /octu/)
        {
                next;
        }
        @input = split("\t", $_);
      $count{$input[0]} = $input[1];
}


$n=0; #mod from 1 to 0, Eric
while (<FILTER>)
{
        chomp;
        @input = split(/\t/, $_);
        $seqLen[$n] = $input[0];
        $dbmatch[$n] = $input[1];
        $bitscore[$n] = $input[2];
        $evalue[$n] = $input[3];
        $matchbases[$n] = $input[4];
        $totalbases[$n] = $input[5];
        $reflength[$n] = $input[6];
        $percentid[$n] = $input[7];
        $n++
```

```
}

$u = 0; #mod from 1 to 0, Eric
while(<TAXA>)
{
	chomp;
	$taxa{$u} = $_;
	$u++;
}

for $key ( sort {$count{$b}<=>$count{$a}} keys %count)
{
	print DEST join("\t", $key, $count{$key}, $seqLen[$key], $bitscore[$key],
$evalue[$key], $matchbases[$key], $totalbases[$key], $reflength[$key],
$percentid[$key], $dbmatch[$key], $taxa{$key}), "\n";
	$u++;
}
```

blastFinalParseQiime.pl

```
#!/usr/bin/perl
# Title: blastFinalParseQiime.pl
# Author: Eric Morrison
# Date: 1/25/11
# Usage: Enter a .blastfinal file and 2 output filenames as command line args.

use strict;
use warnings;

sub usage()
{
  print STDERR q
  (
Usage: blastFinalParseQiime.pl [octulist.blastFinal] [output] [altoutput]

        This script will parse information from .blastfinal file and return
        two cleaned files.

        The alternative output file can be used as input for downstream
        parsing scripts.

);
  exit;
}

if ( ( @ARGV == 0 ) || $ARGV[0] eq "-h" )   #no arguments or help option
{
    &usage();
}

my $input = $ARGV[0];
my $out = $ARGV[1];
my $altOut = $ARGV[2];

open (IN, "$input") || die "Cant open in file. $!.\n";
open (OUT, ">$out") || die "Cant open out file. $!.\n";
open (ALTOUT, ">$altOut") || die "Cant open out file 2. $!.\n";

# capture input data
chomp(my @all = <IN>);
my $headerRow = shift(@all);
my @header = split("\t", $headerRow);

# print headers for both outfiles
```

```
print ALTOUT "OCTU\tGI\tbitScore\teValue\t%ID\n";

while(@all)
{


        # single octu info variable
        my $oc = shift(@all);
        my @oc = split("\t", $oc);

        # create hash for basic info storgae
        my %octu = ("OCTU" => $oc[0], "seqs" => $oc[1], "seqLen" => $oc[2],
"bitScore" => $oc[3],
        "eValue" => $oc[4], "%ID" => $oc[8], "GI" => $oc[10]);
        $octu{"%ID"} = $octu{"%ID"} / 100;
        #$octu{"%ID"} =~ s/\d$//;
        # assess NA hits, if yes, next octu
        if ($oc[9] eq "NA")
                {print OUT $octu{"OCTU"}, "\tRoot\tNA\n";
                print ALTOUT $octu{"OCTU"}, "\t", "NA\tNA\tNA\n";
                next;
                }

        # foreach element, split by colon, assess taxonomic levels, assign to hash
        foreach (@oc)
                {my @element = split(":", $_);
                if (lc$element[0] eq "kingdom")
                        {$octu{lc$element[0]} = $element[1];
                        }
                elsif (lc$element[0] eq "subkingdom")
                        {$octu{lc$element[0]} = $element[1];
                        }
                elsif (lc$element[0] eq "phylum")
                        {$octu{lc$element[0]} = $element[1];
                        }
                elsif (lc$element[0] eq "subphylum")
                        {$octu{lc$element[0]} = $element[1];
                        }
                elsif (lc$element[0] eq "class")
                        {$octu{lc$element[0]} = $element[1];
                        }
                elsif (lc$element[0] eq "subclass")
                        {$octu{lc$element[0]} = $element[1];
                        }
                elsif (lc$element[0] eq "order")
                        {$octu{lc$element[0]} = $element[1];
```

```
            }
        elsif (lc$element[0] eq "family")
            {$octu{lc$element[0]} = $element[1];
            }
        elsif (lc$element[0] eq "genus")
            {$octu{$element[0]} = $element[1];
            }
        elsif (lc$element[0] eq "species")
            {$octu{$element[0]} = $element[1];
            }
        }#foreach bracket

# assess "no rank" elements to three categories in sets of three elements,
# with precedence at end of element (i.e. last "no ranks" printed first")
if ($oc[$#oc - 8] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 8]);
        $octu{$miscElem[0]."c"} = $miscElem[1];
        }
if ($oc[$#oc - 7] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 7]);
        $octu{$miscElem[0]."c"} = $miscElem[1];
        }
if ($oc[$#oc - 6] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 6]);
        $octu{$miscElem[0]."c"} = $miscElem[1];
        }
if ($oc[$#oc - 5] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 5]);
        $octu{$miscElem[0]."c"} = $miscElem[1];
        }
if ($oc[$#oc - 4] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 4]);
        $octu{$miscElem[0]."b"} = $miscElem[1];
        }
if ($oc[$#oc - 3] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 3]);
        $octu{$miscElem[0]."b"} = $miscElem[1];
        }
if ($oc[$#oc - 2] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 2]);
        $octu{$miscElem[0]."a"} = $miscElem[1];
        }
if ($oc[$#oc - 1] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc - 1]);
        $octu{$miscElem[0]."a"} = $miscElem[1];
        }
```

```
if ($oc[$#oc] =~ /no rank:.*/i)
        {my @miscElem = split(":", $oc[$#oc]);
        $octu{$miscElem[0]."a"} = $miscElem[1];
        }

        if($octu{"OCTU"})
        {print OUT $octu{"OCTU"}, "\t";
        print ALTOUT $octu{"OCTU"}, "\t";
        }else
        {print OUT "", "\t";
        }
        if($octu{"GI"})
        {print ALTOUT $octu{"GI"}, "\t";
        }else
        {print ALTOUT "no_hit\t";
        }
        if($octu{"bitScore"})
        {print ALTOUT $octu{"bitScore"}, "\t";
        }else
        {print ALTOUT "", "\t";
        }
        if($octu{"eValue"})
        {print ALTOUT $octu{"eValue"}, "\t";
        }else
        {print ALTOUT "", "\t";
        }
        if($octu{"%ID"})
        {print ALTOUT $octu{"%ID"}, "\t";
        }



        $octu{"taxa"} = "Root;";

        if($octu{"kingdom"})
        {$octu{"taxa"} .= $octu{"kingdom"}.";";
        }
#       if($octu{"subkingdom"})
#       {$octu{"taxa"} .=$octu{"subkingdom"}.";";
#       }
        if($octu{"phylum"})
        { $octu{"taxa"} .=$octu{"phylum"}.";";
        }
        elsif($octu{"subphylum"})
        { $octu{"taxa"} .=$octu{"subphylum"}.";";
```

```
        }
        else
        { $octu{"taxa"} .= "incertae_sedis;";
        }

        if($octu{"class"})
        {$octu{"taxa"} .= $octu{"class"}.";";
        }
        elsif($octu{"subclass"})
        {$octu{"taxa"} .=$octu{"subclass"}.";";
        }
        else
        { $octu{"taxa"} .= "incertae_sedis;";
        }

        if($octu{"order"})
        {$octu{"taxa"} .= $octu{"order"}.";";
        }
        else
        { $octu{"taxa"} .= "incertae_sedis;";
        }

        if($octu{"family"})
        {$octu{"taxa"} .= $octu{"family"}.";";
        }
        else
        { $octu{"taxa"} .= "incertae_sedis;";
        }

        if($octu{"genus"})
        {$octu{"taxa"} .= $octu{"genus"}.";";
        }
        elsif($octu{"no ranka"})
        {$octu{"taxa"} .= $octu{"no ranka"}.";";
        }
        else
        { $octu{"taxa"} .= "incertae_sedis;";
        }
#       if($octu{"species"})
#       {$octu{"taxa"} .= $octu{"species"}.";";
#       }

#       if($octu{"no ranka"})
#       {$octu{"taxa"} .= $octu{"no ranka"}.";";
#       }
```

```
#       if($octu{"no rankb"})
#       {$octu{"taxa"} .= $octu{"no rankb"}.";";
#       }
#
#       if($octu{"no rankc"})
#       {$octu{"taxa"} .= $octu{"no rankc"}.";";
#       }

            if ($octu{"taxa"} =~ /.*samples.*/g || $octu{"taxa"} =~
/.*unclassified.*/g)
                {
                my @splitTax = split(";", $octu{"taxa"});
                $octu{"taxa"} =~ s/incertae_sedis;/$splitTax[6];/g;

                }
            if ($octu{"taxa"} !~ /.*samples.*/g || $octu{"taxa"} !~
/.*unclassified.*/g)
                {
                my @splitTax = split(";", $octu{"taxa"});
                if($splitTax[6] !~ /incertae_sedis/ && $splitTax[6])
                    {
                    $octu{"taxa"} =~
s/incertae_sedis/$splitTax[6]_incertae_sedis/g;
                    }
                elsif($splitTax[5] !~ /incertae_sedis/)
                    {
                    $octu{"taxa"} =~
s/incertae_sedis/$splitTax[5]_incertae_sedis/g;
                    }
                elsif($splitTax[4] !~ /incertae_sedis/)
                    {
                    $octu{"taxa"} =~
s/incertae_sedis/$splitTax[4]_incertae_sedis/g;
                    }
                elsif($splitTax[3] !~ /incertae_sedis/)
                    {
                    $octu{"taxa"} =~
s/incertae_sedis/$splitTax[3]_incertae_sedis/g;
                    }
                elsif($splitTax[2] !~ /incertae_sedis/)
                    {
                    $octu{"taxa"} =~
s/incertae_sedis/$splitTax[2]_incertae_sedis/g;
                    }
                elsif($splitTax[1] !~ /incertae_sedis/)
                    {
```

```
                                $octu{"taxa"} =~
s/incertae_sedis/$splitTax[1]_incertae_sedis/g;
                                }
                        }
                #$octu{"taxa"} =~ s/;$//g;
                $octu{"taxa"} =~ s/ /_/g;
                $octu{"taxa"} .= $octu{"OCTU"};
                print OUT $octu{"taxa"}, "\t";
                print ALTOUT $octu{"taxa"}, "\n";

                if($octu{"%ID"})
                {printf OUT ("%.3f\n", $octu{"%ID"});
                }else
                {print OUT "NA", "\n";
                }
}#while bracket
```

combineITS1_ITS2_otu_table.pl

```
#!/usr/bin/perl
# Eric Morrison
# 3/19/12
# This script combines ITS1 and ITS2 otu_tables with 9 samples. Samples
names are designated with a 1_ or 2_ and ITS2 otu numbers are renamed as
"original OTU # + 10000."
use strict;
use warnings;

my $ITS1=$ARGV[0];
my $ITS2=$ARGV[1];
my $out=$ARGV[2];

open (ITS1, "$ITS1") || die "Can't open ITS1.\n";
open (ITS2, "$ITS2") || die "Can't open ITS2.\n";
open (OUT, ">$out") || die "Can't open out.\n";

chomp(my@its1=<ITS1>);
chomp(my@its2=<ITS2>);

print OUT "# QIIME v1.4.0 OTU table\n";
print OUT "#OTU
ID\t1_ctrl.d2\t1_ctrl.e3\t1_ctrl.e5\t1_low.b3\t1_low.c1\t1_low.d2\t1_high.b3\t1_hig
h.d2\t1_high.d6\t2_ctrl.d2\t2_ctrl.e3\t2_ctrl.e5\t2_low.b3\t2_low.c1\t2_low.d2\t2_h
igh.b3\t2_high.d2\t2_high.d6\tConsensus Lineage\n";


foreach my $i1 (@its1)
        {
        if ($i1=~/#.*/)
                {
                next;
                }
        my@i1=split("\t", $i1);
        my$tax=pop@i1;
        foreach my$count(@i1)
                {
                print OUT $count, "\t";
                }
        print OUT "0\t" x 9, $tax, "\n";
        }
foreach my $i2 (@its2)
        {
        if ($i2=~/#.*/)
```

```
        {
        next;
        }
my@i2=split("\t", $i2);
my$num=shift@i2;
my$tax=pop@i2;
$num+=10000;
print OUT $num, "\t", "0\t" x 9;
foreach my$count(@i2)
        {
        print OUT $count, "\t";
        }
print OUT $tax, "\n";
}
```

autoRunCombineITS1_ITS2_otu_table.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/21/12
# Usage: perl autoRunCombineITS1_ITS2_otu_table.pl
# This script automates combining ITS1 and ITS2 otu tables after removal of
chimeras and June sequences.

use strict;
use warnings;

#print "input loci, comma separated\n";
#chomp(my $loci = <STDIN>);
#my @loci = split(",",$loci);


print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

print "input otu_table name\n";
chomp(my $table = <STDIN>);

system "mkdir working/ITS1_ITS2_combined/";

#foreach my $locus (@loci)
#	{
	foreach my $simi (@sim)
		{
		system "mkdir working/ITS1_ITS2_combined/$simi";
		system "perl scripts/QiimeScripts/combineITS1_ITS2_otu_table.pl
working/ITS1/$simi/$table working/ITS2/$simi/$table
working/ITS1_ITS2_combined/$simi/$table";
		}
#	}
```

APPENDIX E. Scripts for chimera removal. Chimeras were identified with the uchime algorithm of usearch (Edgar RC, 2010) after the writeUclustConsensus.pl step. Data was then run through the subsequent scripts.

writeUclustConsensus.pl

```
#!/usr/bin/perl
# writeUclustConsensus.pl
# Eric Morrison
# 7/5/11
# This script takes a representative sequence file from
# qiime and an otus file from qiime (mothur format) and writes
# a uClust consensus sequences file.

use strict;
use warnings;

sub usage()
        {
        print STDERR q
  (

Usage: writeUclustConsensus.pl <otus.txt> <rep_set.fna>

This script takes a otu mapping file from qiime and a qiime representative
sequence fasta file and returns a representative sequence fasta file in usearch
format.

);
exit;

        }

if ( ( @ARGV == 0 ) || $ARGV[0] eq "-h" )   #no arguments or help option
{
    &usage();
}

my $qOtus = $ARGV[0];
my $qRepset = $ARGV[1];

open (OTUS, "$qOtus") || die "Cant open otu file.\n";
open (REP, "$qRepset") || die "Cant open rep seq file.\n";
open (DEST, ">$qRepset.uclust") || die "Cant open dest file.\n";
```

```
chomp(my @otus = <OTUS>);
chomp(my @rep = <REP>);
my %rep = @rep;

foreach my $otu (@otus)
	{
	my @seqs = split("\t", $otu);
	my $num = shift@seqs;
	my $size = scalar(@seqs);
	foreach my $head(keys %rep)
		{
		my @repHead = split(" ", $head);
		$repHead[0] =~ s/>//;
		if ($num == $repHead[0])
			{
			print DEST ">Cluster$num;size=$size\n$rep{$head}\n";
			}
		}
	}
```

writeUchimeraQiime.pl

```
#!/usr/bin/perl
# This script takes a chimera file from usearch and a qiime otu table file
# as input and returns output with the chimera flag prepended.
# Title: writeUchimeraQiime.pl
# Author: Eric Morrison
# Date: 6/11
# Usage: writeUchimeraQiime.pl [chimera] [qiime otu table] [output.name]

use strict;
use warnings;

sub usage()
{
  print STDERR q
  (
Usage: writeUchimeraQiime.pl [chimera] [qiime otu table] [output.name]


);
  exit;
}

if ( ( @ARGV == 0 ) || $ARGV[0] eq "-h" )   #no arguments or help option
{
    &usage();
}

my $chim = $ARGV[0];
my $otus = $ARGV[1];
my $out = $ARGV[2];

open (CHIMERA, "$chim") || die "Cant open chimera file.\n";
open (OTU, "$otus") || die "Cant open parse file.\n";
open (OUT, ">$out") || die "Cant open out file.\n";

chomp(my @chims = <CHIMERA>);
chomp(my @otu = <OTU>);

my $firstLine = shift@otu;
my $header = shift(@otu);
#print $header, "\n";
print OUT "$firstLine\nChimera\t$header\n";
#print $header, "\n";
```

```
foreach my$ocs(@chims)
        {
        my @oc = split("\t", $ocs);
        $oc[1] =~ /Cluster(\d+);size=.*/;
        my $num = $1;
        #print $num, "\n";
        foreach my$line(@otu)
                {
                if ($line =~ /^#/)
                        {
                        next;
                        }
                my @line = split("\t", $line);
                #print $line[0], "\n";

                if($num =~ /^$line[0]$/)
                        {
                        unshift(@line, $oc[16]);
                        foreach(@line)
                                {print OUT $_, "\t";
                                }
                        print OUT "\n";
                        next;
                        }
                }
        }
```

rmRepSetChimeraQiime.pl

```
#!/usr/bin/perl
# This script takes a chimera file from usearch and a Qiime rep seqs file
# as input and returns a fasta with sequences that were not flagged as chimeras.
# Title: rmRepSetChimeraQiime.pl
# Author: Eric Morrison
# Date: 12/11
# Usage: rmRepSetChimeraQiime.pl [chimera] [qiime rep set] [output.name]

use strict;
use warnings;

sub usage()
{
  print STDERR q
  (
Usage: writeSeq.pl [chimera] [qiime rep set] [output.name]

        This script removes chimeric sequences from a Qiime rep set file.

);
  exit;
}

if ( ( @ARGV == 0 ) || $ARGV[0] eq "-h" )  #no arguments or help option
{
    &usage();
}

my $chim = $ARGV[0];
my $otus = $ARGV[1];
my $out = $ARGV[2];

open (CHIMERA, "$chim") || die "Cant open chimera file.\n";
open (OTU, "$otus") || die "Cant open parse file.\n";
open (OUT, ">$out") || die "Cant open out file.\n";

chomp(my @chims = <CHIMERA>);
chomp(my @otu = <OTU>);
my %otu = @otu;

foreach my$ocs(@chims)
        {
        my @oc = split("\t", $ocs);
        $oc[1] =~ /Cluster(\d+);size=.*/;
```

```
my $num = $1;
#print $num, "\n";
foreach my$head(keys %otu)
        {

        $head =~ />(\d*) .*/;
        my $ocNum = $1;
        print $ocNum, "\n";

        if($num =~ /^$ocNum$/)
                {
                if ($oc[16] eq "N")
                        {
                        print OUT $head, "\n", $otu{$head}, "\n";
                        }
                }
        }
}
```

rmOtuTableChimeraQiime.pl

```
#!/usr/bin/perl
# This script takes a chimera file from usearch and a qiime otu table file
# as input and returns output with chimeras removed.
# Title: rmOtuTableChimeraQiime.pl
# Author: Eric Morrison
# Date: 6/11
# Usage: writeUchimeraQiime.pl [chimera] [qiime otu table] [output.name]

use strict;
use warnings;

sub usage()
{
  print STDERR q
  (
Usage: writeUchimeraQiime.pl [chimera] [qiime otu table] [output.name]


);
  exit;
}

if ( ( @ARGV == 0 ) || $ARGV[0] eq "-h" )   #no arguments or help option
{
    &usage();
}

my $chim = $ARGV[0];
my $otus = $ARGV[1];
my $out = $ARGV[2];

open (CHIMERA, "$chim") || die "Cant open chimera file.\n";
open (OTU, "$otus") || die "Cant open parse file.\n";
open (OUT, ">$out") || die "Cant open out file.\n";

chomp(my @chims = <CHIMERA>);
chomp(my @otu = <OTU>);

my $firstLine = shift@otu;
my $header = shift(@otu);
#print $header, "\n";
print OUT "$firstLine\n$header\n";
#print $header, "\n";
```

```
foreach my$ocs(@chims)
	{
	my @oc = split("\t", $ocs);
	$oc[1] =~ /Cluster(\d+);size=.*/;
	my $num = $1;
	#print $num, "\n";
	foreach my$line(@otu)
		{
		if ($line =~ /^#/)
			{
			next;
			}
		my @line = split("\t", $line);
		#print $line[0], "\n";

		if($num == $line[0])
			{
			if ($oc[16] eq "N")
			{
			print OUT $line;
		#	foreach(@line)
		#		{print OUT $_, "\t";
		#		}
			print OUT "\n";
			next;
			}
			}
		}
	}
```

rearrangeOTUtableRmJune.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/6/12
# This script rearranges the order of sample columns in an otu table from QIIME.
# Usage: rearrangeOTUtableRmJune.pl [otu_table] [output.name]

use strict;
use warnings;

sub usage()
	{
	print STDERR q (

Usage: rearrangeOTUtableRmJune.pl [otu_table] [output.name]

This script rearranges the order of sample columns in an otu table from QIIME.

)
}

if ((@ARGV == 0) | $ARGV[0] eq "-h")
	{
	&usage;
	}

my $in = $ARGV[0];
my $out = $ARGV[1];

open(IN, "$in") || die "Can't open input.\n";
open(OUT, ">$out") || die "Can't open output.\n";

chomp(my@otuTable=<IN>);

my$firstLine=shift@otuTable;
print OUT $firstLine, "\n";
foreach my$otu(@otuTable)
	{
	my@otu=split("\t",$otu);
	my@reOTU;

	$reOTU[0]=$otu[0];
	#if ($otu[0] == 0)
	#	{
	#	$reOTU[0] = '0';
```

```
#       }
        splice(@reOTU,1,3,$otu[2],$otu[3],$otu[4]);
        splice(@reOTU,4,3,$otu[10],$otu[11],$otu[12]);
        splice(@reOTU,7,3,$otu[6],$otu[7],$otu[8]);
#       $reOTU[10] = $otu[1];
#       $reOTU[11] = $otu[9];
#       $reOTU[12] = $otu[5];
        $reOTU[10] = $otu[13];

#       if($reOTU[0] =~ /\#.*/)
#               {
#               foreach(@reOTU)
#                       {
#                       print OUT $_, "\t"
#                       }
#               print OUT "\n";
#               }
#       my $sum = 0;
#       my$i;
#       for($i=1, $i++, $i<10)
#               {
#               $sum += $reOTU[$i];
#               }
#       if ($sum == 0)
#               {
#               next;
#               }
        foreach my$new(@reOTU)
                {
                print OUT $new, "\t";
                }
        print OUT "\n";
        }
```

autoRunRmChimeraJune.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/21/12
# Usage: perl autoRunRmChimeraJune.pl
# This script automates sample rearrangement and removal of chimeras and
unwanted samples from Qiime otu tables.

use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);


print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
        {
        foreach my $simi (@sim)
                {
                system "perl scripts/QiimeScripts/rmOtuTableChimeraQiime.pl
working/$locus/$simi/chime working/$locus/$simi/otu_table.txt
working/$locus/$simi/otu_table_no_chimera.txt";
                system "perl scripts/QiimeScripts/rearrangeOTUtableRmJune.pl
working/$locus/$simi/otu_table_no_chimera.txt
working/$locus/$simi/otu_table_Nov_no_chimera.txt";
                }
        }
```

APPENDIX F. ANOVA scripts. Scripts were run on a table of relative abundances in the following order.

multANOVAR.pl

```
# 7/15/11
# This script uses a modified taxa summary metadata file from qiime
# and accesses R to perform ANOVA  and Tukey's HSD on all
# data points--reports data summary, means, and Tukey's HSD.
# All metadata columns should be removed from metadata file,
# except sample ID and one column should be added that denotes treatment
# as lowercase letters (i.e. "a, b, c").
# Usage: multipleANOVAR.pl [metadataRelAbd.txt] [outputName] [part1 = 1|part2
= 2]

use strict;
use warnings;

sub usage(){
        print STDERR q
        (

Usage part 1: multipleANOVAR.pl [datafile] [output file for R code] [1]

Copy and paste the contents of output file for R code into R

Usage part 2: multipleANOVAR.pl [datafile] [output file for data summary] [2]


Part one of this script takes a data type by sample data file (e.g. variable name in
rows samples names in columns) as input and writes R code to perform ANOVA
and Tukey's HSD on all data points.

Input data should be in the format sample ID, one column that denotes treatment
as lowercase letters (i.e. "a, b, c"), and variables to be assessed in subsequent
columns with the name of the data in the first row of the column.

The ouput from part one can be copy and pasted directly into an R window. R will
perform ANOVA and Tukey's HSD on all data.

The user can then run part two of this script to concatenate the results from R.
```

```
* are used to denote significance values in the concatenated results file for
human readability. * 0.1>=P>0.5; ** 0.5>=P>0.01; *** 0.01>=P>0.001; ****
0.001>=P

);
exit;
}
if((@ARGV==0)||$ARGV[0]eq"-h")
        {
        &usage();
        }

my $metaDat = $ARGV[0];
my $out = $ARGV[1];
my $part = $ARGV[2];
if($part eq "")
        {
        print "You did not pick part one or two. Please select an option.\n";
        exit;
        }

open(MET, "$metaDat") || die "Can't open metadata.\n";
open(DEST, ">$out") || die "Can't open OUT!!!!!\n";


# Reads data from qiime metadata file. Need this for both parts
chomp(my @met = <MET>);
#$met =~ s/\r\n|\r|\n/:::::/g;

#my @met = split(":::::",$met);

my @lines = @met;
for(my $i = 0;$i < @met; $i++)
        {
        my @line = split("\t", $met[$i]);
        $lines[$i] = [ @line ];
        }


# Make directories for R input and R output. Part 1 only.
if ($part == 1)
        {
        system "rm -R Rin\n";
        system "rm -R Rout\n";
        system "mkdir Rin\n";
        system "mkdir Rout\n";
```

```
        for(my $u=2;$u<@{$lines[1]};$u++)
                {

                # Print input data file for R one for each entry
                open(TEMP, ">Rin/R.$u.txt") || die "Cant open temp dest\n";
                for(my $n = 0; $n < @lines; $n++)
                        {
                        print TEMP $lines[$n][0], "\t", $lines[$n][1], "\t",
$lines[$n][$u], "\n";
                        }

                # Print to R parameters file. New data is appended to existing file
                system "touch Rout/tukes.$u.txt\n";
                system "touch Rout/means.$u.txt\n";
                print DEST "data.$u = read.table(\"Rin/R.$u.txt\",header=F)\n",
                "aov.$u = aov(V3~V2,data=data.$u)\n",
                "Tukes.$u = TukeyHSD(aov.$u)\n",
                "sink(file = \"Rout/tukes.$u.txt\")\n",
                "Tukes.$u\nsink()\n",
                "Means.$u = model.tables(aov.$u, \"means\")\n",
                "sink(file = \"Rout/means.$u.txt\")\n",
                "Means.$u\nsink()\n";
                }
        #print DEST "q()\nn\n";
        }


# Test R output files and check for significant values
if ($part == 2)
        {
        for(my $u=2;$u<@{$lines[1]};$u++)
                {
                open(TEMPTUKES, "Rout/tukes.$u.txt") || die "Can't open R tukes
file $u.\n";
                chomp(my @rTukes = <TEMPTUKES>);
                print DEST $lines[0][$u], "\n";
                print DEST "\tdiff\tlwr\tupr\tp adj\n";
                splice(@rTukes,0,7);
                pop @rTukes;
                foreach(@rTukes)
                        {
                        my @sig = split(" ", $_);
                        if ($sig[$#sig] <= 0.05 && $sig[$#sig] > 0.01)
                                {
                                push(@sig, '**');
                                }
```

```
                elsif($sig[$#sig] <= 0.01 && $sig[$#sig] > 0.001)
                        {
                        push(@sig, '***');
                        }
                elsif($sig[$#sig] <= 0.001)
                        {
                        push(@sig, '****');
                        }
                elsif($sig[$#sig] <= 0.1 && $sig[$#sig] > 0.05)
                        {
                        push(@sig, '*');
                        }
                foreach(@sig)
                        {
                        print DEST "$_\t";
                        }
                print DEST "\n";
                }
        print DEST "\n";
        open(TEMPMEAN, "Rout/means.$u.txt") || die "Can't open R
means file $u.\n";
        chomp(my @rMeans = <TEMPMEAN>);
        splice (@rMeans,0,7);
        foreach(@rMeans)
                {
                print DEST $_, "\n";
                }
        print DEST "\n";
        }
}
```

parseMultANOVAForChart3sitesITS1ITS2.pl

```
#!/usr/bin/perl
# Eric Morrison
# 3/19/12

use strict;
use warnings;

my$in=$ARGV[0];
my$out=$ARGV[1];

open(IN, "$in") || die "Can't open input.\n";
open(OUT, ">$out") || die "Can't open output.\n";

chomp(my@in=<IN>);
print OUT "control\tlow N\thigh N\tb-a\tc-a\tc-b\ttaxonomy\n";
while(@in)
	{
	my $taxonomy = $in[0];
	my $ba = $in[2];
	my $ca = $in[3];
	my $cb = $in[4];
	#foreach (@sigs)
	#	{
	#	print $_, "\n";
	#	}
	#print "\n\n\n\n\n";
	splice(@in,0,6); #####THIS LINE CONTROLS HOW MANY LINES ARE
SPLICESD FROM EACH ENTRY
	my@avgs=split(" ",$in[0]);
	foreach(@avgs)
		{
		if($_ > 0.000000000000000001)
			{
			print OUT $_, "\t";
			}else{
			print OUT "0", "\t";
			}
		}
	my @ba = split(" ", $ba);
#	foreach(@ba)
#		{
#		print $_, "\n";
#		}
	print OUT $ba[5], "\t";
```

```
        my @ca = split(" ", $ca);
        print OUT $ca[5], "\t";
        my @cb = split(" ", $cb);
        print OUT $cb[5], "\t";
#       print OUT "\n";
        my @taxonomy = split(";", $taxonomy);
        foreach (@taxonomy)
                {
                        print OUT $_,"\t";
                }
        print OUT "\n";
        splice(@in,0,3);
        }
```

parseMultANOVAForChartOnlySig3sitesITS1ITS2.pl

```
#!/usr/bin/perl
# Eric Morrison
# 3/19/12

use strict;
use warnings;

my$in=$ARGV[0];
my$out=$ARGV[1];

open(IN, "$in") || die "Can't open input.\n";
open(OUT, ">$out") || die "Can't open output.\n";

chomp(my@in=<IN>);
print OUT "control\tlow N\thigh N\tb-a\tc-a\tc-b\ttaxonomy\n";
while(@in)
	{

	my $ba = $in[2];
	my $ca = $in[3];
	my $cb = $in[4];
	my @ba = split(" ", $ba);
	my @ca = split(" ", $ca);
	my @cb = split(" ", $cb);
	if($ca[5] || $ba[5] || $cb[5])
		{

		my $taxonomy = $in[0];
		splice(@in,0,6);
		my@avgs=split(" ",$in[0]);


		foreach(@avgs)
			{
			if($_ > 0.000000000000000001)
				{
				print OUT $_, "\t";
				}else{
				print OUT "0", "\t";
				}
			}

		print OUT $ba[5], "\t";
```

```
                print OUT $ca[5], "\t";

                print OUT $cb[5], "\t";
#       print OUT "\n";
                my @taxonomy = split(";", $taxonomy);
                foreach (@taxonomy)
                        {
                                print OUT $_,"\t";
                        }
                splice(@in,0,3);
                print OUT "\n";
                }
        else{
                splice(@in,0,9);
                }
        }
```

APPENDIX G. Scripts for automation of operations in the QIIME software package run on the Amazon AWS Cloud server.

autoRunQiimeAlphaDivCharts.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/22/12

use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input alpha diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
	{
	print "input $locus max for rarefactions\n";
	chomp(my$max=<STDIN>);

	foreach my $simi (@sim)
		{
		system "multiple_rarefactions.py -i $locus/$simi/otu_table* -o
$locus/$simi/rarefactions/ -m 10 -x $max -s 10";
		foreach my $metric (@metric)
			{
			system "alpha_diversity.py -i $locus/$simi/rarefactions/ -o
$locus/$simi/$metric/ -m $metric";
			system "collate_alpha.py -i $locus/$simi/$metric/ -o
$locus/$simi/$metric"."_collated";
```

```
                        system "make_rarefaction_plots.py -i
$locus/$simi/$metric"."_collated -m metadata".$locus."* -b Treatment,SampleID -
o $locus/$simi/$metric"."_plots";
                        }
                }
        }
```

autoRunQiimeBetaDivCharts2D.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/22/12



use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

print "need to rarify? (y|n)\n";
chomp(my$rareYN = <STDIN>);

print "need to make tree? (y|n)\n";
chomp(my $tre=<STDIN>);

my $max;

foreach my $locus (@loci)
	{
	if ($rareYN eq "y")
		{
		print "input $locus max for rarefactions\n";
		chomp($max=<STDIN>);
		}
	foreach my $simi (@sim)
		{
		if ($rareYN eq "y")
			{
			system "multiple_rarefactions.py -i
$locus/$simi/otu_table_Nov* -o $locus/$simi/rarefactions/ -m 10 -x $max -s 10";
			}
		if ($tre eq "y")
```

```
                        {
                        system "make_phylogeny.py -i
$locus/$simi/rep_set_no_chimera_aligned.fasta -o
$locus/$simi/rep_set_no_chimera_aligned.tre";
                        }
                foreach my $metric (@metric)
                        {
                        if ($metric =~ /.*unifrac/)
                                {
                                system "beta_diversity.py -i $locus/$simi/rarefactions/
-o $locus/$simi/$metric/ -m $metric -t
$locus/$simi/rep_set_Nov_no_chimera_*.ph";
                                }
                        else
                                {
                                system "beta_diversity.py -i $locus/$simi/rarefactions/
-o $locus/$simi/$metric/ -m $metric";
                                }

                        system "principal_coordinates.py -i $locus/$simi/$metric/ -o
$locus/$simi/$metric"."_pcoa";
                        system "make_2d_plots.py -i $locus/$simi/$metric"."_pcoa -
m metadata$locus".".txt -b Treatment -o $locus/$simi/$metric"."_2d";
                        }
                }
        }
```

autoRunQiimeBetaDivCharts3dTaxa.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/22/12



use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

print "need to rarify? (y|n)\n";
chomp(my$rareYN = <STDIN>);

print "need to make tree? (y|n)\n";
chomp(my $tre=<STDIN>);

my $max;

foreach my $locus (@loci)
	{
	if ($rareYN eq "y")
		{
		print "input $locus max for rarefactions\n";
		chomp($max=<STDIN>);
		}
	foreach my $simi (@sim)
		{
		if ($rareYN eq "y")
			{
			system "multiple_rarefactions.py -i $locus/$simi/otu_table* -o
$locus/$simi/rarefactions/ -m 10 -x $max -s 10";
			}
		if ($tre eq "y")
```

```
                        {
                        system "make_phylogeny.py -i
$locus/$simi/rep_set_no_chimera.aln -t muscle -o $locus/$simi/";
                        }
                foreach my $metric (@metric)
                        {
                        if ($locus eq "LSU")
                                {
                                system "beta_diversity.py -i $locus/$simi/rarefactions/
-o $locus/$simi/$metric/ -m $metric -t
$locus/$simi/rep_set_Nov_no_chimera*.ph";
                                }
                        else
                                {
                                system "beta_diversity.py -i $locus/$simi/rarefactions/
-o $locus/$simi/$metric/ -m $metric";
                                }
                        system "summarize_taxa.py -i $locus/$simi/otu_table* -L 6 -
o $locus/$simi/";
                        system "principal_coordinates.py -i $locus/$simi/$metric/ -o
$locus/$simi/$metric"."_pcoa";
                        system "make_3d_plots.py -t $locus/$simi/otu_table*_L6.txt -
i $locus/$simi/$metric"."_pcoa -m metadata$locus".".txt -o
$locus/$simi/$metric"."_3d";
                        }
                }
        }
```

autoRunQiimeBetaDivChartsANOSIM.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/22/12



use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

#print "need to rarify? (y|n)\n";
#chomp(my$rareYN = <STDIN>);

print "need to make tree? (y|n)\n";
chomp(my $tre=<STDIN>);

my $max;

foreach my $locus (@loci)
	{

	foreach my $simi (@sim)
		{
		print "input rarefaction depth for $locus/$simi\n";
		chomp(my$depth=<STDIN>);
		system "single_rarefaction.py -i $locus/$simi/otu_table_Nov* -o
$locus/$simi/otu_table_Nov_no_chimera_rare.txt -d $depth";
		if ($tre eq "y")
				{
				system "make_phylogeny.py -i
$locus/$simi/rep_set_no_chimera_aligned.fasta -o
$locus/$simi/rep_set_no_chimera_aligned.tre";
				}
```

```
                foreach my $metric (@metric)
                        {
                        if ($metric =~ /.*unifrac/)
                                {
                                system "beta_diversity.py -i
$locus/$simi/otu_table_Nov_no_chimera_rare.txt -o $locus/$simi/$metric/ -m
$metric -t $locus/$simi/rep_set_Nov_no_chimera_*.ph";
                                }
                        else
                                {
                                system "beta_diversity.py -i
$locus/$simi/otu_table_Nov_no_chimera_rare.txt -o $locus/$simi/$metric/ -m
$metric";
                                }
                                system "compare_categories.py -i
$locus/$simi/$metric/$metric"."_otu_table_Nov_no_chimera_rare.txt -m
metadata$locus".".txt -c Treatment --method adonis -n 1000 -o
$locus/$simi/$metric"."_ADONIS";
                                system "compare_categories.py -i
$locus/$simi/$metric/$metric"."_otu_table_Nov_no_chimera_rare.txt -m
metadata$locus".".txt -c Treatment --method mrpp -n 1000 -o
$locus/$simi/$metric"."_MRPP";
                                system "compare_categories.py -i
$locus/$simi/$metric/$metric"."_otu_table_Nov_no_chimera_rare.txt -m
metadata$locus".".txt -c Treatment --method anosim -n 1000 -o
$locus/$simi/$metric"."_ANOSIM";                        }
                }
        }
```

autoRunQiimeBetaDivChartsDistanceHistograms.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/22/12



use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

#print "need to rarify? (y|n)\n";
#chomp(my$rareYN = <STDIN>);

print "need to make tree? (y|n)\n";
chomp(my $tre=<STDIN>);

my $max;

foreach my $locus (@loci)
	{

	foreach my $simi (@sim)
		{
		print "input rarefaction depth for $locus/$simi\n";
		chomp(my$depth=<STDIN>);
		system "single_rarefaction.py -i $locus/$simi/otu_table_Nov* -o
$locus/$simi/otu_table_Nov_no_chimera_rare.txt -d $depth";
		if ($tre eq "y")
				{
				system "make_phylogeny.py -i
$locus/$simi/rep_set_no_chimera_aligned.fasta -o
$locus/$simi/rep_set_no_chimera_aligned.tre";
				}
```

```
		foreach my $metric (@metric)
			{
			if ($metric =~ /.*unifrac/)
				{
				system "beta_diversity.py -i
$locus/$simi/otu_table_Nov_no_chimera_rare.txt -o $locus/$simi/$metric/ -m
$metric -t $locus/$simi/rep_set_Nov_no_chimera*.ph";
				}
			else
				{
				system "beta_diversity.py -i
$locus/$simi/otu_table_Nov_no_chimera_rare.txt -o $locus/$simi/$metric/ -m
$metric";
				}

			system "make_distance_histograms.py -d
$locus/$simi/$metric/* -m metadata".$locus.".txt --fields SampleID,Treatment --
monte_carlo_iters 1000 -o $locus/$simi/$metric"."_distance_histograms";
			}
		}
	}
```

autoRunQiimeCombinedTaxaSummary.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/6/12
# Usage: perl autoRunQiimeTaxa.pl
# This script automates creating QIIME taxa plots and relative abundance
summaries.

use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input taxonomic orders, comma separated\n";
chomp(my$taxa = <STDIN>);
my @taxa = split(",",$taxa);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

#print "input chart type\n";
#chomp(my $chart = <STDIN>);


foreach my $locus (@loci)
        {
                foreach my $simi (@sim)
                        {
                        foreach my $tax (@taxa)
                                {
                                my $l;
                                if ($tax eq "genus")
                                        {
                                        $l = 7;
                                        }elsif($tax eq "family"){
                                        $l = 6;
                                        }elsif($tax eq "order"){
                                        $l = 5;
                                        }elsif($tax eq "OTU"){
                                        $l = 8;
                                        }elsif($tax eq "class"){
                                        $l = 4;
```

```
                                    }elsif($tax eq "phylum"){
                                    $l = 3;
                                    }


                                    #system "summarize_taxa.py -i
$locus/$simi/avg_otu_table_Nov_no_chimera* -L $l -o $locus/$simi/$tax";
                                    system "summarize_taxa.py -i
$locus/$simi/otu_table_Nov_no_chimera* -L $l -o $locus/$simi/$tax -m
metadataRelAbd".$locus.".txt";
                                    #system "plot_taxa_summary.py -i
$locus/$simi/$tax/avg_otu_table* -l $tax -c $chart -o $locus/$simi/$tax/$chart/";
                                }
                            }
        }
```

autoRunQiimeMakeOtuTable.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/21/12
# Usage: perl autoRunMakeOtuTable.pl
# This script automates creating QIIME otu tables.

use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);


print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
	{
	foreach my $simi (@sim)
		{
		system "make_otu_table.py -i $locus/$simi/*_otus.txt -o
$locus/$simi/otu_table.txt -t $locus/$simi/*taxonomyQiime.txt";
		}
	}
```

autoRunQiimePickOtus.pl

```
#!/usr/bin/perl
# Eric Morrison
# 2/6/12
# Usage: perl autoRunQiimePickOtus.pl
# This script automates creating QIIME otu maps and rep sets.

use strict;
use warnings;

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);


print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
	{
	foreach my $simi (@sim)
		{
		my $simVal = $simi/100;
		system "pick_otus.py -i $locus/seqs.fna -o $locus/$simi/otus.txt -s
$simVal --optimal -m uclust";
		system "pick_rep_set.py -i $locus/$simi/$locus"."_anoise_otus.txt -f
$locus/seqs.fna -o $locus/$simi/rep_set.fna";
		}
	}
```

autoDldAlphaDivCharts.pl

```
#!/usr/bin/perl
# Eric Morrison
# 1/13/12



use strict;
use warnings;

print "input public dns\n";
chomp(my$pubdns = <STDIN>);

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input alpha diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);



foreach my $locus (@loci)
        {
        foreach my $simi (@sim)
                {
                foreach my $metric (@metric)
                        {

                        #system "rm -R
aws$locus/$noise/$simi/taxa_summary/$tax";
                        system "mkdir working/$locus/$simi/alpha_div";
                        system "mkdir working/$locus/$simi/alpha_div/$metric";

                        system "mkdir
working/$locus/$simi/alpha_div/$metric/average_plots";
                        system "mkdir
working/$locus/$simi/alpha_div/$metric/average_tables";
                        system "mkdir
working/$locus/$simi/alpha_div/$metric/html_plots";
```

```
                    #system "mkdir
working/$locus/$simi/alha_div/$metric/charts";

                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_plots/*
working/$locus/$simi/alpha_div/$metric/";

                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_plots/average_plots/*
working/$locus/$simi/alpha_div/$metric/average_plots/";
                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_plots/average_tables/*
working/$locus/$simi/alpha_div/$metric/average_tables/";
                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_plots/html_plots/*
working/$locus/$simi/alpha_div/$metric/html_plots/";
                    #system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_plots/charts/*
working/$locus/$noise/$simi/taxa_summary/$tax/charts/";

                    }
                }
        }
```

autoDldBetaDivAnosim.pl

```
#!/usr/bin/perl
# Eric Morrison
# 1/13/12



use strict;
use warnings;

print "input public dns\n";
chomp(my$pubdns = <STDIN>);

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input significance test, comma separated\n";
chomp(my$sig = <STDIN>);
my @sig = split(",",$sig);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
        {
        foreach my $simi (@sim)
                {
                foreach my $metric (@metric)
                        {
                        foreach my $signi (@sig)
                                {
                                system "mkdir working/$locus/$simi/beta_div";
                                system "mkdir working/$locus/$simi/beta_div/$metric";
                                system "mkdir
working/$locus/$simi/beta_div/$metric/$signi";
                                system "mkdir
working/$locus/$simi/beta_div/$metric/$signi/ctrl-low";
```

```
                                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_$signi/*
working/$locus/$simi/beta_div/$metric/$signi/ctrl-low";
                                }
                        }
                }
        }
```

autoDldBetaDivCharts.pl

```
#!/usr/bin/perl
# Eric Morrison
# 1/13/12



use strict;
use warnings;

print "input public dns\n";
chomp(my$pubdns = <STDIN>);

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
        {
        foreach my $simi (@sim)
                {
                foreach my $metric (@metric)
                        {
                        print "input level 1 directory name for
$locus/$simi/$metric"."_3d\n";
                        chomp(my $l1DIR = <STDIN>);
                        print "input level 2 directory name for
$locus/$simi/$metric"."_3d\n";
                        chomp(my $l2DIR = <STDIN>);
                        system "mkdir working/$locus/$simi/beta_div";
                        system "mkdir working/$locus/$simi/beta_div/$metric";

                        system "mkdir working/$locus/$simi/beta_div/$metric/jar";
                        system "mkdir
working/$locus/$simi/beta_div/$metric/$l1DIR";
```

```
                    system "mkdir
working/$locus/$simi/beta_div/$metric/$l2DIR";

                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_3d/*
working/$locus/$simi/beta_div/$metric/";

                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_3d/$l1DIR/*
working/$locus/$simi/beta_div/$metric/$l1DIR/";
                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_3d/$l1DIR/*
working/$locus/$simi/beta_div/$metric/$l2DIR/";

                    system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_3d/jar/*
working/$locus/$simi/beta_div/$metric/jar/";
                    #system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$noise/$simi/$metric"."_3d_noTaxa/html_plots/*
/Users/ericmorrison/aws$locus/$noise/$simi/alpha_div/$metric/html_plots/";
                    #system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$noise/$simi/$metric"."_plots/charts/*
/Users/ericmorrison/aws$locus/$noise/$simi/taxa_summary/$tax/charts/";
                    }
                }
        }
```

autoDIdBetaDivCharts2d.pl

```
#!/usr/bin/perl
# Eric Morrison
# 1/13/12



use strict;
use warnings;

print "input public dns\n";
chomp(my$pubdns = <STDIN>);

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);


foreach my $locus (@loci)
        {
        foreach my $simi (@sim)
                {
                foreach my $metric (@metric)
                        {
                        print "input level 1 directory name for
$locus/$simi/$metric"."_2d\n";
                        chomp(my $l1DIR = <STDIN>);
                        print "input level 2 directory name for
$locus/$simi/$metric"."_2d/$l1DIR/\n";
                        chomp(my $l2DIR = <STDIN>);
                        system "mkdir working/$locus/$simi/beta_div";
                        system "mkdir working/$locus/$simi/beta_div/$metric"."2d";

                        system "mkdir
working/$locus/$simi/beta_div/$metric"."2d/js";
                        system "mkdir
working/$locus/$simi/beta_div/$metric"."2d/$l1DIR";
```

```
                    system "mkdir
working/$locus/$simi/beta_div/$metric"."2d/$l1DIR/$l2DIR";

                    system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_2d/*
working/$locus/$simi/beta_div/$metric"."2d/";

                    system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_2d/$l1DIR/*
working/$locus/$simi/beta_div/$metric"."2d/$l1DIR/";
                    system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_2d/$l1DIR/$l2DIR/*
working/$locus/$simi/beta_div/$metric"."2d/$l1DIR/$l2DIR/";

                    system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_2d/js/*
working/$locus/$simi/beta_div/$metric"."2d/js/";
                    #system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$noise/$simi/$metric"."_3d_noTaxa/html_plots/*
/Users/ericmorrison/aws$locus/$noise/$simi/alpha_div/$metric/html_plots/";
                    #system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$noise/$simi/$metric"."_plots/charts/*
/Users/ericmorrison/aws$locus/$noise/$simi/taxa_summary/$tax/charts/";
                    }
                }
        }
```

autoDIdBetaDivChartsDistanceHistograms.pl

```
#!/usr/bin/perl
# Eric Morrison
# 1/13/12

use strict;
use warnings;

print "input public dns\n";
chomp(my$pubdns = <STDIN>);

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input beta diversity metrics, comma separated\n";
chomp(my$met = <STDIN>);
my @metric = split(",",$met);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

foreach my $locus (@loci)
	{
	foreach my $simi (@sim)
		{
		foreach my $metric (@metric)
			{
			print "input level 1 directory name for
$locus/$simi/$metric"."_distance_histograms\n";
			chomp(my $l1DIR = <STDIN>);

			system "mkdir working/$locus/$simi/beta_div";
			system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms";

			system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms/monte_carlo_grou
p_distances";
			system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms/histograms";

			system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms/js";
```

```
                system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms/$l1DIR";
                system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms/$l1DIR/group_dist
ances_pairs";
                system "mkdir
working/$locus/$simi/beta_div/$metric"."_distance_histograms/$l1DIR/group_dist
ances_single";

                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/";

                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/$l1DIR/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/$l1DIR/";
                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/$l1DIR/group_
distances_pairs/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/$l1DIR/group_dist
ances_pairs/";
                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/$l1DIR/group_
distances_single/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/$l1DIR/group_dist
ances_single/";

                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/js/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/js/";
                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/histograms/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/histograms/";
                system "scp -i qiimeKey/qiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$metric"."_distance_histograms/monte_carlo_g
roup_distances/*
working/$locus/$simi/beta_div/$metric"."_distance_histograms/monte_carlo_grou
p_distances/";

                }
            }
        }
```

autoDldFile.pl

```
#!/usr/bin/perl
# Eric Morrison
# 1/13/12



use strict;
use warnings;

print "input public dns\n";
chomp(my$pubdns = <STDIN>);

print "input loci, comma separated\n";
chomp(my $loci = <STDIN>);
my @loci = split(",",$loci);

print "input %similarity levels, comma separated\n";
chomp(my$sim = <STDIN>);
my @sim = split(",",$sim);

print "input file name\n";
chomp(my $file = <STDIN>);

print "need to make destination directory? (y|n)";
chomp(my $mkdir = <STDIN>);

print "taxonomic directories? y|n\n";
chomp(my $yn = <STDIN>);
if ($yn eq "n")
{

foreach my $locus (@loci)
        {
        foreach my $simi (@sim)
                {
                if ($mkdir eq "y")
                        {
                        system "mkdir working/$locus/$simi";
                        }
                system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$file /Users/ericmorrison/working/$locus/$simi/
";
                }
        }
```

```
}

elsif($yn eq "y")
{

print "Enter taxonomic levels, comma separated\n";
chomp(my$tax = <STDIN>);
my @tax = split(",", $tax);

#my @noise = ("denoised","noisy");

foreach my $locus (@loci)
	{
	foreach my $simi (@sim)
		{
	#	foreach my $noise (@noise)
	#		{
			foreach my $taxon (@tax)
				{
				if ($mkdir eq "y")
					{
					system "mkdir working/$locus/$simi/";
					system "mkdir working/$locus/$simi/$taxon/";
					}
				system "scp -i qiimeKeyNew/newQiime.pem
ubuntu\@$pubdns:~/$locus/$simi/$taxon/$file working/$locus/$simi/$taxon/ ";
				}
	#		}
		}
	}
}
```