GEOTHERMAL SOIL ECOLOGY IN YELLOWSTONE NATIONAL PARK

by

James Francis Meadow

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Land Resources and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

April 2012

APPROVAL

of a dissertation submitted by

James Francis Meadow

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to The Graduate School.

Dr. Catherine A. Zabinski

Approved for the Department of Land Resources and Environmental Sciences

Dr. Tracy M. Sterling

Approved for the The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U. S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

James Francis Meadow

April 2012

ACKNOWLEDGEMENTS

Funding was provided by a GK-12 Graduate Fellowship from the NSF and the Big Sky Institute, a Boyd Evison Graduate Fellowship from the Grand Teton Association, a dissertation fellowship from the Institute on Ecosystems, and also from a USDA NRICG to Dr. Zabinski. I would like to thank Rosie Wallander, Galena Ackerman, John Dore, Christian Klatt, Dana Skorupa, Scott Klingenpeel, Jake Beam and Mark Kozubal for endless help with sample processing, analysis, supplies, and dumb questions. Thanks to Dr. Ylva Lekberg for helping me to step into the world I am now in, and for introducing me to molecular biology. Thanks to the incredibly patient YNP Permitting Staff. Thanks to my graduate committee, Dr. Tim McDermott, Dr. David Roberts, Dr. Cathy Cripps, Dr. Tom Deluca, Dr. Billie Kerans and Dr. Rick Engel for support throughout the process. I especially have my major advisor, Dr. Cathy Zabinski, to thank for helping me maintain balance and perspective during my five years at MSU, and for being a terrific example of quiet stability in academia. Thanks to my lab mates for well-timed discussions on everything *but* science, and to my great friend, Adam Sigler, who will often drop everything when we both need to have good time. Most of all, thanks to my amazing wife, Kelly Meadow, for working with me through all of the challenges of graduate school and for her constant support.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Microbial communities in soil are among the most diverse and species-rich of any habitat, but we know surprisingly little about the factors that structure them. Geothermal soils present unique and relatively unexplored model systems in which to address ecological questions using soil microbial communities, since harsh conditions in these soils exert strong filters on most organisms. This work represents two very different approaches to studying soil ecology in geothermal soils in Yellowstone National Park: 1) Arbuscular mycorrhizal fungal (AMF) communities living in the roots of *Mimulus gutattus* in contrasting plant community types were compared to assess a link in community structure between plants and their AMF symbionts; and 2) soil microbial communities were surveyed across multiple spatial scales in an unstudied diatomaceous biological soil crust in alkaline siliceous geothermal soils, using bar-coded 454 pyrosequencing of 18S and 16S rDNA.

Mycorrhizal communities living in plant roots from contrasting community types showed a striking difference in taxon richness and diversity that appears to transcend soil-chemical differences, though robust conclusions are difficult since plant and fungal communities are structured by some of the same confounding soil conditions. Cluster and discriminant analyses were employed to compare drivers of AMF community structure.

Eukaryotic and prokaryotic communities in a diatomaceous biological soil crust differ significantly from that of an adjacent sinter soil, and along a photic depth gradient. Along with a description of this unique system, extensive multivariate community analyses were used to address outstanding questions of soil microbial community spatial heterogeneity and the methodologies best suited to the unique assumptions of these datasets. Depending on the intended scope of inference, much detail can be gained by investigation of microbial communities at the aggregate or soil particle scale, rather than through composite sampling. Additionally, β-diversity patterns are apparent with relatively few sequences per sample.

CHAPTER 1.

INTRODUCTION

This dissertation explores several aspects of soil microbial communities and their ecology in Yellowstone National Park (YNP) geothermal soils. These soils present, with caveats, elegant model systems in which to address outstanding questions in the field of soil microbial ecology; they are embedded within a mosaic landscape with highly variable conditions that generally exert strong filters on inhabitants, and thus an investigator is able to work with organisms living at the limits of environmental tolerance. In some ways the complex chemical and physical conditions within these systems confound one another in such a way as to complicate study and analysis, while on the other hand, life in these soils is so difficult for potential denizens that the noise and complexity associated with ecological studies in less harsh conditions is cut away to expose the core necessities of life within an interacting community. The work presented herein takes advantage of these remarkable soil environments to better understand ecological interactions, community assembly, and the inherent heterogeneity of soil microbial communities using a variety of molecular methods aimed at sampling and studying environmental DNA. Additionally, ecological statistical methods that were designed for, and have long been used to understand, communities of macroorganisms are increasingly being applied to massive microbial ecology datasets, and

this dissertation presents statistical approaches that are useful in exploring microbial communities.

Geothermal Soils

The research presented in this dissertation was conducted exclusively in the geothermally-influenced soils of Yellowstone National Park (YNP), WY, USA. Geothermal soils present an elegant opportunity to study ecological concepts in a model system context, because abiotic factors dictate harsh conditions for soil organisms and plants in a highly variable mosaic landscape. These soils are often a direct product of geothermal waters and their aqueous chemical constituents. Conditions unique to geothermal soils include elevated soil temperature that increases with depth, periodic inundation of chemically diverse waters, belowground influence from steam and geothermal water, and aqueous saturation of chemical elements that act as soil parent material and drivers of soil-chemical processes.

Channing *et al.* (2004) described an alkaline-silceous geothermal soil, in close proximity and similar to that presented in Chapters 3 & 4 of the present work, that was composed of taphonized plant and microbial biomass and opaline silica precipitate from Si-saturated thermal water, and they also found that this biomass taphonization can occur within the span of a single year (Channing & Edwards 2004). Channing (2001) suggested similar aqueous taphonization as the primary mechanism responsible for the well-preserved fossils in the Rhynie Chert formation. The high Si content in thermal water found in these systems results from geothermally-heated groundwater

moving through high-Si igneous substrata. "Sinter" is the name commonly used to refer to the opaline-siliceous deposits that form when aqueous Si precipitates at the spring surface, which eventually breaks down into soils, while travertine deposits and their mineral soils result from calcareous parent materials in combination with thermal conditions (Rodman *et al.* 1996). All soils used in the present research, with the exception of diatomaceous residuum, are primarily composed of sinter material and are subsequently referred to as such.

Soil-chemical conditions in thermal areas are often highly variable across a relatively fine spatial scale due to source waters that experience different localized subsurface chemical concentrations, groundwater temperature, pressure, and residence time (Morgan 2007). For instance, Burr *et al.* (2005) reported NH_4^+-N concentrations up to 800 mg L^{-1} in a geothermal soil within 2 km of the diatomaceous system reported in Chapters 3 & 4, even though NH_4^+ and NO_3^- concentrations in my study system were below detection limits in the source spring water, and total N levels (as detected by combustion) closely followed a predictable 10:1 C:N ratio for all soils tested herein; full chemical results are presented in relevant chapters and appendices.

Relative to the wide attention garnered by geothermal spring systems in microbial ecology research, little work has been done in the adjacent soils, with a few notable exceptions. Burr *et al.* (2005) studied biologically mediated N transformations in an acidic soil, and found potential for most steps of a soil N-cycle. Geothermal heating events select for a particular subset of the soil microbial community following a shift

of geothermal activities in a forested soil (Norris *et al.* 2002), including a shift to a less complex, and more thermophilic community. A similarly discriminant fungal community was reported by Redman *et al.* (1999), who were able to culture several thermophilic and thermotolerant fungal species from an acidic thermal soil. Geobiology and phototrophic ecology in endolithic microbial communities from geothermal lithic deposits have been investigated (Ciniglia *et al.* 2004; Walker *et al.* 2005; Norris & Castenholz 2006), and advances in molecular biology have been made using thermal soil microbial communities (e.g., Botero *et al.* 2005). Consistent with efforts in extreme geothermal springs, some novel microorganisms have also been cultured and described in thermal soils (e.g., Stott *et al.* 2008). Some work has also been done to explore the more apparent constituents of thermal soils, plants, and their ecological strategies (Stout *et al.* 1997; Stout & Al-Niemi 2002; Tercek & Whitbeck 2004), and arbuscular mycorrhizal fungal communities colonizing thermal plants have also been investigated under a variety of soil conditions (Bunn & Zabinski 2003; Bunn *et al.* 2009; Appoloni *et al.* 2008). Conditions imposed upon plants and soil organisms living within these soils, whether temperature, chemistry, or isolation, clearly select for those specially capable of life in harsh conditions. Microbial communities in this context often contain organisms whose strengths might lie not in fierce competitive ability, but rather in existing beyond the comfort levels of other prospective organisms. The relatively small spatial scale of extreme soil variability is also a major

factor in the selection of these soils as models for soil ecological studies, and this dissertation was able to take advantage of exactly that in two different systems.

Arbuscular Mycorrhizal Fungi

A majority of plants in all terrestrial ecosystems engage in a root symbiosis with arbuscular mycorrhizal fungi (AMF). The fungal component of this symbiosis is a monophyletic group (*Glomeromycota*) containing ≈ 300 described taxa (Schüßler *et al.* 2001; Rosendahl 2008). The primary symbiotic mechanism entails exchange of plant photosynthates for soil P, though plants have also been observed to benefit from increased water and nutrient status, as well as some pathogen resistance (Parniske 2008). AMF have been recognized as obligate symbionts, while mycorrhizal host plants exist on a continuum ranging from facultative to obligate (Klironomos 2003; Johnson *et al.* 2006). All members of the *Glomeromycota* are microscopic and putatively asexual, and they do not produce fruiting bodies. Instead, dispersal depends mainly on the production of subterraneous, and in some cases endorhizal, multinucleate spores, or on extension of fungal hyphae from colonized roots and root fragments (Smith & Read 2008). There is currently no formalized species concept for AMF, and it is unknown whether their high degree of single-spore genetic variation is the result of hetero- or homokaryosis (Pawlowska & Taylor 2004; Bever & Wang 2005); this, along with their coenocytic, anastomosing nature, challenges formal definitions of individuals, populations, and communities (Rosendahl 2008). Mycorrhizal fungi can be especially beneficial for plants living under harsh environmental conditions

(Schechter & Bruns 2008; Bunn *et al.* 2009; Fitzsimons & Miller 2010; Lekberg *et al.* 2011), and, especially relevant to the current work, are often found heavily colonizing thermal-tolerant plants in geothermal soils (Bunn & Zabinski 2003; Appoloni *et al.* 2008).

Given their paramount role in ecosystem function and plant community ecology, AMF are a well-studied group of organisms and the AM symbiosis has often been used in addressing larger ecological questions concerning symbioses and community ecology (e.g., Dumbrell *et al.* 2010; Lekberg *et al.* 2007, 2011). Some studies have shown specificity and conditional benefit in the AM symbiosis (Klironomos 2003; Sanders 2003; Helgason *et al.* 2007; Öpik *et al.* 2009), though some of these observed differences can be linked to taxon-level environmental limitations of either plants or fungi. While most AMF research has revolved around single-species pairings in greenhouse experiments, mixed plant and fungal community studies reveal that the AM symbiosis in natural systems likely influences, and is affected by, plant density, biodiversity, and age structure (van der Heijden *et al.* 1998; Bever 2002; van der Heijden *et al.* 2003; Johnson *et al.* 2003; Schroeder & Janos 2004; van der Heijden 2004; Hausmann & Hawkes 2009, 2010; van de Voorde *et al.* 2010), and seasonal variation and inequality of resources also affect AMF communities (Dumbrell *et al.* 2011; Collins & Foster 2009).

Biological Soil Crusts

Soils in arid ecosystems often exhibit low plant cover due to limited precipitation, and this results in increased solar radiation at the soil surface. These systems are often inhabited by a phototrophically-driven community of soil microbes collectively known as biological soil crusts (BSC) (Belnap & Lange 2003; Belnap 2003; Whitford 2002; Büdel 2005; Rosentreter *et al.* 2007; Ward 2009). BSCs are variously composed of cyanobacteria, lichens, mosses, algae, free-living fungi, and other bacteria and archaea. These assemblages can be conceptually thought of as microbial mat systems, since their cohesive exudative properties effectively hold surface soil particles together in a C- and N-rich matrix of microbial biomass, cyanobacterial filaments, exudates, hyphae, rhizoids, and rhizines. Biological soil crusts are distinct from chemical or physical soil crusts, such as mineral salt evaporites or desert pavements, in that the surface is held together by biotic activity rather than chemical or physical properties. Metabolism of BSC constituents is strongly regulated by pulse-precipitation events, and most BSC organisms are anhydrobiotic (Belnap 2002, 2003; Cardon *et al.* 2008). Organisms associated with BSCs are responsible for a major portion of C and N fixed in some arid systems where plant activity is limited, and this is largely a function of cyanobacterial activity at and near the soil surface (Belnap 2002; Hawkes 2003; Evans & Lange 2003). Cyanobacteria and their exudates give BSC-affected soils most of their recognizable surface texture, and this helps to stabilize otherwise highly erodible desert soil (Chaudhary *et al.* 2009). This visible surface texture, along with soil

type, is diagnostic of BSC microbial community composition (Bowker & Belnap 2008; Rosentreter *et al.* 2007; Belnap & Lange 2003). Given their roles in nutrient cycling, soil-water relations and soil stability, it is not surprising that BSCs influence plant nutrition and communities (Schwartzman & Volk 1989; Harper & Belnap 2001; Hawkes 2003; Housman *et al.* 2007). Organisms associated with BSCs are uniquely suited to life in arid soils, and as such, several novel microorganisms have been cultured from crusted soils (e.g., Gundlapally & Garcia-Pichel 2005; Bates *et al.* 2006). Much work has also gone into characterizing the communities of cyanobacteria associated with BSCs (Garcia-Pichel *et al.* 2001; Schlesinger *et al.* 2003; Gundlapally & Garcia-Pichel 2006; Bhatnagar *et al.* 2008), as well as other constituents of BSCs and associated lichen communities (Soule *et al.* 2009; Bates *et al.* 2011a, 2012). Biological soil crusts also present an opportunity to study ecological concepts in microbial communities, since their constituents are relatively well-studied, and multiple trophic levels and biogeochemical processes co-occur on a relatively fine scale (Bowker *et al.* 2009).

High-Throughput DNA Sequencing of Soil Microbial Communities

Soils hold an unimaginable diversity of microorganisms that vary with environmental conditions at the μm scale (Sylvia *et al.* 2004; Paul 2007; Madigan *et al.* 2008; Fierer & Jackson 2006; Fierer *et al.* 2009a; Fierer & Lennon 2011). These organisms were classically studied with culture-dependent methods, though it is increasingly clear that our ability to grow soil microbes is still limited; dependence

on these techniques historically skewed impressions of microbial biodiversity (Venter *et al.* 2004; Martiny *et al.* 2006). Thus culture-independent molecular characterization techniques, such as DNA sequencing, and especially those designed to deal with the heterogeneity of environmental samples, has the potential to greatly improve our understanding of soil microbial communities, processes, and biodiversity (Fulthorpe *et al.* 2008; Lemos *et al.* 2011). Molecular biology techniques have followed Moore's Law (technological output per unit cost increases exponentially over time; Carlson 2003), and the current cost of sequencing DNA from environmental samples has fallen far below 1¢/Kb sequenced; this has had a tremendous effect on the study of soil microbial ecology, since thousands of samples can now be sequenced together for relatively little cost. Chapter 2 of the current work employs single-colony direct Sanger-sequencing (Sanger *et al.* 1977), since I was able to utilize AMF colonization partitioning *in situ* in plant roots with *Glomeromycota*-specific primers. Chapters 3 & 4, on the other hand, use 454 bar-coded pyrosequencing (Hutchinson 2007; Hamady *et al.* 2008; Meyer *et al.* 2008), which allowed simultaneous sequencing of prokaryotic and eukaryotic sequences from nearly 100 aggregate-sized soil samples. The difference between the two methods are vast in terms of effort, time consumption, and the quantity of data generated; the similarity is that both give a relatively clear, culture-independent identification of microbial taxa present in a sample. Thus all microbial community data in this dissertation are derived from DNA sequencing.

High-throughput DNA sequencing (including massively-parallel 454 pyrosequencing, mentioned above) has allowed the large-scale study of complex soil microbial communities not previously feasible, and this has expanded our understanding of the processes and factors that control microbial community assembly in a variety of soil environments (Fierer *et al.* 2007a; Roesch *et al.* 2007; Ramette *et al.* 2009; Lemos *et al.* 2011; Fierer *et al.* 2011); the application of ecological theory in microbial research has accelerated as a result (Martiny *et al.* 2006; Prosser *et al.* 2007; Fierer *et al.* 2007b, 2009a; Philippot *et al.* 2011). Using 16S bar-coded sequences from biomes across the North American continent, Lauber *et al.* (2009) showed pH to be a strong predictor of bacterial community composition, after accounting for biome covariates such as plants, plant functional types, elevation, and climate, though all of these factors are inherently interrelated; these findings have been reinforced by additional work (Rousk *et al.* 2010a, b). Though some ecological principles can be generalized to include plants, animals, and microbes (Ramirez *et al.* 2010; Nemergut *et al.* 2010), additional research has illustrated the fundamental ecological differences in biodiversity and biogeographical patterns between macro- and microorganisms (Fulthorpe *et al.* 2008; Chu *et al.* 2010; Fierer *et al.* 2011). In fact strong evidence points to the ecological coherence of higher taxonomic groups of bacteria (Philippot *et al.* 2011). Network analysis has been used to show that co-ocurrences among bacterial groups in soils are non-random (Berberán *et al.* 2011), and Bates *et al.* (2011b) used archaeal 16S sequences to relate community patterns to soil-chemical conditions in ecosystems

across the globe. Most pyro-tagged studies conducted in soils have utilized 16S ribosomal sequences; to my knowledge, no study has yet been able to capture the whole eukaryotic soil microbial community, and part of the reason is the limited ability for primers within the 18S region to capture all of Eukarya. Methods for detecting groups within Eukarya have been developed and used in other, primarily aquatic, environments (Amaral-Zettler *et al.* 2009; Medinger *et al.* 2010; Chariton *et al.* 2010; Behnke *et al.* 2011). Fungi have probably been investigated most often compared to other groups of soil eukaryotes, owing to their important role in soil processes. Fungi have also been targeted more often because identification of fungal primers that capture species-level distinctions within specific groups, while difficult for fungi as a whole, is more feasible than for all eukaryotes (e.g., Taylor *et al.* 2008; Öpik *et al.* 2009; Rousk *et al.* 2010a, b).

Along with the rapid creation and accumulation of unprecedented volumes of microbial community nucleotide data, methods for dealing with these data have also increased rapidly (Gilbert *et al.* 2011; Lemos *et al.* 2011). Several bioinformatics pipeline software packages are currently available for use with different types of metagenomic data (e.g., Caporaso *et al.* 2010b; Sun *et al.* 2010; Giardine *et al.* 2005), and these incorporate many choices for analyzing sequence data. The barcoded 454 data presented in Chapters 3 & 4 of this dissertation were processed using QIIME (Caporaso *et al.* 2010b). Phylogenetic measures of dissimilarity have become

a common way to summarize multivariate community data, and the UniFrac distance (Lozupone *et al.* 2006) is an index that allows for statistical comparisons of communities using clustering and ordination techniques, and provides insight into community differences that are not only based on community-organism content, but also the evolutionary interrelatedness of potentially interacting community members. With choices in every step of analysis, it has become increasingly clear that analytical methods can have a tremendous effect on results and inference (Schloss 2010; Liu *et al.* 2007). Along with the infusion of ecological community analyses into microbial ecology, concerns over experimental design, replication, and meaningful inference are being discussed in a way reminiscent of past discussions in macroecology (Hurlbert 1984; Prosser 2010; Lennon 2011; Lemos *et al.* 2011); these concepts were especially taken into consideration during the design of experiments presented in Chapters 3 & 4, and it is my hope that this work adds to the ongoing conversation.

Overview of Dissertation

Surprisingly little research has been conducted in geothermal soils, even though the study of more charismatic thermal features has yielded breakthroughs in the understanding of life on Earth. The formation of these soils is unique in soil science, since the primary factor determining the elemental content of particles, the chemical conditions, and thus biotic inhabitants, is geothermal-influenced water. Arbuscular mycorrhizal fungi and biological soil crust communities, the two major groups of soil organisms investigated in this dissertation, are both well-studied, though some basic

ecological questions stand out since *in situ* studies are difficult for both groups, and little research has gone into their habitation in geothermal soils.

Chapter 2 entails the study of arubuscular mycorrhizal communities inhabiting the roots of *Mimulus guttatus* in paired, adjacent contrasting plant community types – a situation that illustrates the utility of geothermal soils as model systems. Using 28S ribosomal sequences from fungi within plant roots, I assessed a potential link in symbiont community structures. This study represents a step in our understanding of community assembly of symbiotic communities, and especially for arbuscular mycorrhizal fungi.

Chapter 3 describes a previously unstudied geothermal diatomaceous biological soil crust, as well as its eukaryotic microbial component. This system is a product of a combination of geothermal conditions, and while the formation appears to have a small geographic extent and perhaps minimal influence on the surrounding biome, this is a very unique soil situation that challenges conventional definitions of parent material, soil particles, porosity, and soil development. I happened across this unique thermal environment while sampling plants for another project (Chapter 2). Every time I walked through the area, I was intrigued by the texture of the soil surface, the fibrous cohesion of the soil material, and the visibly active phototrophic community in the top few centimeters of the epipedon. After brief correspondence with leaders in the field of biological soil crusts and geothermal microbial ecology, I obtained funding to survey the microbial communities inhabiting the soils. In this study, I was able to

take advantage of differences in soil material and microbial communities across spatial scales in order to investigate questions regarding microbial community heterogeneity and metagenomic sampling design.

Chapter 4 expands on the findings presented in Chapter 3, and here I analyze the prokaryotic component of the diatomaceous biological soil crust community. Additional steps were taken in this study to incorporate phylogeny so that evolutionary history is added to community analysis. I also compare occurrences of ecologically-important bacterial phyla, cyanobacteria, verrucomicrobia, and planctomycetes along a photic gradient, and across soil types. Additional work concerning sampling effort assessment and rarefaction of pyrosequencing datasets is included as an appendix.

CHAPTER 2.

LINKING SYMBIONT COMMUNITY STRUCTURES IN A MODEL ARBUSCULAR MYCORRHIZAL SYSTEM

Contribution of Author and Co-Author

Manuscript in Chapter 2

Author: James F. Meadow

Contributions: Conceived the study, obtained partial funding, collected and analyzed output data, and wrote the manuscript.

Co-author: Catherine A. Zabinski

Contributions: Conceived the study, obtained partial funding, assisted in collecting data, and assisted with study design and discussed the results and implications and edited the manuscript at all stages.

Manuscript Information Page

James F. Meadow and Catherine A. Zabinski
Journal Name: New Phytologist
Status of Manuscript:
___ Prepared for submission to a peer-reviewed journal
___ Officially submitted to a peer-reviewed journal
_✓Accepted by a peer-reviewed journal
___ Published in a peer-reviewed journal
Accepted January 31, 2012

Title: Linking symbiont community structures in a model arbuscular mycorrhizal system

Short Running Title: Linking symbiont community structures

Key words: species-area; symbiotic; community structure; arbuscular mycorrhizal fungi; Yellowstone National Park; habitat diversity; host diversity.

Authors: James F. Meadow* and Catherine A. Zabinski

Author Affiliations:

Department of Land Resources and Environmental Sciences

Montana State University

334 Leon Johnson Hall

Bozeman, MT 59717

phone: (406) 994-4227

fax: (406) 994-3933

Email: jfmeadow@gmail.com*; cathyz@montana.edu

Summary

- The influence of plant communities on symbiotic arbuscular mycorrhizal fungal (AMF) communities is difficult to study in situ since both symbionts are strongly influenced by some of the same soil and environmental conditions, and thus we have a poor understanding of the potential links in community composition and structure between host and fungal communities.
- AMF were characterized in colonized roots of thermal soil *Mimulus guttatus* in both isolated plants supporting AMF for only a few months of the growing season, and in plants growing in mixed plant communities composed of annual and perennial hosts. Cluster and discriminant analysis were used to compare competing models based on either communities or soil conditions.
- *M. guttatus* in adjacent contrasting plant community situations harbored distinct AMF communities with few fungal taxa occurring in both community types. Isolated plants harbored communities of fewer fungal taxa with lower diversity than plants in mixed communities. Host community type was more indicative of AMF community structure than pH.
- Our results support an inherent relationship between host plant and AMF community structures, though pH-based models were also statistically supported.

Key words: symbiotic; community structure; arbuscular mycorrhizal fungi; Yellowstone National Park; habitat diversity; host diversity.

Introduction

Understanding the drivers and controls that structure biotic communities is among the most fundamental goals in ecology (Begon *et al.* 2006), and a major focus of microbial ecology in recent years (Fierer & Jackson 2006; Ramette *et al.* 2009). Soil microbes, in particular, exist in a complex heterogeneous environment and their presence is often a result, to varying degrees, of dispersal, environmental tolerance, and ecological interactions (Martiny *et al.* 2006; Miransari 2011; Unterseher *et al.* 2011). The relative importance of these controls is poorly understood for the majority of soil organisms, especially microorganisms (Fierer *et al.* 2009b).

Arbuscular mycorrhizal fungi (AMF) form a root symbiosis with a majority of vascular plants in all terrestrial soil systems, and influence processes from the scale of individual microbial interactions to ecosystems (Smith & Read 2008). The relationship between the two symbionts is generally regarded as mutualistic and primarily entails the exchange of plant photosynthates for immobile soil nutrients, especially phosphorus. Plants might receive additional benefits in improved water relations, pathogen resistance, and improved survival when establishing in existing mycorrhizal networks (Ruiz-Luzano 2003; Newsham *et al.* 1995; van der Heijden 2004), and plant growth and community composition are affected by mycorrhiza and AMF community structure (van der Heijden *et al.* 1998, 2003). A small number of AMF taxa (<300 described) appear to engage in mycorrhiza rather non-selectively with the majority of plant taxa; this taxonomically lopsided relationship might be one reason why AMF communities have previously been assumed to be somewhat less influenced by plant

community composition. Two-way controls on communities have been shown using feedback experiments (e.g., Bever 2002), and host plant identity, as well as biome, have been cited as important in affecting AMF community composition and diversity (Mummey *et al.* 2005; Hausmann & Hawkes 2010; van de Voorde *et al.* 2010; Davidson *et al.* 2011). Recent advances in the molecular technology required to study AMF communities *in situ* have shown that AMF diversity and community structure can change with differences in soil conditions such as pH, texture, and nutrient levels, and that fungal communities are sometimes controlled more by environmental characteristics than by dispersal limitations (Redecker 2006; Öpik *et al.* 2006; Lekberg *et al.* 2011). Meta-analysis of molecular AMF studies has revealed potential generalists and specialists within the Glomeromycota (Öpik *et al.* 2010), and Kivlin *et al.* (2011) reported that plant community type, soil temperature and moisture were all associated with changes in AMF community composition. The relative importance of host versus soil environment, however, is difficult to study in natural systems since plant and fungal communities both respond to variation in soil conditions; the result is that we have an incomplete understanding of the relative influences of soil conditions and host communities as as drivers of AMF community structure and composition.

Geothermal soils in Yellowstone National Park (YNP) are a unique example in soil formation in that these soils are largely a product of geothermal influences, including steam and elevated temperatures, periodic inundation by chemically diverse geothermal waters, and rapid taphonomy of plant and microbial biomass by mineral-saturated waters; mineral precipitates are the primary parent materials for many of

these soils (Channing & Edwards 2004; Rodman *et al.* 1996). These complex soil formation factors result in a mosaic of highly-variable soil conditions across a relatively small spatial scale. The heterogenous and depauperate nature of biotic communities in these thermal systems presents an elegant study system for addressing ecological questions without some of the confounding influences of distance, atmospheric differences and dispersal barriers. Plants living in these soils often exist in conditions that are far beyond the tolerance limits of most vascular plants (Bunn & Zabinski 2003), resulting in a subset assemblage of thermal-tolerant plants within the larger context of the Greater Yellowstone Ecosystem. One of these thermal-tolerant plants, a facultative-thermal, annual forb, is *Mimulus guttatus* DC., and when *M. guttatus* grows in geothermal soils it is often heavily colonized by AMF (Bunn & Zabinski 2003). *Mimulus guttatus* appears to have a wide tolerance for geothermal soil conditions and shows up in two very different plant community situations in geothermal soils: as a patch within an existing plant community consisting of grasses and other forbs, including annual and perennial plants (subsequently referred to as 'communal' sites; Fig. 1); and as individual, isolated plants emerging in essentially bare soil with few or no other plants (subsequently referred to as 'isolated' sites; Fig. 2.1). Additionally, these disparate plant communities often occur within 10m of one another, offering a unique opportunity to analyze the influence of plant community differences and edaphic conditions on the soil-microbial community. Vectors for AMF dispersal have previously been identified in YNP thermal soils by Lekberg *et al.* (2011), indicating that dispersal limitations between paired sites are not likely to play a major role

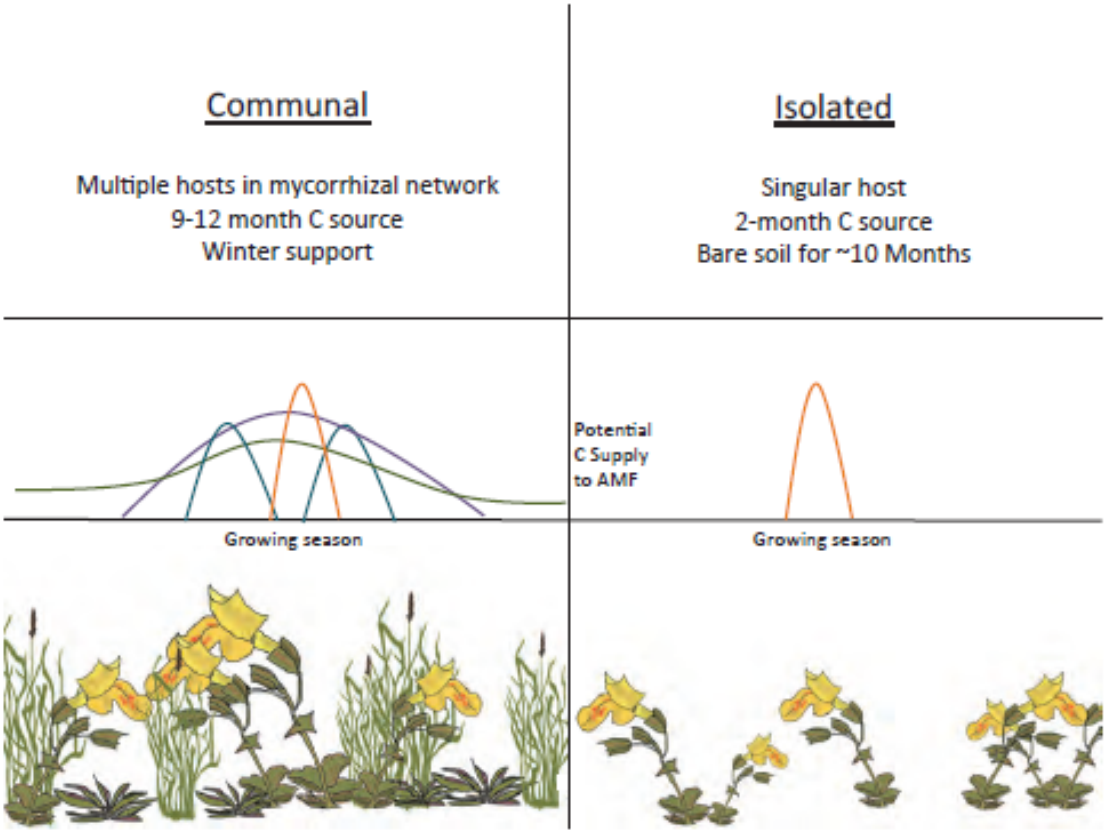

Figure 2.1. Paired site framework. All paired sites consisted of a communal (left) and an isolated (right) component. Hypothesized AMF community differences are based on the discrepancy in carbon supply (central panels) between these two host community situations, where communal sites might foster fungal networks throughout the growing season with some winter support, while isolated plant mycorrhizal support consists of C from a single plant during a short growing season window.

in structuring these communities. The reason for the difference in plant community types is not known.

From the perspective of the obligate mycorrhizal fungal community interacting with *M. guttatus*, these plant community types present two very different symbiont life history situations (Fig. 2.1). In the case of AMF colonizing hosts in mixed communities, host availability comes during the majority of the year from different host-plant

species, and with some element of winter root support in the case of sites that remain unfrozen. For those in isolated sites, all host support comes in a relatively short several month window, the duration of the *M. guttatus* life cycle, and symbiosis then shuts down for the remaining majority of the year. Given interspecific differences in sporulation and spore tolerance to harsh environmental conditions (Klironomos *et al.* 2001), the AMF communities in these two situations might differ either due to taxon-level variation in dispersal limitations (e.g., insufficient sporulation rates for some taxa to adequately disperse into isolated soils before *M. guttatus* germination) or in spore tolerance to storage during the months between host life cycles. AMF taxa in both host community types experience some degree of harsh, geothermal soil conditions, though mixed plant host communities are present throughout most of the year and for the entire regular growing season, and this host community structure allows for year-round mycorrhizal networks rather than obligatory spore production and storage. The severe filter exerted by isolated plants existing for only two months in otherwise bare soil would, on the other hand, likely favor those AMF taxa that are either prodigious spore producers that are dispersing into isolated sites at some point before *M. guttatus* germinates, or are especially tolerant of long-term storage as spores in geothermal soil conditions. Either way, the communities associated with these isolated sites are potentially a subset of AMF taxa occurring in communal plants, or, alternatively, a unique set of AMF that are selected for by conditions imposed by an isolated host, or a combination of the two. A substantive difference in AMF communities between host community types, if sufficiently independent of soil characteristics, would indicate an inherent link in symbiont community structures

associated with taxon richness and some predilection among AMF taxa in terms of association with ephemeral or perennial plant community types.

In the present study, we assess the relative roles of pH and host community structure as controls on AMF community composition in this model system in geothermal soils in Yellowstone National Park by contrasting AMF communities in the roots of *M. guttatus* living in either communal or isolated plant community types. We hypothesize that: a) isolated *M. guttatus* will harbor depauperate AMF communities in comparison with communal *M. guttatus*, and that b) fungal communities associated with isolated plants will be distinct from those associated with communal plants.

Materials and Methods

Plant and Soil Collection

Five paired sites were included in this study from Imperial Meadow (M1, M2, and M3; centered around 44°33′04″ N, 110°51′05″ W) and the Rabbit Creek (R1 and R2; 44°30′55″ N, 110°49′16″ W) drainage in Lower and Midway Geyser Basins, respectively, in Yellowstone National Park (WY, USA). Sites were selected based on the presence of appropriate adjacent paired patches of *M. guttatus*. Isolated sites consisted of either sparse or clumped *M. gutattus* plants with no other plant taxa growing within 1m. All of these plants were small relative to optimal growth conditions ($\leq$ 10cm tall; Dorn 2001), and though mycorrhizal networks could conceivably extend beyond 1m, the distance to surrounding plants helps to reduce this possibility. Between 6 and 20 plants were sampled individually from each patch (5 pairs

of patches), and sample number was determined by the patch size to avoid sampling more than 20% of the individual plants from each patch. Plants sampled from a given patch were all with 10m of one another; though *M. gutattus* density was comparable between isolated and communal sites, total plant density was inevitably higher in communal sites. Soil temperature was measured in the rooting zone of each plant, and whole plants were extracted with 5–10g of rhizosphere soil, transported to the laboratory within 8 hours, and frozen (-80°C) until being processed.

For processing, individual plants were allowed to thaw at room temperature (≈ 23°C) for 30 minutes before roots were carefully separated from soil. Soil was oven dried overnight (60°C) and analyzed for pH (Hendershot *et al.* 2008, 2:1 water extract), total C and N (LEKO combustion; Yeomans & Bremner 1991), and texture (micro-pipette extraction; Miller & Miller 1987). Due to the heavy influence from adjacent geothermal spring features, pH is considered to be a driving variable in these soils (Rodman *et al.* 1996), and is preferentially used during analysis in this study. The use of pH as a primary variable in ecological analyses of soil microbial communities is also well established in non-geothermal systems (e.g., Rousk *et al.* 2010a).

DNA Extraction, PCR

AM fungal colonization rates were estimated for plants from each site (Koske & Gemma 1989), and washed roots were cut into small segments, with size (≈ 0.5 cm) based on colonization rates in that more than half of all segments at a given length

were colonized at least once. Eight root segments were randomly picked per plant for molecular analysis.

We extracted DNA from root pieces by first denaturing plant and fungal tissues in 80 μl TBE buffer (95°C for 2 minutes) after which root pieces were manually crushed using sterile micro-pestles. A Chelex 100 suspension (20 μl) was added and samples were vortexed and placed on ice for 2 minutes before a second denaturation, a final vortexing, and a final icing. Samples were then centrifuged (8000 rpm for 5 min) to pellet cell contents, and supernatant was drawn and diluted 50x for use as template in nested PCR.

The first PCR was conducted with eukaryote specific primers (NDL22-0061) to target the variable D2 region of the large ribosomal subunit, and the second PCR utilized the Glomeromycota-specific FLR3–FLR4 primer pair (van Tuinen *et al.* 1998; Gollotte *et al.* 2004). Even though this primer combination can amplify non-Glomeromycotan sequences, and some AMF groups might be missed (Mummey & Rillig 2007), we detected taxa from most major AMF groups and no non-Glomeromycetes. All PCR was performed using GoTaq Green Master Mix (Promega, Madison, WI, USA), and with the following thermocycling conditions for PCR-1: 1 minute at 95°C; 30 cycles of 1 minute at 95°C, 1 minute at 53°C, and 1 minute at 72°C adding 4 seconds to the elongation step for each cycle; followed by a final elongation of 5 minutes at 72°C. The PCR-2 program was the same but with an annealing temperature of 56°C and only 25 cycles without the stepped elongation times. All PCR was performed with 50μl volume, including 2μl 50x diluted DNA extract or 50x diluted

PCR-1 product, for PCR-1 and PCR-2, respectively. This sequencing approach, including small-root-segment PCR coupled with direct sequencing, was used to avoid cloning and to attempt to capture single colonizations. The trade-off associated with direct sequencing is that more root segments can be analyzed with less effort and cost compared to cloning and sequencing, though mixed sequences are lost. Using this approach, approximately half of all root segments are expected to result in positive PCR products. Additionally, the majority of these positive products are expected to return singular AMF sequences. The first indication of multiple AMF sequences in the same PCR product is the appearance of multiple bands on agarose gel, though some combinations of AMF taxa produce segments that are nearly equal length. Thus PCR products were visualized using electrophoresis on agarose gel, and products showing single bands were cleaned using the QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA, USA) and submitted for direct sequencing at the Idaho State University Molecular Research Core Facility (www.isu.edu/bios/MRCF). Some PCR products, even those with apparently singular bands on agarose gel, were composed of multiple sequences, and these were filtered out upon sequence trace visualization.

Phylogeny

All sequence traces were manually screened using FinchTV version 1.4.0 for Mac (www.geospiza.com), and sequences showing signs of contamination, indicating mixed colonizations, were discarded. All screened sequences were queried using BLAST (Altschul *et al.* 1990). Operational taxonomic units (OTUs) were designated by clustering at 97.5% sequence similarity based on a neighbor joining tree; representative

sequences from each taxon cluster were used in final alignment and phylogenetic reconstruction. Though 97% similarity effectively separated most OTUs, a bimodal clade within *Glomus A* was resolved at 97.5%, and this had identical results for all other OTUs. Since our rDNA dataset was relatively small, and since known difficulties in assigning AMF OTUs based on sequence similarities occur with all candidate primer sets, we felt that manual clustering and assessment of an effective similarity cutoff was a more conservative approach to OTU selection than the use of automated clustering algorithms and a hard similarity cutoff. This also preserves potentially novel OTUs, since AMF in YNP thermal have not been extensively studied. Close GenBank sequences, as well as INVAM and BEG isolates, were included with each OTU representative to better distinguish clades. Sequences were aligned with ClustalW in Mega 5.0 for Mac (www.megasoftware.net), and maximum likelihood model tests were performed using PhyML (PhyML 3.2 for Linux; Guidon *et al.* 2010; Paradis *et al.* 2004) in R (version 2.11.1 for Linux; R Development Core Team 2010) with the `ape` package (Paradis *et al.* 2004). Bayesian inference phylogenetic trees were produced using BEAST (http://beast.bio.ed.ac.uk) for Linux and phylogenetic tree images were produced in R using `ape`. Trees were run under a general time-relaxed model with a 5-parameter Γ site distribution and with some sites assumed as invariant (G+Γ+I) based on PhyML results (AIC = 15928). Trees were run for 10^7 iterations, saving every thousandth result, and compiling after a 20% 'burn in.' Tentative OTU names were assigned based on newly proposed AMF phylogeny (Schüßler & Walker 2010; Krüger *et al.* 2011). Partial rRNA sequences from these samples were deposited in

GenBank under accession numbers JN836499, JN836501–9, and JN836511–28 (Table 3; Supplemental table in accepted manuscript).

Statistical Analysis

Relative abundance data were calculated for each fungal taxon in each site by combining presence in plants and applying a sample total standardization (Equations 1 & 2 in Appendix A); sample total standardized data were used for all analyses other than initial rarefaction. All statistical analyses were done in R. Network visualizations were performed using the `bipartite` package (Dormann *et al.* 2008), and individual sites were summarized with α-diversity metrics (richness, Shannon-Wiener diversity, and Shannon-Wiener evenness). Tests of these indices were performed as Kruskal-Wallis rank-sum tests. We tested for a bias in sampling effort by comparing the total number of sequences per site, the number of plants that returned clean AMF sequences per site, and the average numbers of sequences per plant for each site, all with two-sample t-tests. We also created species accumulation curves for each community type using the 'rarefaction' method implemented in the `vegan` package (Oksanen *et al.* 2011).

Raup-Crick probabilistic dissimilarity values were computed for multivariate analyses as instituted in the the `vegan` package (Raup & Crick 1979; Chase *et al.* 2011). Agglomerative hierarchical clustering was conducted using the flexible-β approach (β = -0.25) to cluster sites by community similarity. Silhouette and heat plots were created with the `optpart` package (Roberts 2010c) to assess goodness-of-clustering

for each solution. Analysis of similarities (ANOSIM), to assess significance of clustering solutions, was performed with the `vegan` package. Four clustering solutions were tested with agglomerative hierarchical clustering, and these correspond to Fig. 2.6: a) community type; b) bimodal pH, with M2c in the high-pH cluster; c) balanced ranked pH, exchanging M2c and M1i; and d) unbalanced ranked pH, including M3i in the low pH cluster. The latter three clustering solutions (b, c, & d) represent three different views of 2-cluster solutions based on pH rather than community type.

Results

Soil Analysis

All soil physical and chemical characteristics measured, with the exception of soil temperature, were predictably correlated with pH (Pearsons correlation coefficient r = 0.72, 0.64 & 0.57; for pH compared to total N, total C, and clay, respectively) and none were distributed as evenly as pH (Tables 2.1 & 2.2). While temperature has previously been explored as a primary driver of plant and fungal communities in YNP thermal soils and has been noted to fluctuate during growing seasons (Bunn & Zabinski 2003; Bunn *et al.* 2009), soils used in this study had a relatively narrow range of average temperature, from 24.2°C to 29.8°C. This narrow range, and the fact that we collected soil temperature data at only a single time point, did not allow the complete investigation of temperature as a primary gradient of interest in driving AMF community patterns in this study. There was also no difference in average soil temperature between the two community types (t = -0.62, P=value = 0.56; from a

two-sample t-test). Isolated plant communities were all found growing in high-pH soils (mean = 8.97) with low percentages of total C and N (mean = 1.17% & 0.16%, respectively), while plants growing in communal sites were generally in more neutral to slightly acidic soils with a single notable exception. Though pH is quite different between plant community types (P = 0.015; Table 2.1), one communal site, M2c, has a high mean pH (pH = 9.0) that is more consistent with isolated sites in this study, and its other measured soil characteristics follow suit.

Molecular Identification and Phylogeny

DNA was extracted from 1,128 root segments for PCR, and, by design, approximately half of all root segments showed positive PCR products on agarose gel, of which 478 showed singular bands and were submitted for sequencing. After manually quality-filtering for mixed sequences and putative chimeras, 228 sequences were used for final analysis. These were individually searched using BLAST and assigned tentative taxonomic identifiers based on nearest BLAST result or known INVAM or BEG isolate. Pairwise alignments and phylogenetic construction with sequences from known isolates resulted in twenty-eight OTUs in eight genera. Posterior support for every genera-level clade is 1.0, while similarly high support is consistent throughout most major phylogenetic groupings (Fig. 2.3). We also found no difference between isolated and communal sites in the number of plants returning AMF sequences (t = -1.23, P-value = 0.26; from a two-sample t-test), but communal plants did harbor more OTUs per plant (1.69 & 1.27 OTUs per communal and isolated plant, respectively; t = -2.38, P = 0.056; from a two-sample t-test of average OTUs per plant

in each site). Isolated plants, returned an average of 0.52 more clean sequences per plant than communal plants (mean sequences per plant = 2.63 & 2.11, for isolated and communal plants, respectively; t = 1.96, P-value = 0.09; from a two-sample t-test), and an average of 14 more clean sequences were found per site in isolated sites as a result (29.8 and 15.8 sequences per site for isolated and communal sites, respectively; t = 2.04, P = 0.09; from a two-sample t-test). These tests together indicate an approximately even sampling effort across community types in terms of the number of plants sampled, but the large difference in the number of sequences returned per site is perhaps an indication of the larger number of mixed colonizations in communal plants that were discarded (Fig. 2.2a & b). More sequences per plant and per site could potentially bias richness and diversity toward isolated plants, but this was certainly not the case in the present study, as communal plants were far more OTU rich and diverse. It is also likely, given the number of singletons detected in communal plants, that this sequencing approach resulted in an underestimate of OTU richness in communal plants, as indicated by the steep terminus of the communal species accumulation curve (Fig. 2.2c).

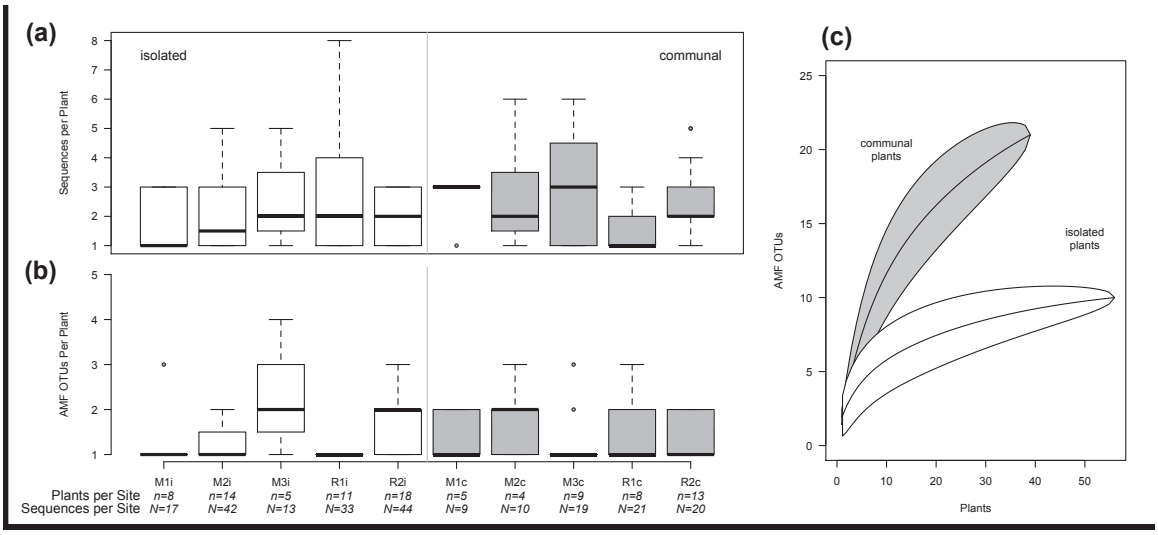

Figure 2.2. Sampling Effort Assessment. a) Boxes show the distribution of sequences returned by individual plants within each site; and b) the distribution of OTUs detected per plant (n = number of plants per site; N = number of clean sequences per site). We found no difference between isolated and communal sites in the number of plants returning AMF sequences (t = -1.23, P-value = 0.26; from a two-sample t-test of n), but communal plants did reveal more OTUs per plant (1.69 & 1.27 OTUs per communal and isolated plant, respectively; t = -2.38, P = 0.056; from a two-sample t-test of average OTUs per plant in each site). Isolated plants returned an average of 0.52 more clean sequences per plant than communal plants (mean sequences per plant = 2.63 & 2.11, for isolated and communal plants, respectively; t = 1.955, P-value = 0.088; from a two-sample t-test), and an average of 14 more clean sequences were found per site in isolated sites as a result (29.8 and 15.8 sequences per site for isolated and communal sites, respectively; t = 2.04, P = 0.09; from a two-sample t-test on N). c) Species accumulation curves for isolated (white confidence interval) and communal (gray confidence interval) plants show that the sampling effort likely captured only a subset of AMF OTUs from communal sites, but that most isolated AMF OTUs were detected.

Table 2.1. Mean Soil Parameter Values. Values represent averages of measurements (± SE) from rooting-zone soil under individual plants within communal or isolated sites. Footnotes reference method used.

Site	Community Type	*n* Plants	*N* Sequences	Soil Temp (°C)	pH*	% N**	% C**
M1	isolated	8	17	26.75 (±0.49)	8.12 (±0.11)	0.21 (±0.01)	1.09 (±0.05)
M2	isolated	14	42	29.79 (±1.13)	9.75 (±0.10)	0.09 (±0.00)	0.89 (±0.02)
M3	isolated	5	13	24.60 (±0.68)	8.67 (±0.24)	0.15 (±0.01)	1.04 (±0.04)
R1	isolated	11	33	25.91 (±0.51)	9.34 (±0.10)	0.18 (±0.01)	1.54 (±0.06)
R2	isolated	18	44	24.22 (±0.55)	8.99 (±0.07)	0.18 (±0.01)	1.22 (±0.04)
M1	communal	5	9	28.44 (±0.80)	6.18 (±0.22)	0.33 (±0.08)	3.41 (±1.26)
M2	communal	4	10	26.25 (±1.03)	9.00 (±0.12)	0.19 (±0.02)	1.64 (±0.19)
M3	communal	9	19	25.22 (±0.36)	6.36 (±0.15)	0.29 (±0.02)	2.63 (±0.18)
R1	communal	8	21	27.38 (±1.19)	6.97 (±0.27)	0.53 (±0.06)	7.01 (±0.93)
R2	communal	13	20	26.77 (±0.68)	6.72 (±0.22)	0.45 (±0.02)	4.95 (±0.26)

*Hendershot *et al.* (2008)

**Yeomans & Bremner (1991)

Table 2.2. Mean Soil Parameter Values (cont.) Values represent averages of measurements (± SE) from rooting-zone soil under individual plants within communal or isolated sites.

Site	Community Type	% Sand†	% Silt†	% Clay†
M1	isolated	84.38 (±0.84)	14.50 (±0.76)	1.12 (±0.23)
M2	isolated	89.50 (±0.54)	10.50 (±0.54)	0.07 (±0.07)
M3	isolated	74.60 (±1.78)	22.60 (±1.36)	2.60 (±0.68)
R1	isolated	76.45 (±1.40)	22.82 (±1.20)	0.73 (±0.27)
R2	isolated	78.72 (±0.63)	20.22 (±0.57)	1.22 (±0.10)
M1	communal	81.57 (±1.67)	16.86 (±1.32)	1.57 (±0.37)
M2	communal	84.00 (±2.35)	15.75 (±2.53)	0.50 (±0.29)
M3	communal	65.55 (±0.87)	31.45 (±1.05)	3.18 (±0.31)
R1	communal	58.00 (±2.04)	39.12 (±1.92)	2.88 (±0.52)
R2	communal	53.67 (±1.92)	36.20 (±1.32)	10.53 (±0.95)

†Miller & Miller (1987)

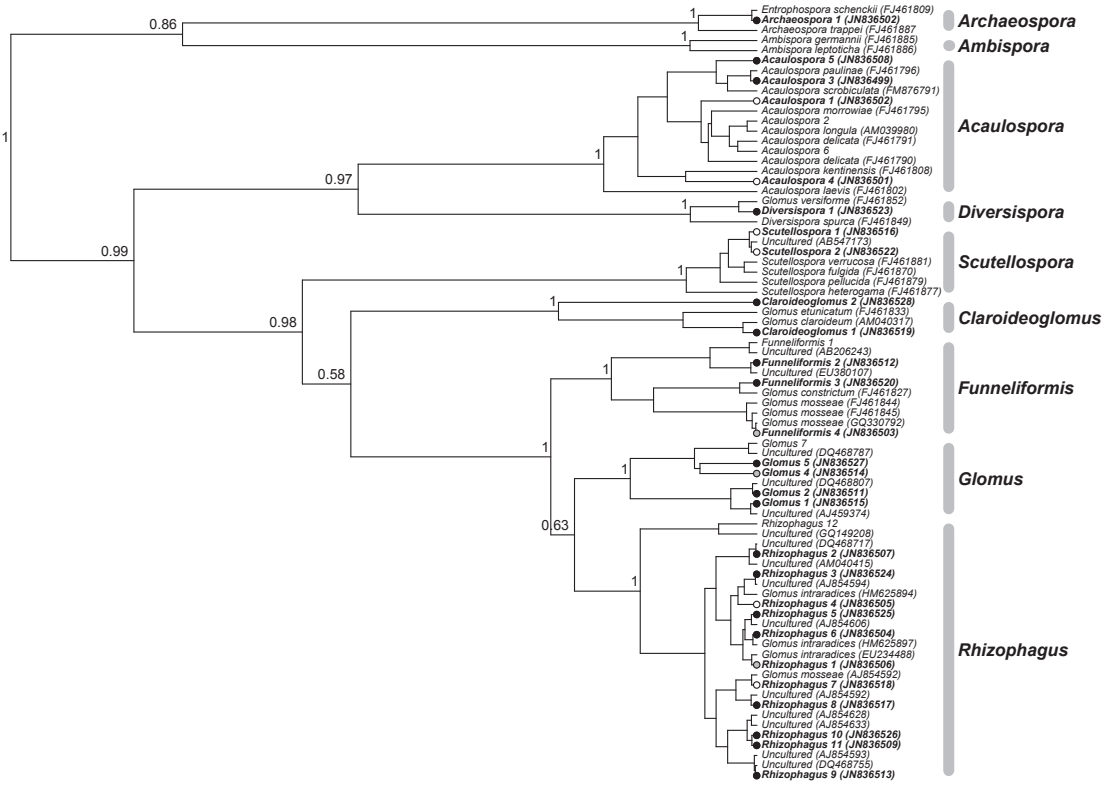

Figure 2.3. Bayesian Inference Phylogeny. AMF taxa used in these analyses are bold and with points at branch tips indicating plant community affiliation; black were found only in communal sites, white were found only in isolated sites, and gray were found in both community types. Clade bars indicate genera *sensu* Schüßler and Walker (2010) and Krüger *et al.* (2011). Names assigned during this study follow this convention, while isolates used for alignment retain names given in NCBI database. Node support values are posterior probabilities resulting from 10^7 iterations, saving every thousandth result, and compiling after a 20% 'burn in'. An ascomycete (*Mortierella polycephala*; NCBI accession number AF113464) was used as an out-group and is not shown.

Table 2.3. Operational Taxonomic Unit Metadata. Results from BLAST searches for each of the taxon clusters used as OTUs in this study. Shown are the most similar result from BLAST search with similarity, as well as most similar known isolate (isolated cultures with full epithets) and their similarity. Accession numbers assigned to sequences from this study are in OTU accession column (JN836499, JN836501–9, and JN836511–28).

OTU	Closest GenBank Match		Closest GenBank Known Isolate		OTU accession
	Accession	Similarity (%)	Taxon (Accession)	Similarity (%)	
Acauospora_3	JF717533	98	Acaulospora paulinae (FJ461796)	96	JN836499
Acaulospora_4	AJ459357*	86	Acaulospora delicata (FJ461790)*	87	JN836501
Archaeospora_1	FJ461809	98	Enterophospora schenckii (FJ461809)	98	JN836502
Funneliformis_4	GQ330792	99	Glomus mosseae (GQ330792)	99	JN836503
Rhizophagus_6	HM625897	99	Glomus intraradices (HM625897)	99	JN836504
Rhizophagus_4	HM625894	95	Glomus intraradices (HM625894)	95	JN836505
Rhizophagus_1	FM865595.2	100	Glomus intraradices (EU234492)	99	JN836506
Rhizophagus_2	JF717564	99	Glomus intraradices (JF439101)	97	JN836507
Acaulospora_5	AB610835	93	Acaulospora cavernata (FR692348)	93	JN836508
Rhizophagus_11	EF066679	99	Glomus intraradices (FJ235569)	98	JN836509
Glomus_2	DQ468807	97	Glomus intraradices (JF439212)*	82	JN836511
Funneliformis_2	EU380107	97	Glomus aggregatum (FJ461812)	93	JN836512
Rhizophagus_9	JF717562	100	Glomus intraradices (JF439169)	99	JN836513

*BLAST sequence query coverage <97%

Table 2.4. Operational Taxonomic Unit Metadata (cont.) Results from BLAST searches for each of the taxon clusters used as OTUs in this study. Shown are the most similar result from BLAST search with similarity, as well as most similar known isolate (isolated cultures with full epithets) and their similarity. Accession numbers assigned to sequences from this study are in OTU accession column (JN836499, JN836501–9, and JN836511–28).

OTU	Closest GenBank Match		Closest GenBank Known Isolate		OTU accession
	Accession	Similarity (%)	Taxon (Accession)	Similarity (%)	
Glomus_4	EU379995	98	Glomus microaggregatum (AF389021)	90	JN836514
Glomus_1	AB643635	98	Glomus microaggregatum (AF389021)*	85	JN836515
Scutellospora_1	AB547173	97	Scutellospora verrucosa (AY900509)	95	JN836516
Rhizophagus_8	JF717537	99	Glomus intraradices JF439196)	95	JN836517
Rhizophagus_7	FM992381	99	Glomus mosseae (AY769968)	99	JN836518
Claroideoglomus_1	AB548569	99	Glomus claroideum (AM040317)	98	JN836519
Funneliformis_3	EU380038	95	Glomus constrictum (JF439167)	95	JN836520
Acaulospora_1	AF389006	94	Acaulospora longula (AF389006)	94	JN836521
Scutellospora_2	AB547173	97	Scutellospora nodosa (FM876836)	93	JN836522
Diversiformis_1	FN547635	96	Glomus versiforme (FN547635)	96	JN836523
Rhizophagus_3	EF554562	99	Glomus intraradices (HM625896)	98	JN836524
Rhizophagus_5	EU380034	99	Glomus intraradices (HM625892)	97	JN836525
Rhizophagus_10	AJ854628	97	Glomus intraradices (FJ235569)	96	JN836526
Glomus_5	JF717467	96	Glomus constrictum (JF439180)	91	JN836527
Claroideoglomus_2	FN643146	98	Glomus claroideoglomus (AY541846)*	83	JN836528

*BLAST sequence query coverage <97%

Diversity and Community Composition

Of the AMF taxa found, eighteen taxa were found only in communal sites and seven only in isolated sites, while just three taxa were found in both site types, including *Rhizophagus intraradices* and *Funneliformis mosseae*, (synonymous with *Glomus intraradices* and *Glomus mosseae*, respectively; Schüßler & Walker 2010; Krüger *et al.* 2011). Bipartite network visualization of the dataset (Fig. 2.4) reveals a striking pattern of diversity differences between communal and isolated sites, regardless of their spatial proximity in the study area. Sites and fungal OTUs (boxes along the top and bottom, respectively, of Fig. 2.4) are ordered using a CCA-based χ^2 algorithm to reduce the number of interaction cross-overs, and is thus an indication of underlying relationships. Isolated sites appear to be dominated by only a single or two fungal OTUs (represented by the number and relative width of connections in Fig. 2.4), while communal sites show little sign of domination by any one OTU but rather a more even and diverse fungal community. Kruskal-Wallis rank-sum tests of taxon richness, Shannon-Wiener diversity, and Shannon-Wiener evenness all conclusively illustrate these differences (Fig. 2.5; Supplemental figure in accepted manuscript). Taxon richness in communal sites is twice as high as that in isolated sites, with 3.8 more fungal taxa in communal sites ($\chi^2 = 7.45$, $P = 0.006$), and Shannon-Wiener diversity and evenness are 1.9 and 1.2 times higher, respectively, in communal than in isolated sites ($\chi^2 = 6.82$, $P = 0.009$ for diversity & $\chi^2 = 5.77$, $P = 0.016$ for evenness). It is notable here that the single uncharacteristically high-pH communal site (M2c) has diversity indices that are all consistent with other communal sites rather than with comparably high-pH isolated sites (Fig. 2.4).

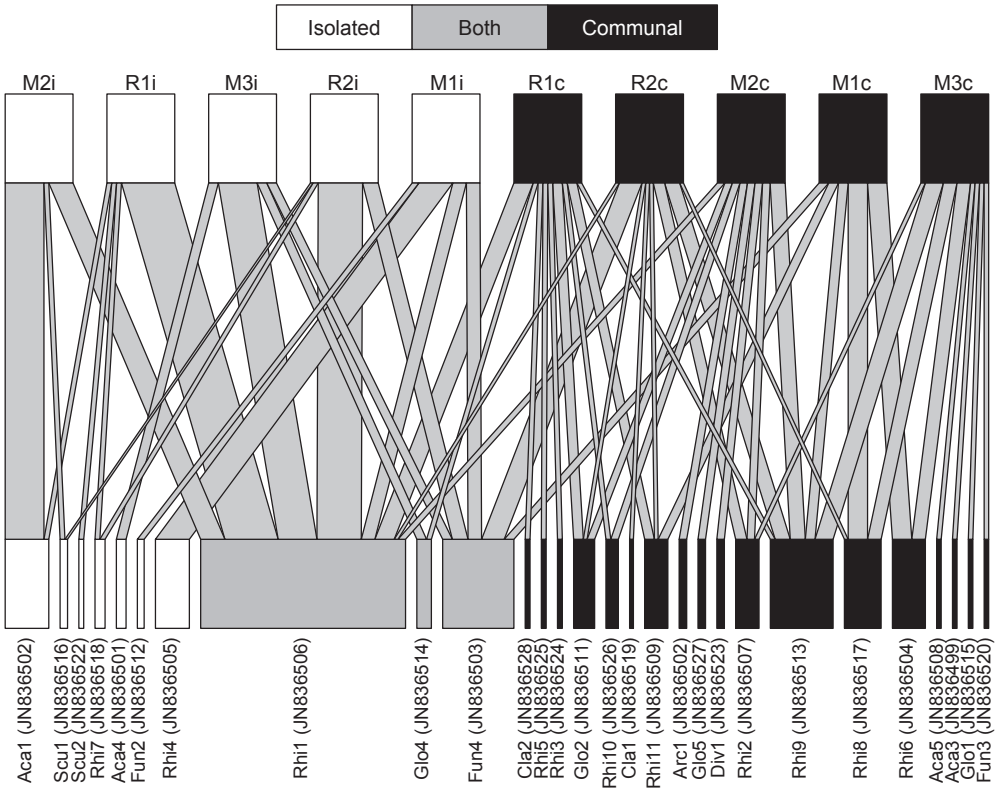

Figure 2.4. Full bipartite network visualization of all sites in the study. Boxes above are sites composed of all plants sampled from that site. Boxes below are individual AMF OTUs. Interaction width is proportional to relative abundance of fungal OTUs found at each site after sample total standardization (Supplemental equations 1 & 2). White sites were isolated, black sites were communal. White OTUs were found only in isolated sites, black were found only in communal sites, and gray were found in both site types. The order of sites and OTUs in the graphic is a result of a CCA-based χ^2 algorithm that minimizes the number of interaction cross-overs.

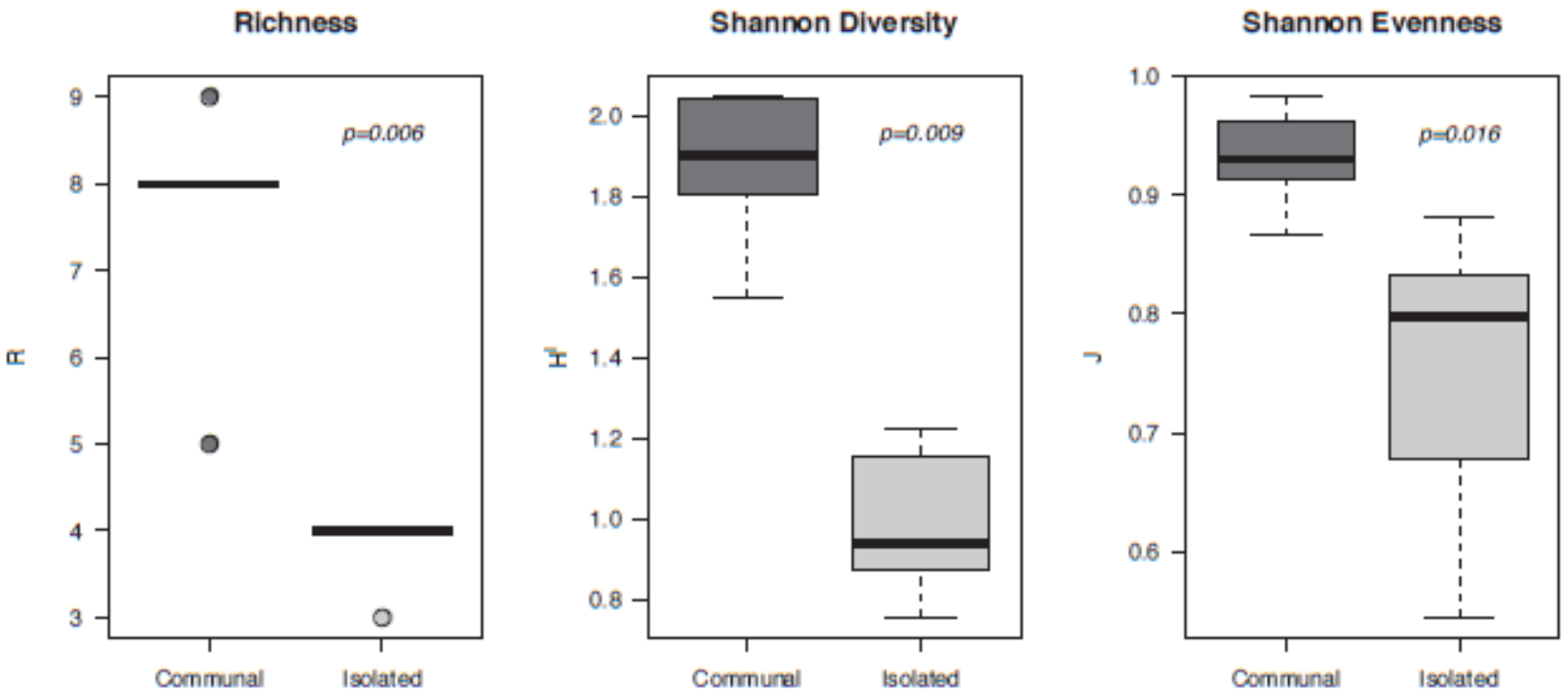

Figure 2.5. Diversity comparisons. Taxon richness, Shannon-Wiener diversity, and Shannon-Wiener evenness of sites separated by plant community type. P-values are from Kruskal-Wallis tests.

Cluster and Discriminant Analysis

Two different clustering approaches were employed for comparison: one relying on community dissimilarity data using the Raup-Crick probabilistic metric, and another on environmental data. The first strategy resulted in clustering consistent with plant community type (Fig. 2.6a), but when pH was used instead, clusters were created with plant community types mixed in two-cluster solutions. The community-type-based clustering solution was tested against three other possible combinations of pH-based two-cluster solutions with analysis of similarities (ANOSIM), and while all four models are supported to some extent, the community-type solution showed the lowest P-value ($P = 0.009$), though the R-statistic is perhaps more telling (Fig. 2.6). The R-statistic from ANOSIM is based on the difference of mean ranks between groups and within groups, and ranges from -1 to 1, with 0 indicating completely random grouping and positive values indicating systematic grouping (Legendre & Legendre 1998). The amount of variation explained by the clustering solutions is highest for the community-based solution, though all pH-based solutions return positive R-statistics.

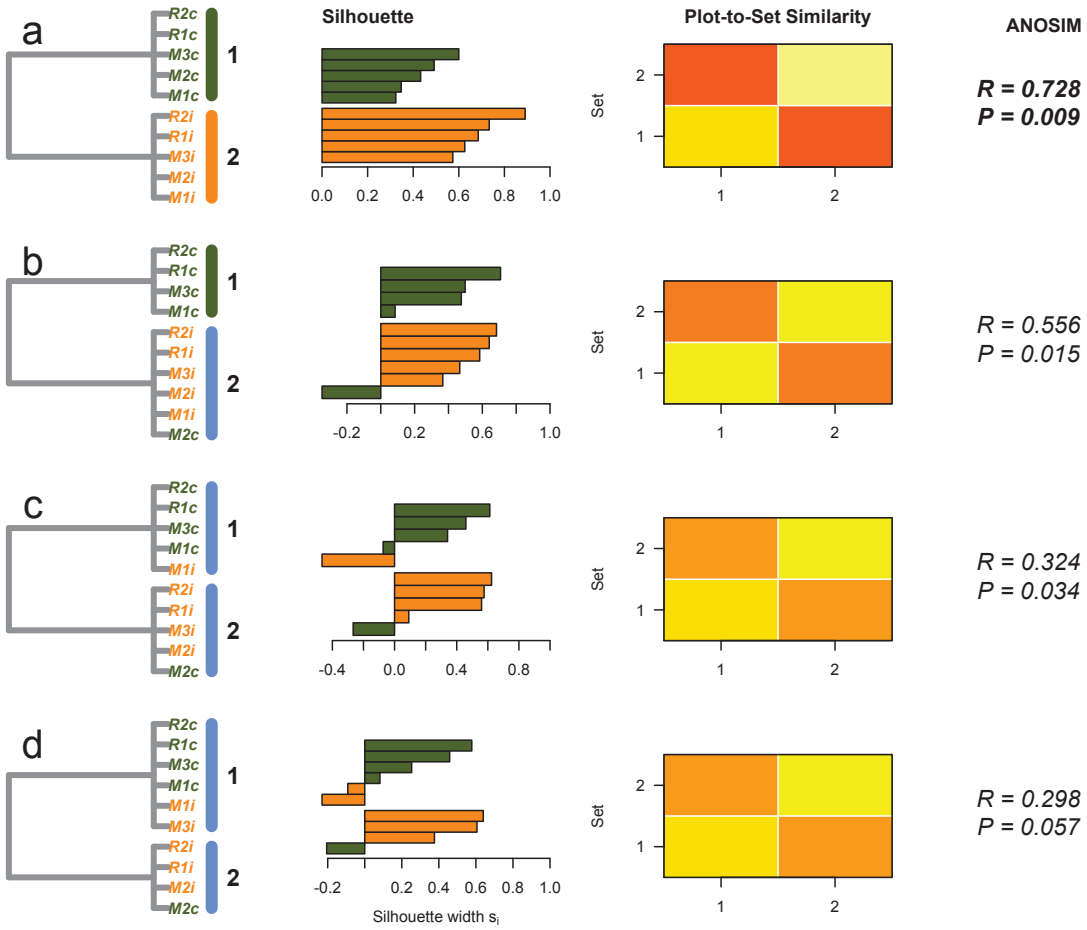

Figure 2.6. Comparison of Clustering Solutions. Trees represent clustering solutions based on: a) plant community type; b) bimodal pH; c) balanced ranked pH; and d) unbalanced ranked pH. Clade bars on trees represent community type homogeneity within each cluster; green are communal sites, orange are isolated sites, and blue include both community types. Silhouette plots (second column) represent within-group goodness-of-clustering from Raup-Crick dissimilarity. 'Set-to-Set' heat plots indicate goodness-of-clustering across clusters, with white representing the highest degree of similarity and red representing the highest degree of dissimilarity. P-values are from Analysis of Similarities (999 permutations); R-statistic represents variance of dissimilarity explained by the clustering solution, with possible values from -1 to 1, where 0 equals random assignment and 1 equals systematic grouping.

Discussion

Symbiotic organisms are limited not only by their own environmental tolerances, but also by the environmental tolerances of their potential symbionts, and this is especially true of plants and their obligately biotrophic AMF colonizers. Given the complexities of these relationships, the relative importance of either factor on the assembly of symbiotic communities is poorly understood. Our objective in this study was to elucidate drivers of AMF community composition and structure to assess a structural link between symbiont communities, and thus to assign relative influence to these drivers. Because the traits of plant and fungal communities tend to be driven by some of the same soil conditions, disentanglement of the relative influences from soil and ecological interactions becomes intractable. Thus studies attempting to elucidate drivers of AMF community composition often focus on either of the two drivers: edaphic factors irrespective of plant community compositional properties (Lekberg *et al.* 2011; Schechter & Bruns 2008), and across an edaphic gradient where plant communities are constant (Wu *et al.* 2007); or conversely as a factor of plant community compositional (van de Voorde *et al.* 2010) and structural differences (Börstler *et al.* 2006) across relatively constant edaphic conditions. Missing from this literature is an *in situ* comparison of the relative influences of soil and host community structure on AMF communities. The current study was an attempt to address this problem; the differences we observed in AMF communities were better predicted by host community type than by soil pH. Isolated plant communities used in this study

were all found in high-pH soils, while only a single communal site (M2c) fit this description, and all other communal sites were found in near-neutral pH; thus these two predictors are, unfortunately, nearly completely confounded. Although limited to a single site, and therefore insufficient for robust conclusions, the overlap between pH and plant community types in the high pH communal site provides an opportunity to compare the relative influences of soil and plant community structure on AMF community composition.

We hypothesized that AMF communities existing in the roots of communal *M. guttatus* would be distinct from isolated communities such that a more rich and diverse AMF community would be found compared to the relatively depauperate AMF community living in isolated plant roots. The most striking example of such a difference in AMF community composition and structure can be seen in Fig. 2.4. While taxon richness is clearly a factor in distinguishing community types, dominance by one or two fungal OTUs in isolated sites is perhaps more salient for identifying fungal community structural differences between isolated and communal sites. This characteristically even distribution was also observed in the lone communal site with more harsh soil characteristics (M2c), an indication that effects of host vegetation type on fungal communities are more important than the influence of pH. Though the overlap in pH between isolated and communal sites was limited to a single example, this site sheds light on a larger pattern of AMF community assembly; these findings are consistent with van de Voorde *et al.* (2010), who showed experimentally that plant community assembly history can have a major influence on symbiotic communities living in the roots of those plant communities.

Additionally, we found surprisingly little overlap between fungal communities in these contrasting host situations. In fact, only three of the twenty-eight fungal taxa were found in both plant community types, while seven taxa were found only in isolated sites and eighteen were found only in communal sites, even though paired sites are adjacent and assumed to have few barriers to AMF dispersal. Our findings can likely be attributed, at least in part, to island effects since AMF are not neutral in their dispersal abilities, though this sheds light on the autecological differences between AMF taxa, especially given the low degree of OTU overlap between community types (3 OTUs), relative to OTUs found multiple times in only one community type (10 OTUs). One of the three fungal taxa found in both community types, *Rhizophagus intraradices* (Rhi 1), was found primarily in isolated sites and made only rare appearances in communal sites, indicating that this taxon is either well suited to high-pH soil conditions or most successful in short-lived plant communities where dispersal and aggressive colonization are favored. Given that this taxon is found world-wide in highly variable soil conditions (Rosendahl *et al.* 2009; Öpik *et al.* 2010), the latter seems more apt, and the fact that *Rhizophagus intraradices* was differentially dominant in the two community types might indicate a more conserved 'niche' for this taxon than is often assumed. The other two of the three shared taxa show no particular affinity for either community type, and one of these, *Funneliformis mosseae* (Fun 4), was among the most commonly detected fungal taxa in this study. This seems to contradict the idea that isolated AMF communities would be a subset of the surrounding communal assemblages, but rather lends credence to the hypothesis that AMF taxa that show up in isolated plants are particularly suited to these conditions,

seemingly more so than to life in communal sites where competition among other colonizing AMF might play a bigger role, though some common taxa might be able to establish in either situation.

Since this study was designed around detecting differences in AMF community composition as a function of host community type, a major hurdle was the separation of soil characteristics, which likely play a major role in structuring these thermal plant communities, from the effects of the plant community structure itself. Host community type was only reasonably broken into two clusters, and thus we presented a comparison of the community-based clustering solution with three different options for clustering based on ranked pH. While all three pH-based clustering solutions resulted in positive R-values, indicating that communities were clustered by pH better than expected by random assignment, even the most significant pH-based clustering solutions (Fig. 2.6b; $R = 0.556$; $P = 0.015$) accounted for substantially less of the variation explained by the community-based solution ($R = 0.728$; $P = 0.009$), which was the most indicative of fungal community clustering. Clearly these results cannot be interpreted to infer that pH has no effect on AMF communities since all four models do show some degree of support. In fact, ANOSIM tests of all three pH-based clustering solutions resulted in P-values that were low; the R-statistic provides a better interpretation of the goodness-of-clustering, and both measures (R-statistic and the resulting P-value) indicate that host vegetation type plays a role in structuring AMF communities, and in this case, was more closely associated with variation in AMF communities than was pH.

In addition to the modest soil condition overlap, the scope of our study was also limited by our use of only a single plant species at a given point in time, which was inherent in our study design; our choice of *M. guttatus* was based on its ephemeral life history strategy and its appearance in completely isolated thermal soils near simultaneously emerging communal patches. Given that this study only explored the mycorrhizal community associated with a single plant species, more information will certainly be gained as more plant taxa are investigated for similar community structure links. Use of a single plant species helped to reduce the possibility that the differences in fungal community composition were simply due to potential *M. guttatus* specificity issues. Since geothermal soils can be subject to seasonal and diurnal variation in environmental conditions, it is important to note that our study represents only a snapshot of conditions encountered during the flowering period of *M. guttatus*. Differences in AMF community structure and composition are likely attributable to some combination of three factors discussed thus far: 1) long term spore-storage in harsh soil conditions limiting the establishment of some AMF taxa in isolated plant roots; 2) short isolated growing season limiting the accumulation of AMF taxa in isolated plant roots; or 3) host specificity issues associated with surrounding plant taxa in communal sites. This study was not designed to disentangle these three potential contributions to fungal community characteristics, but rather to detect a community-structure link between host and fungal communities.

By comparing the relative strengths of associations between AMF community and either soil-chemical characteristics or plant community traits, we found host community structure was more predictive of the observed differences in AMF community

structure than was pH, and that AMF communities in isolated plant patches were less species-rich, less even, and less diverse. While this is consistent with our first hypothesis, a competing pH-based model was also supported, and a more robust conclusion will rely on more overlap in soil conditions and more host species than just one.

This study adds to the growing body of evidence that plant communities exert some control over their associated AMF communities, and that AMF taxa are not neutral in their appearance in ecological communities. Our results also illustrate a potential link in symbiont community structure that might help soil ecologists understand AMF communities and their assembly patterns in both natural and managed systems.

Acknowledgments

We are grateful to the Yellowstone National Park permit personnel for help with research permits, to Ylva Lekberg for help with molecular methods, and to Rosie Wallander for help with soil analysis. We also thank three anonymous referees for their valuable comments. JM was supported with a GK-12 Graduate Fellowship from the National Science Foundation and the Big Sky Institute, and with additional support from USDA NRICG to CZ.

CHAPTER 3.

SPATIAL HETEROGENEITY OF EUKARYOTIC MICROBIAL COMMUNITIES IN AN UNSTUDIED GEOTHERMAL DIATOMACEOUS BIOLOGICAL SOIL CRUST: YELLOWSTONE NATIONAL PARK, WY, USA

Abstract

Knowledge of microbial communities and their inherent heterogeneity has dramatically increased with the widespread use of high-throughput sequencing technologies, and we are learning more about the ecological processes that structure microbial communities across a wide range of environments, as well as the relative scales of importance for describing bacterial communities in natural systems. Little work has been done to assess fine-scale eukaryotic microbial heterogeneity in soils. Here, we present findings from a bar-coded amplicon (18S rRNA) survey of the eukaryotic microbial communities in a previously unstudied geothermal diatomaceous biological soil crust in Yellowstone National Park, WY, USA, in which we explicitly compare microbial community heterogeneity at the aggregate-scale within soil cores. Multivariate analysis of community composition showed that while subsamples from within the same soil core clustered together, community differences between aggregates in the same core was unexpectedly high. This study describes an unstudied soil microbial environment and also adds to our growing understanding of microbial heterogeneity, and the scales relevant to the study of soil microbial communities.

Introduction

Soils hold an immense diversity of prokaryotic and eukaryotic microorganisms; these microbial communities affect processes from the molecular to the ecosystem scale (Sylvia *et al.* 2004). The substantial physiochemical heterogeneity of the soil habitat contributes to their biodiversity (Fierer & Lennon 2011), and soils are among the most taxon-rich of any microbial habitat (Madigan *et al.* 2008). Attempts at understanding the drivers of microbial community composition have met with mixed success, though one unifying pattern observed in microbial communities of soil and other complex environments is that these communities are structured along pH gradients (Lauber *et al.* 2009; Rousk *et al.* 2010b; Fierer *et al.* 2009b). The inherent complexity of these communities is one of the most difficult aspects of environmental microbial ecology, and this has become increasingly more analytically manageable as high-throughput DNA sequencing technologies have emerged (Lauber *et al.* 2009; Sogin *et al.* 2006; Roesch *et al.* 2007). These technologies have mitigated some of the problems associated with culture-dependent techniques, since environmental DNA can be extracted and sequenced from microorganisms regardless of their autecology. This has resulted in the rapid production of massive community datasets, allowing for complex ecological analysis of microbial data from hundreds, or thousands, of samples, which was not feasible prior to high-throughput sequencing. The study of soil microbial communities has been particularly helped in this regard, given the relatively small spatial scale containing a wide array of functions, and thus the opportunity to study complex interactions on a fine scale.

Ecological questions regarding soil microbial communities can be more easily addressed in extreme environments, where harsh environmental conditions impose strict habitat filters on resident communities (de Bruijn 2011). Arid soils, arctic ice, and geothermal features have been helpful in attempting to elucidate some of the drivers of microbial speciation, population dynamics, and community composition and structure (Hollister *et al.* 2010; Miller *et al.* 2009; Becraft *et al.* 2011; Yergeau *et al.* 2011).

Biological soil crusts (BSC) have been proposed as model systems for studying ecological principles in microbial communities since their biologically-driven functional attributes are well-studied compared to other soil systems, and since their communities are often functionally diverse in such a way as to be relatively independent of vegetation influences (Bowker *et al.* 2009). BSCs are well known for their pivotal role in arid soil ecology, and have been documented and studied in arid soils around the world. Major features of BSCs include anhydrobiotic microbial growth and exudative carbon sheath formation, N-fixing cyanobacteria, heterotrophic microorganisms, and lichenous growth in more well-developed BSCs. Although diatoms and other free-living eukaryotic algae do occur as BSC constituents in arid soils, both groups are subject to environmental restraints, especially desiccation, that limit their occurrence in BSC microbial communities (Belnap & Lange 2003).

In the present study, we analyzed community composition of eukaryotic microbes in a previously unreported and unstudied diatomaceous biological soil crust in geothermal-influenced soils in Yellowstone National Park, WY, USA. This unique BSC system occurs as the result of a combination of factors: 1) periodic inundation with alkaline-siliceous geothermal water and subsurface geothermal influence; 2)

harsh soil-chemical conditions that limit plant growth and coverage, and thus expose the soil surface to full solar radiation; and 3) rapid accumulation of diatomaceous residuum acting as soil parent material. The primary source of geothermal water in this site, and thus arguably the primary soil formation factor in shaping these geothermal soils, is a single alkaline siliceous spring (YNP Thermal Inventory ID LFMNN020; Fig. 3.1). From the source pool until water exits the system and flows into a larger stream, approximately 200m downstream from the source, water pH remains near 10, and Si concentration far exceeds concentration thresholds required to allow for diatom dominance in aquatic environments (Egge & Aksnes 1992). Though several other important elements are in relatively high concentrations (such as Na, B, As and Cl), N and P concentrations in water are very low throughout the site (Tables 3.1 & 3.2).

The study site can be logically broken into three distinct parts, each approximately 1ha in size: 1) Surrounding the pool is a coarse, unconsolidated soil subsequently referred to as "sinter" (Fig. 3.2a). This white, siliceous soil material has been described in similar geothermal systems, and is composed primarily of amorphous opaline silica and taphonized microbial biomass (Channing & Edwards 2004). Soil particles mostly range from silt sized to coarse fragments, averaging between 2-3mm diameter. This sinter area receives periodic inundation of geothermal waters, but the surface ≈1cm remains dry throughout the summer months. 2) A wet meadow collects water from the spring, after flowing through shallow cyanobacterial mat systems. The wet meadow is colonized exclusively by a single vascular plant species, *Triglochin maritimus*. This wet meadow area appears to be the site of most diatom

Figure 3.1. Source Geothermal Spring (YNP Thermal Inventory ID LFMNN020). Geothermal water flows from this source spring (55°C; <1 m^3 $second^{-1}$) into an open wet meadow, the apparent site of most diatomaceous growth and accumulation. Infrequent inundation events have also been recorded that flood the area in geothermal water.

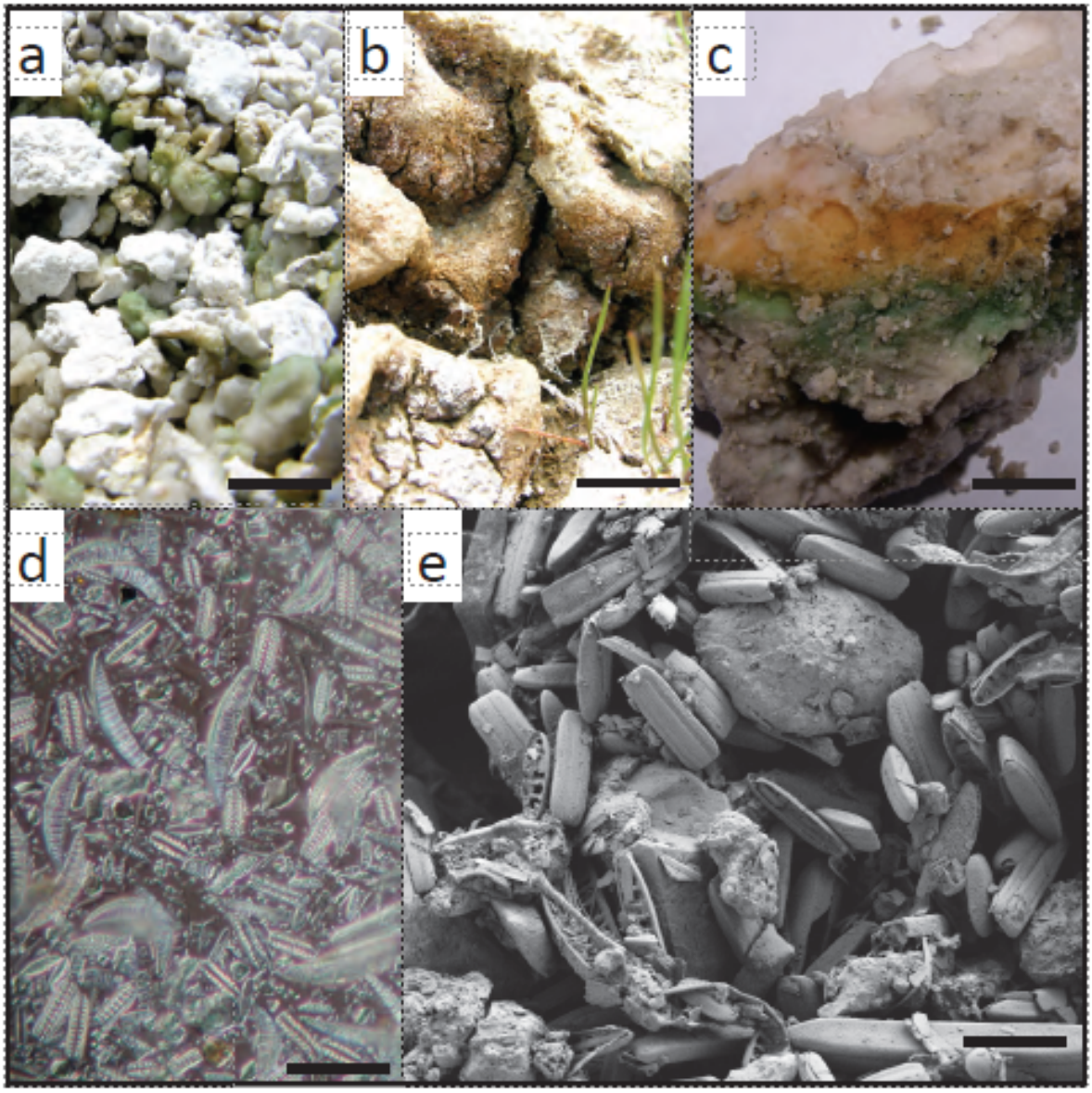

Figure 3.2. Diatomaceous geothermal soils: (a) Sinter particles (disturbed) with white on the surface above orange and green photic zones (scale bar = 1cm); (b) Fibrous, dense diatomaceous soil crust (scale bar = 3cm). This is the general appearance of crust samples; (c) Vertical photic layering (scale bar = 1cm). This striation was observed in similar appearance in both sinter and crust soils, across the study area; (d) Differential interference contrast (DIC) micrograph of crust soil material (20x; scale bar = 40μm). Throughout the study area, the majority of soil particles are comprised of diatom frustules and their fragments; (e) Field emission scanning electron micrograph (FE-SEM) of diatomaceous soil material from crust samples (scale bar = 15μm).

accumulation. Results from the wetland are not presented here. 3) The third part of this system is a unique and visually striking upland crusted soil system. This soil, subsequently referred to as "crust," is a remnant that was once under a wet meadow, but as these diatomaceous soils develop very rapidly, it is now above the direct influence of the flowing geothermal water. The pronounced surface texture in this crusted soil appears to be the product of a combination of seasonal swelling with microbial biomass and exudates, and heterogeneous erosion of the dense, fibrous remnant wetland diatomaceous accumulation (Fig. 3.2b). There is a conspicuous vertical coloration gradient of microbial communities in the top few centimeters of both soil types, with orange in the top active zone, green in the middle, and gray below the photic zone (Fig. 3.2c).

Microscopy of these soils indicates that the majority of silt- to fine-sand-sized particles are diatom frustules, either complete or fragmented; the remaining portion is amorphous opaline silica and taphonized biomass (Channing & Edwards 2004; Kyle & Schroeder 2007) (Figs. 3.2d & e). Live diatoms have been observed throughout the site, though the majority of diatomaceous residuum is composed of non-living frustules. The diatom community is morphologically dominated by *Denticula spp.* and *Rhopalodia spp.* (E. Thomas, personal communication).

This soil development sequence presents a unique opportunity to study the differentiation of eukaryotic microbial assemblages across multiple spatial scales and without dispersal limitations that might affect community structure, as previous work has shown bison to be dispersal vectors in these thermal soils (Lekberg *et al.* 2011). These two soils, the sinter and crust, are subsequently referred to as "soil types."

Vertical coloration is also exhibited in both soil types and is generalized as surface (top), photic subsurface (mid), and subphotic (deep) samples in each soil type.

The objective of this study was to compare microbial community compositions across several spatial scales relevant to soil microbial ecology, from the particle or aggregate level up to the soil type scale. Specifically we addressed three hypotheses: 1) eukaryotic microbial communities are more similar within individual samples on the single aggregate scale than they are across the site or along a photic depth gradient; 2) eukaryotic microbial communities differ significantly along a vertical photic gradient; and 3) eukaryotic microbial communities differ significantly between soil types (i.e., sinter vs. crust). The latter two hypotheses were also applied to the diatom communities.

Experimental Procedures

Sample Collection and DNA Extraction.

All soil samples were taken from Imperial Meadow (centered around 44°33′04″ N, 110°51′05″ W) in the Lower Geyser Basin, Yellowstone National Park, WY, USA. The spring at the center of this study system is classified as an alkaline siliceous geothermal pool (see Tables 3.1 & 3.2 for chemical description; Rowe *et al.* 1973). To characterize the chemical properties of thermal water from the spring, we measured water temperature and pH in the field with an Accumet AP63 portable water analyzer, and filter-sterilized water samples to 0.2μm, which were refrigerated until analyzed for chemistry. Total dissolved organic C (DOC), total N (TN) were measured using a

Shimadzu TOC-VCHS analyzer. Trace elements and dissolved anions were measured using inductively coupled plasma optical emission spectrometry (ICP-OES) on an Optima 5300 DV and ion chromatography (IC) on a Metrohm-Peak compact ion chromatograph, respectively.

All samples used in this study were taken in October, 2010. Three individual sinter particles ($\approx$ 3 mm diameter) were sampled from each of the representative photic depths using sterile forceps. All sinter particles were sampled from within 10m of one another, with one replicate set containing a particle from each photic depth taken from within a 1m area. Mid-depth sinter samples also consist of an extra set of mid-depth samples, since sinter particles were more ambiguous in orange versus green coloration (Fig. 3.2b). These were preliminarily tested for community differences, and were subsequently lumped into one mid-depth group for final analysis. Crust samples were taken from three different locations within the upland crusted soils, each within 10m of another and approximately 100m from sinter sampling locations. Crust soils were sampled with a sterilized soil core ($\approx$ 2 cm diameter), and were sampled to 5 cm depth. All samples were frozen on dry ice, on site, and archived at -80°C until DNA extraction.

Crust soil cores were cut along conspicuously colored photic boundaries with a sterile scalpel. Edges of the photic slices that touched the soil corer, as well as transition regions between conspicuous colors were removed so as to reduce cross-contamination from adjacent photic levels. These photic slices ($\approx$ 0.3 cm thick by 1cm wide) were cut into three equal pieces for subsampling. All subsampling was

performed in a sterile hood. Whole genomic DNA was extracted from each sinter particle and crust subsample using the FastDNA Spin Kit for Soil from MPBio (www.mpbio.com) following manufacturer's instructions, with an additional ethanol precipitation step.

Bar-coded Pyrosequencing.

A portion of the 18S small-subunit ribosomal gene was amplified using the F-907 primer (Troedsson *et al.* 2008) containing a 10-bp bar-code sequence with a Roche 454-A pyrosequencing adapter (Titanium Lib-L adapters), and the R-1428 primer with a Roche 454-B adapter. Unique bar-code sequences were used to distinguish between all samples and subsamples. PCRs were conducted with 10μM of each forward and reverse primer, 3μl template DNA, 22μl GoTaq Green Master Mix (Promega, Madison, WI, USA), and an additional 1μl BSA. PCR was conducted with the following thermocycler conditions: 5 minute denaturation at 95°C; 30 cycles of 30 seconds at 95°C, 30 seconds at 45°C, and 1 minute at 72°C; and an additional 5 minute annealing step at 72°C. Each sample and subsample was amplified in triplicate, pooled and cleaned using a QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA, USA). Equal concentrations of cleaned PCR product were combined and sent to the Center for Integrated BioSystems at Utah State University to be run on a Roche FLX Ti 454 pyrosequencing machine (www.biosystems.usu.edu). Raw 454 data files have been submitted to NCBI SRA (SRP010732.3) and bar-code mapping files are included in supplemental information as tab-delimited text files.

Sequence Processing.

All sequence processing was performed using QIIME (Caporaso *et al.* 2010b). Quality-filtered sequences were clustered at the 97% similarity level against a representative SILVA 18S database (Pruesse *et al.* 2007) designed specifically to be used with the QIIME pipeline. Taxonomy was also assigned to representative sequences from each cluster using the SILVA taxonomic classifications, and taxa that were represented by fewer than 10 high-quality sequences were removed from analysis. Sequences were then assigned to samples using their unique 10-bp bar-codes. Sequences were also culled from the phyla Chordata and Streptophyta, to reduce the number of non-microbial eukaryotes; the only Streptophytes detected in the study were from *T. maritimus*. Non-chordate metazoan sequences were retained. Twelve crust subsamples were removed from analysis due to low sequence numbers or dominance by sequences from *T. maritimus*. After trimming, the subsample dataset was rarefied to 533 sequences per subsample before proceeding with β-diversity analysis. The combined sample dataset was rarefied to 850 sequences per sample. An additional combined dataset included only diatoms and was constructed by selecting only OTUs which included the taxonomic distinction "Bacillariophyta." It is worth noting that since we used a SILVA database to cluster and assign taxonomy to OTUs, it is likely that diatom taxa were discarded if not represented in SILVA; this is potentially a major drawback in this method, but it is a conservative approach to assigning taxonomy. This dataset was rarefied to 330 sequences per combined sample.

Statistical Analysis.

All analysis and figure production following sequence processing was performed in R (R Development Core Team 2010). For ordination and β-diversity assessments, we transformed the abundance data using the Canberra quantitative metric as it is implemented in the `vegan` package (Oksanen *et al.* 2011); this version of the metric contains a non-zero correction as a way of de-emphasizing joint-absences between samples (Legendre & Legendre 1998). To test for influence from amplification, cell count and ribosomal copy-number biases, we compared both the dissimilarities and first axes of non-metric multidimensional scaling (NMDS) ordinations from Canberra with those of the Ochiai coefficient, a geometric qualitative dissimilarity metric (i.e., Canberra takes OTU abundance into account while Ochiai relies only on which OTUs were present or absent). Subsamples were clustered using a flexible-β agglomerative hierarchical clustering method (β = -0.25) as implemented in the `agnes()` procedure in `cluster` (Maechler *et al.* 2005). A parsimonious solution to cluster analysis was optimized with a combination of `partana()` and `silhouette()`, in the `optpart` and `cluster` packages, respectively (Roberts 2010c; Kaufman & Rousseeuw 1990). We used distance-based redundancy analysis (DB-RDA), as implemented by the `capscale()` procedure in `vegan` for assessing the relative strengths of each spatial gradient (sinter vs. crust; and photic gradient) in predicting community dissimilarities. Subsample- and gradient-based groups were tested with analysis of similarities (ANOSIM) in `vegan`. The relationship between spatial distance and community dissimilarity were tested using a quadratic model ($d_{Canberra} \sim log(Distance) + log(Distance)^2$) where *Distance*

is generalized spatial distance; spatial distances were approximated as follows: Sub-samples = 1cm; Samples = 1m (since these encompass sinter particles as well as crust soil cores); Within Soil Type = 10m; Across Soil Type = 100m. It is important to emphasize that these are not actual spatial distances, but rather distance rounded to the nearest logged-distance category; this allowed for the comparison of groups rather than assigning spatial distance to each pairwise calculation.

Results

Chemical properties of the spring (Tables 3.1 & 3.2) appeared to be relatively constant over the three sampling times (June, October and February). The spring periodically inundates the sinter area with a surge of geothermal water; we have measured two such events in 11 months of monitoring, but this inundation seems to be infrequent (Fig. 3.4). Concentrations of most elements and ions increased as water travels through the system, since steam is evaporating from the surface near the source. The notable exception to this is silicon, which decreased as it is precipitated and taken up. DOC and TN (Table 3.1) also show a considerable jump in the 200m sample, possibly indicating C and N being fixed in and near the stream throughout the microbially active crust study area, or perhaps a virus-shedding event, though these measurements are not replicated and are only intended for chemical characterization of the geothermal water moving through the system.

Table 3.1. Chemical characterization of geothermal water. Water temperature and pH were measured in the field with a handheld water analyzer. Dissolved organic C (DOC) and total N (TN) were measured by combustion catalytic odixation/NDIR. Anions were measured using ion chromatography. All concentration values are reported in mg L^{-1}. IDs indicate stream distance from source pool.

			C and N		IC Anions				
ID	Temp (°C)	pH	DOC	TN	F^-	Cl^-	NO_3^-	PO_4^{3-}	SO_4^{2-}
Source	54.7	9.7	1.328	0.167	30.46	320.90	ND	ND	21.75
12m	46.8	9.8	1.202	0.139	31.93	337.23	ND	ND	23.08
32m	16.2	10.2	1.096	0.120	36.98	375.64	ND	ND	30.82
72m	12.4	10.3	0.917	0.044	37.25	396.57	ND	ND	31.97
200m	12.1	10.1	5.428	0.400	42.67	448.61	ND	ND	44.48

ND = not detectable (<0.5 mg L^{-1})

Table 3.2. Chemical characterization of geothermal water (cont.). Trace elements were measured using inductively cooled plasma optical emission spectrometry (ICP-OES). All values are reported in mg L^{-1}. IDs indicate stream distance from source pool.

	ICP-OES										
ID	Al	As	B	Ca	Fe	K	Li	Na	P	S	Si
Source	0.442	2.94	5.17	0.236	0.009	27.13	11.44	467.4	0.066	14.89	187.9
12m	0.48	3.18	5.60	0.262	0.012	28.24	12.39	492.6	0.073	16.14	203.4
32m	0.373	3.64	6.39	0.242	0.006	32.94	14.1	557.3	0.087	18.61	194.9
72m	0.408	3.83	6.99	0.174	0.008	34.74	15.91	606.8	0.089	19.99	203.9
200m	0.129	4.45	7.47	0.599	0.026	36.34	16.81	717.6	0.113	28.84	110

Sequence Processing

After quality filtering, 173,381 18S sequences were retained (535-bp average length), from which 19,188 sequences were randomly selected by rarefaction of the subsample dataset (533 sequences per subsample). For the combined dataset, crust subsamples were analytically recombined while sinter samples remained singular; samples were rarefied to 850 sequences for each of 21 samples, totaling 17,850 sequences. The final diatom dataset was rarefied to 330 sequences for each of 21 samples, totaling 6,930 sequences.

A breakdown of class distribution in the rarefied subsample dataset (Fig. 3.3) shows that diatom OTUs (Bacillariophyceae) dominate most samples, though sinter particles show a relatively more even taxonomic distribution compared to crust subsamples. Crust surface subsamples were heavily dominated by diatom sequences, with decreasing diatom abundance with depth. Sinter particles also appeared to contain more rotifer sequences than did most crust subsamples. The target region of the small ribosomal subunit used in this study, unfortunately, gives poor resolution for diatoms, though we were able to assign tentative OTU species names based on BLAST results (Table 3.4).

Sample Dissimilarity

Given inherent abundance biases due to differential amplification, ribosomal copy number, and cell number, we compared the influence of abundance data using the Canberra and Ochiai metrics (quantitative and qualitative dissimilarity calculations, respectively) on the whole eukaryotic dataset. We found a high degree of agreement

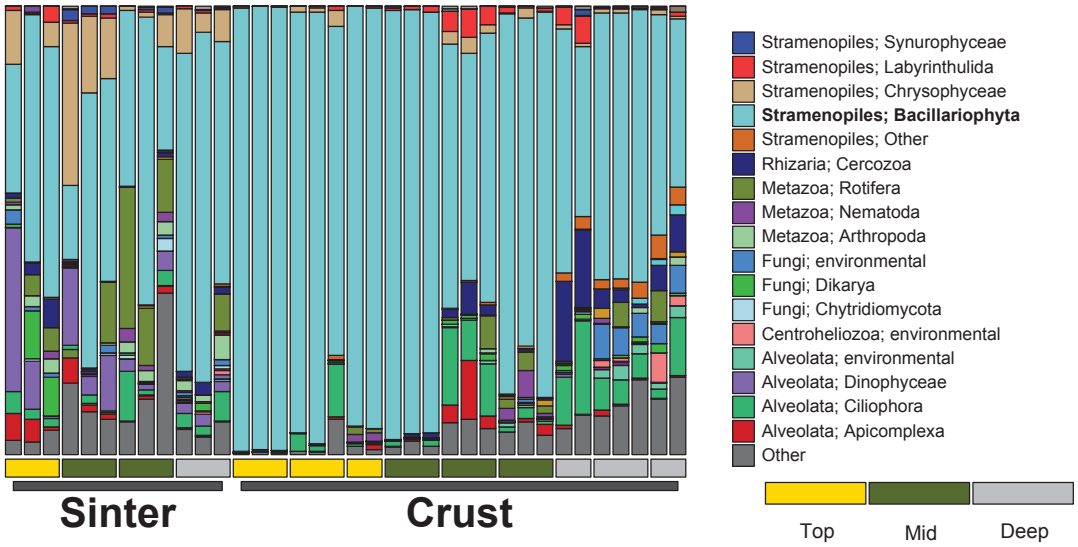

Figure 3.3. Class-level taxonomic distribution of OTUs after rarefaction. The most abundant class detected in nearly all samples was Bacillariophyta (diatoms), excepting some sinter particles. The legend includes only the 18 classes comprising more than 2.4% of any given sample for simplification.

between qualitative and quantitative indices when comparing dissimilarities (Pearson's correlation $r = 0.907$; Fig. 3.5a) and when comparing the subsequent first axes of NMDS ($r = 0.901$; Fig. 3.5b), indicating that some information was gained when considering OTU abundance, but this additional information does not skew dissimilarity based on abundance biases. This being the case, Canberra dissimilarity was used for all additional analyses.

Subsample Clustering

Agglomerative hierarchical clustering and ANOSIM were used to address our first hypothesis that eukaryotic communities are more similar within a sample than across

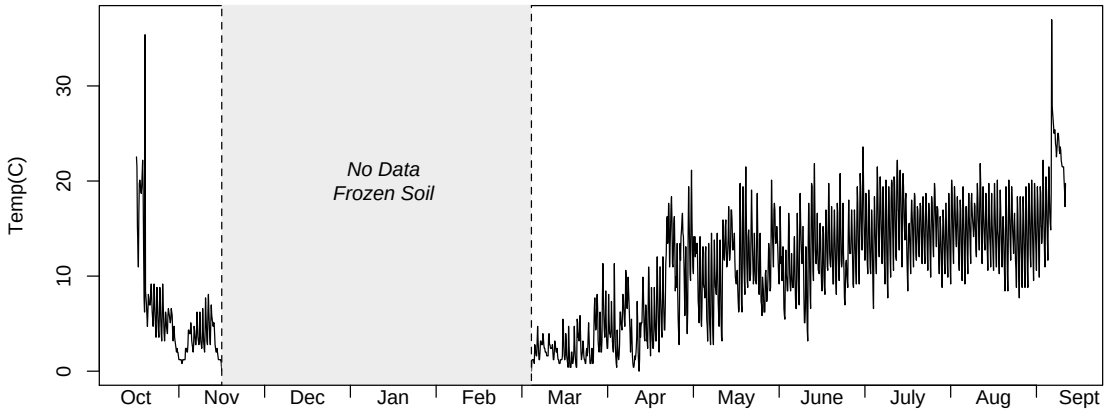

Figure 3.4. Soil temperature measured at 10cm depth in sinter ($\approx$ 10m from source pool). Data were recorded using an Onset StowAway XTI temperature logger. Dates when soil was frozen have been removed. The spikes in October 2010 and September 2011 indicate inundation with geothermal water. We were only able to capture two such events, though it is possible that inundation events during winter months were not captured.

environmental gradients. When subsampled data were assessed by clustering, subsamples are predictably regrouped within their respective samples (Fig. 3.6), and this clustering model explained a substantial portion of variability (ANOSIM R-statistic = 0.971, P-value < 0.001; Table 3.3). This 12-cluster solution consolidated all samples and gave the most parsimonious solution when considering partana (partana ratio = 2.57) and silhouette ($\bar{s}_i = 0.12$) results, with one cluster consolidating two replicate "deep" samples ($j = 12$ in Fig. 3.7, where j is cluster number). Although sinter and crust samples are not mixed in any of the 12 clusters, the highest bifurcation of the clustering tree puts deep and mid crust samples with all of the sinter samples, perhaps indicating greater degree of community partitioning between surface

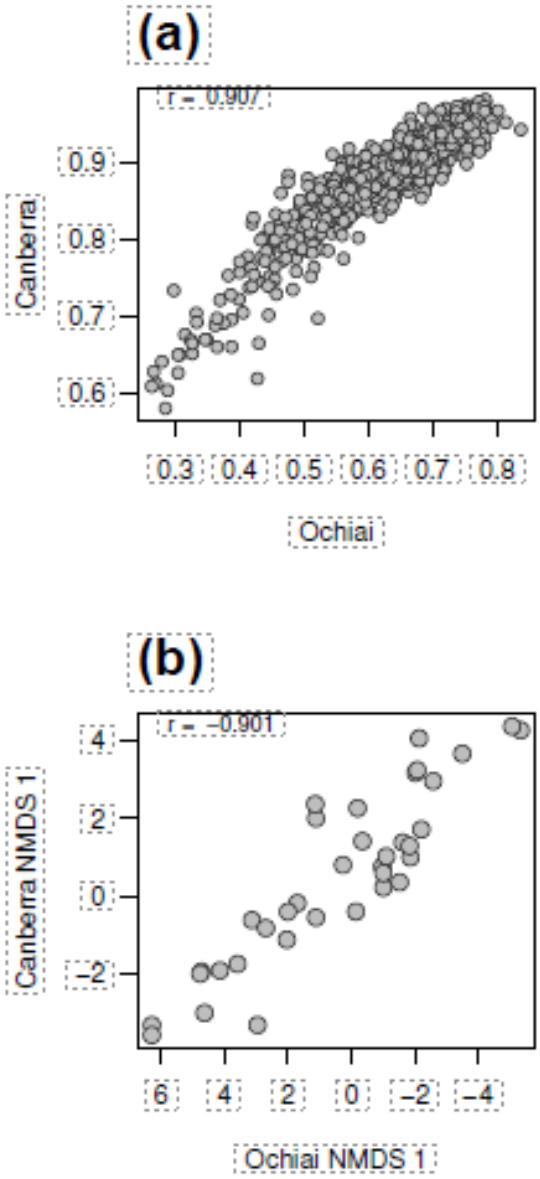

Figure 3.5. Comparisons between quantitative and qualitative dissimilarities: (a) and subsequent first axes of NMDS (b). Strong correlations in both cases indicates that abundance biases from amplification, cell numbers and ribosomal gene copy numbers do not substantially affect dissimilarity calculation or resulting ordinations, though additional information is clearly considered when using the quantitative Canberra metric. The values on the x-axis of the lower graph (b) have been reversed for consistency.

and phototrophically active communities than between those in subphotic samples in either soil. This might be understandable given moisture levels at depth in both soils and the influence of geothermal water on both subphotic communities, more so than in the surface samples.

Fig. 3.7a shows silhouette width representing goodness-of-clustering for each of the 12 clusters. The right panel of Fig. 3.7 (b & c) show results from partitioning analysis where grayscale value indicates average similarity of each cluster (top) and subsample (bottom) to other clusters, with white indicating greatest mean within-cluster similarity; an ideal result will show white on the ascending diagonal with black elsewhere. For example, the diagonal cell of cluster $j = 9$ is the closest to white, and this indicates that mean within-cluster similarity of $j = 9$ was higher than the mean within-cluster similarity of any other cluster, and since there are no negative bars in this cluster ($j = 9$, in Fig. 3.7a), all subsamples within this cluster were more similar to each other than to any other cluster. As well, cluster $j = 5$, composed of crust subsamples from the same surface sample, had very little in common with any sinter clusters compared to other crust clusters, as indicated by the dark cells to the left and below.

Convincing results from ANOSIM, along with general agreement from partana and silhouette analysis indicate that within-sample community similarity was greater than across any other combination. Thus subsamples were combined into their respective samples to address the second and third hypotheses.

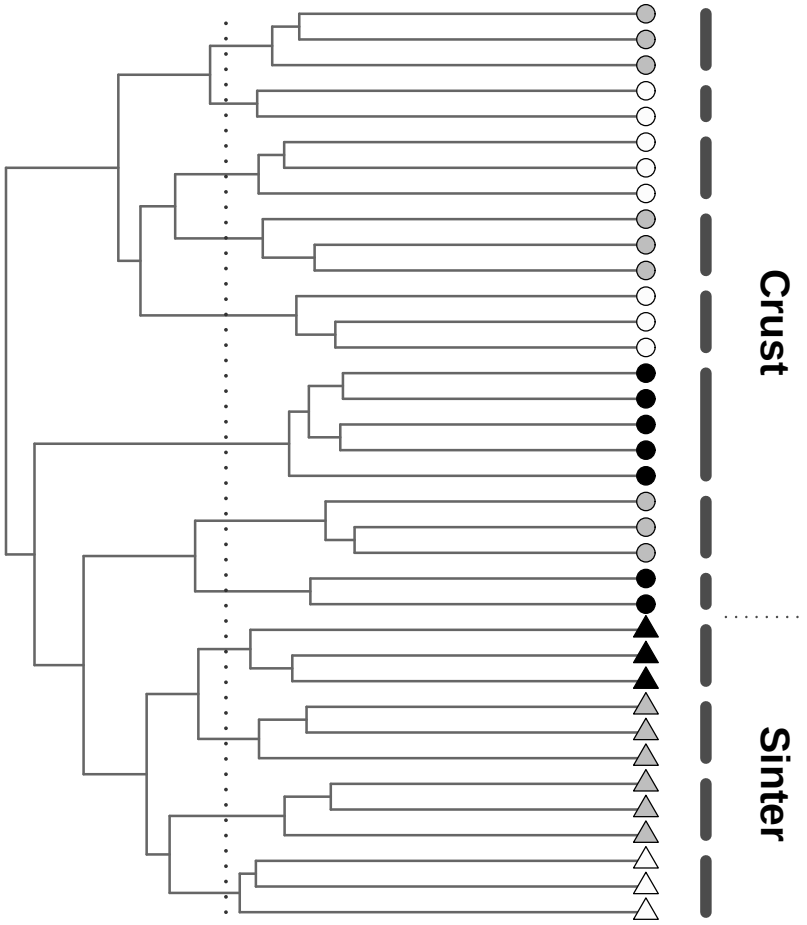

Figure 3.6. Agglomerative hierarchical clustering of subsamples. Samples were clustered using a flexible-β approach, and separated into clusters by optimizing results from both partition and silhouette analysis, resulting in 12 clusters (vertical dotted line). Crust subsamples are represented by circles, while sinter subsamples are triangles; top, mid and deep photic depths are white, gray and black filled symbols, repectively. Clade bars indicate clusters separated in a 12 cluster solution. The resulting clustering solution completely separates sinter and crust subsamples, while also grouping subsamples surprisingly well. The only exception to the original members of each sample is the 5-subsample "deep" cluster (seventh cluster from the bottom), which includes subsamples of two deep samples.

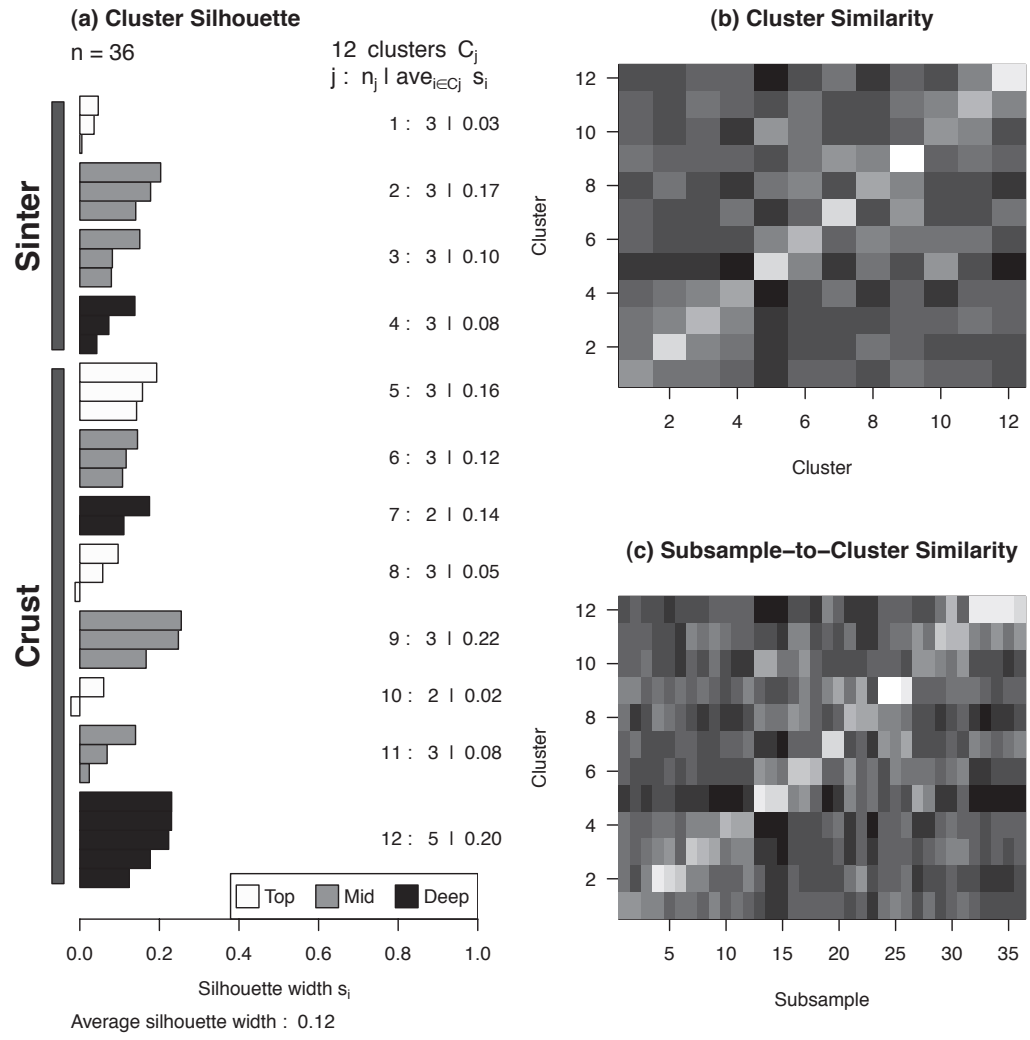

Figure 3.7. Silhouette and partition analysis results for subsample data. (a) The left panel shows silhouette widths for each of 12 clusters created by agglomerative hierarchical clustering. C_j = number of clusters j; n_j = number of members in each cluster; $ave_{i \in Cj}\ s_i$ = average within-cluster similarity agreement. Clusters 1–4 are composed of sinter particles, while all other clusters are composed of crust subsamples. White, gray and black bars represent top, mid and deep samples, respectively. (b & c) The two panels on the right show results from partition analysis (partana). Numbers on the axes refer to n_j and n, for clusters and subsamples, respectively. The ratio of within-cluster similarity versus between-cluster similarity is indicated in grayscale, with white indicating a high partana ratio and black indicating a low partana ratio. An ideal result would show white across the ascending diagonal with black elsewhere. The 12 cluster solution optimized the values from both analyses. Cluster (j) order in silhouette graph (a) is not meaningful, but does correspond to axes of partana graphs (b & c).

Table 3.3. Results from Analysis of Similarities (ANOSIM). R-statistic is based on the difference of mean ranks between groups and within groups, and ranges from -1 to 1 with 0 indicating completely random grouping and positive values indicating systematic grouping (Legendre & Legendre 1998). P-values are based on 999 iterations, and are only reported down to the P = 0.001 level. "Ungrouped" subsamples correspond to the clustering solution presented in Figure 3.7. "All Grouped" samples contain all eukaryotic OTUs, and "Diatoms Grouped" contains only diatoms.

Data	Grouping	P-value	R-statistic
Ungrouped	Subsamples	<0.001	0.97
All Grouped	Sinter/Crust	<0.001	0.46
	Vertical	0.003	0.26
	Both	<0.001	0.63
Diatoms Grouped	Sinter/Crust	0.033	0.18
	Vertical	0.016	0.22
	Both	0.002	0.40

Eukaryotic Combined Samples

DB-RDA of combined samples (Fig. 3.8), constrained by both spatial variables, reveals thorough separation of sinter and crust samples, and a tight grouping of photic levels within soil type. Unconstrained dissimilarity accounted for most of the inertia (76% versus 24% constrained inertia; inset pie-chart of Fig. 3.8); the majority of information used in ordination mapping was from community dissimilarity data, and the two variables used as constraints did not markedly disagree. ANOSIM was used to test these groupings, and all three clustering solutions were convincingly significant (Table 3.3), indicating that both scales were associated with shifts in community composition, but the combination of the two variables into six clusters resulted in the best fit ($R = 0.63$; $P < 0.001$).

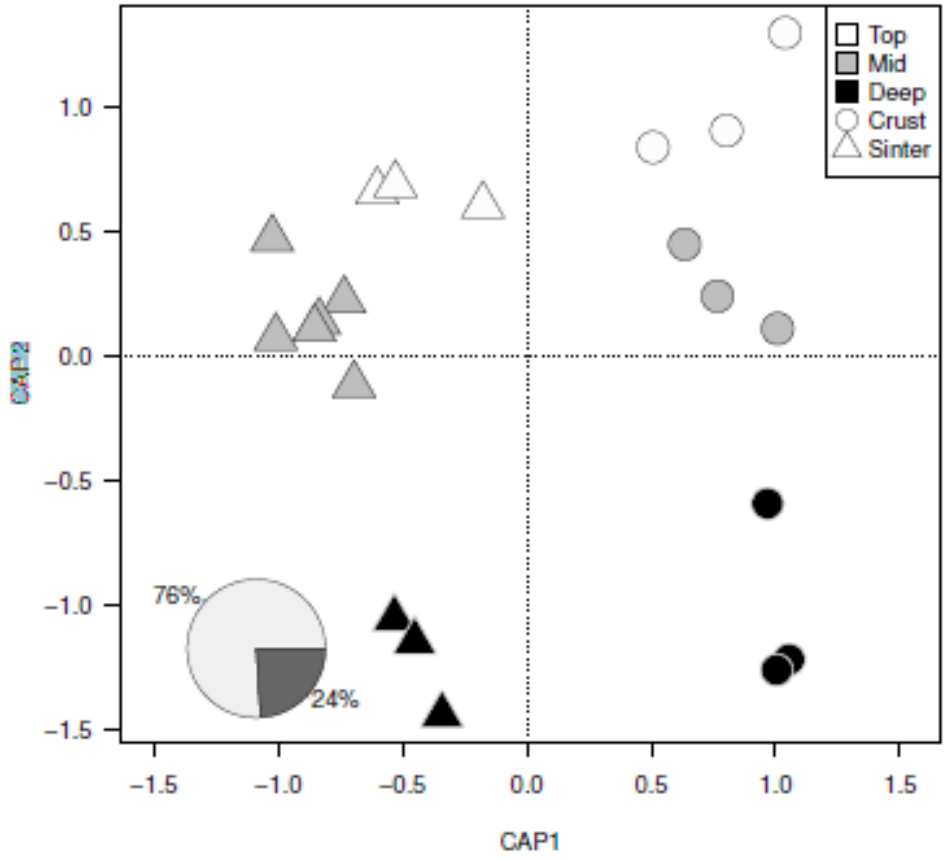

Figure 3.8. Distance Based Redundancy Analysis (DB-RDA) for all eukaryotes. DB-RDA is constrained ordination, and, in this case, Canberra dissimilarity is constrained by both soil type and photic depth. Inset pie-chart shows the percentage of total inertia by unconstrained (76%; light gray) and constrained (24%; dark gray) inertia.

Diatoms

The diatom-only dataset was substantially smaller (6930 total sequences and just 13 taxa), and results were somewhat more nuanced compared to the whole eukaryotic community. Fig. 3.9 shows a pattern similar to that seen for the whole eukaryotic community with less complete grouping of samples, and some mixing between both soil types and photic levels, though "deep" samples were well clustered. ANOSIM

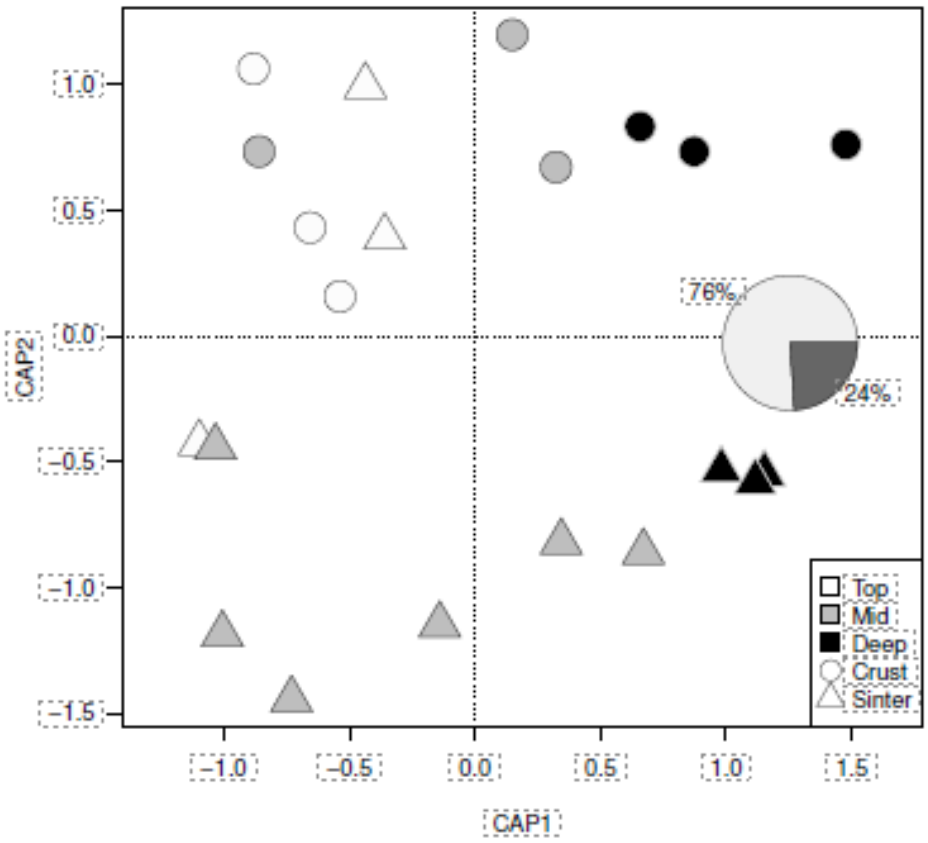

Figure 3.9. Distance Based Redundancy Analysis (DB-RDA) for diatoms. Separation of samples along CAP1 (first axis) indicates that the most important variable considered (in terms of constraining sample mapping positions) is photic depth rather than soil type, though soil type separates almost completely along CAP2. This is in contrast to both the axis orientation and sample grouping displayed in Figure 3.8.

tests of the three same clustering solutions were also all significant, though the soil type-based grouping was less convincing ($R = 0.18$; $P = 0.033$; Table 3.3).

Spatial Heterogeneity

As additional support for our first hypothesis that eukaryotic microbial community similarity was highest within soil samples and increased with distance, a quadratic

Table 3.4. BLAST Identification of Diatoms. Each of the 13 taxa used in analysis of the diatom-only dataset was identified by closest matching isolate using BLAST.

Isolate Species	GB Accession	Similarity (%)
Anomoeneis sphaerophora	AJ535153	99
Cocconeis stauroneiformis	AB930614	98
Cymbella signata	JN033245	98
Denticula kuetzingii	HQ912610	99
Gyrosigma aciminatum	HQ912598	97
Haslea crucigera	AY485482	98
Navicula phyllepta	FJ624253	98
Nitzschia communis	AJ867014	99
Nitzschia cf. supralitorea	AJ867020	99
Pinnularia brebissonii	HQ912604	98
Pinnularia rupestris	AJ867027	98
Pinnularia viridiformis	AM501985	98
Sellaphora pupula	EF151982	98

model was fit to describe the relationship between the log of spatial distance and Canberra Dissimilarity using subsample data; the model was highly significant ($P <$ 0.0001 for all coefficients; $t = 229.1$, 18.49 and -7.49, for intercept, first-order slope, and squared term, repectively; Fig. 3.10), and the estimated trend increases asymptotically over the spatial scale of the study area. This helps to confirm that eukaryotic communities were relatively similar within soil cores, but that a microbial community survey on the single aggregate scale could be more informative in terms of fine-scale community heterogeneity than a whole soil core, depending on the context and the desired inference of the study.

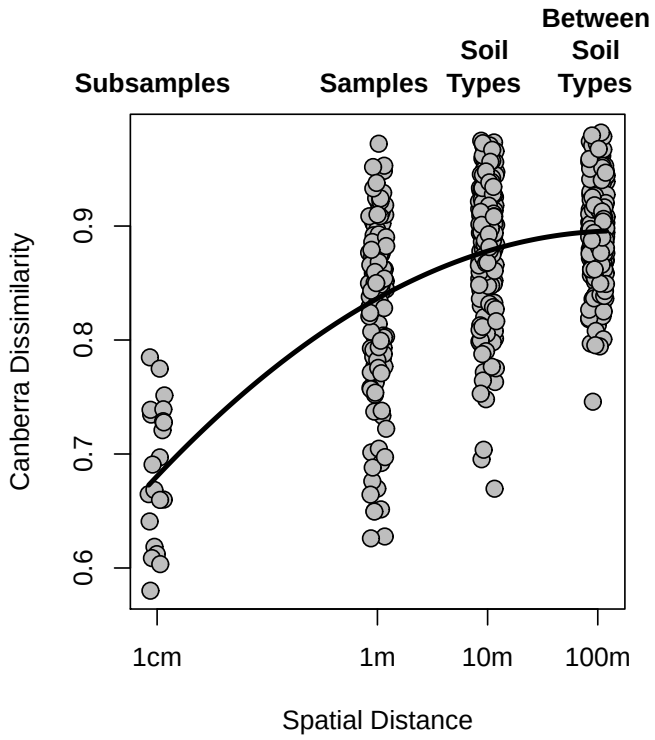

Figure 3.10. Subsample dissimilarities regressed against distance. The curve is from a quadratic model ($d_{Canberra} \sim log(Dist) + log(Dist)^2$), where *Dist* is generalized spatial distance (corresponds to Table 3.5). The relationship displayed was highly significant for all model terms.

Table 3.5. Results from quadratic model ($d_{Canberra} \sim log(Dist) + log(Dist)^2$) where *Dist* is generalized spatial distance (corresponds to Figure 3.10). *P*-values were multiplied by the total number of pairwise comparisons (640) to account for non-independence.

Model Term	Estimate	Std. Error	t-value	P-value
Intercept	0.836	0.004	229.08	<0.001
log(Distance)	0.023	0.001	18.49	<0.001
log(Distance)^2	-0.002	0.0003	-7.49	<0.001

Discussion

One of the most unique features of this site is that it is dominated by diatoms, both in OTU abundance and frustular residuum. Taxa from most major eukaryotic microbial lineages were detected, though 67% of all sequences analyzed were from diatoms. To our knowledge, a frustule-based geothermal soil system has never been described, though diatom communities have been studied in acidic geothermal springs (Seckbach & Kocielek 2011; Owen *et al.* 2008; Hobbs *et al.* 2009). The eye-catching crusted area of this study site displays many of the defining characteristics of biological soil crusts (BSC), except that the soils remain moist throughout the year and the fibrous taphonized diatomaceous soil matrix is not prone to the same erodible instability. We also found no lichens or mosses, either visually or by 18S sequence; both are major constituents, as well as indicators, of well-developed arid BSCs. These features, along with the fact that the entire study area is dominated by diatoms, are in stark contrast to any other BSC system previously studied. Throughout the study area, the soil is exposed to full sunlight since essentially no plants shade the surface, and this is evident in the conspicuous phototrophic microbial communities. Though porosity was not directly measured here, frustule residuum and porous opaline precipitation, along with high soil-C content ($\approx$ 5%; unpublished data) both improve water-holding capacity, and likely contribute to the moist nature of the soil which in turn helps to support the microbial communities that distinguish this site. Sinter soils surrounding the geothermal source pool are visually different from crusted areas (Fig. 3.2a & b), and eukaryotic microbial community data indicate that the visual

differences are associated with a biotic shift. FE-SEM imaging reveals a complex and highly porous sinter surface mostly composed of opaline-Si, taphonized microbial biomass, and diatomaceous residuum. While this is qualitatively similar to the crust parent material, diatoms appear to make up far less of the sinter soil material, both visibly and in 18S sequences. Since geothermal water from a common source is a major soil formation factor in these soils, chemical conditions are similar throughout the site, and are not presented here. Rather, the major difference between sinter and crust soils appears to be history; sinter particles comprise a younger, more porous and aerated soil, while crusts have the added history of diatom and wetland soil material accumulation, creating a more dense soil matrix. This historical and physical disparity likely contributes to the observed differences in microbial communities between the soil types. To our knowledge, this is the first whole-eukaryotic survey of a soil microbial system using high-throughput sequencing, though similar work has been conducted in marine (Edgcomb *et al.* 2011) and freshwater (Brate *et al.* 2010) systems and for subsets of soil (Lim *et al.* 2010) and aquatic (Amaral-Zettler *et al.* 2009) eukaryotic microbial communities. Pyrosequencing has been recently used to describe the prokaryotic communities associated with the lichen symbiosis from rocks and soils (Bates *et al.* 2011a), as well as the lichen associated eukaryotic communities of BSCs (Bates *et al.* 2012), but whole eukaryotic communities have otherwise not been described for either BSCs or other soil microbial systems. Our bar-coded pyrosequencing approach, using the F-907 and R-1428 primer set (Troedsson *et al.* 2008), was successful at detecting eukaryotic sequences from most major lineages, and thus represents a model for evaluating eukaryotic microbial communities in other

systems using high-throughput sequencing. The target 18S region, though, allows for poor taxonomic resolution for many taxa considered here, including diatoms, and thus might be a limiting factor for some applications; this is a common problem when analyzing short sequence reads.

We conducted this study not just to describe an unstudied soil system, but also to utilize this model phototrophic soil system to investigate the spatial heterogeneity of eukaryotic soil microbial communities. As high-throughput sequencing technologies become increasingly utilized, questions have been raised as to the scales of inherent heterogeneity in microbial communities being studied, in terms of ecological relevance, as well as statistical independence and inference (Prosser *et al.* 2007). Though this debate has received more attention in microbial ecological studies recently (Prosser 2010; Lennon 2011), macroecologists have long dealt with these issues (Hurlbert 1984). Our goal was to compare the degree of within-sample community resemblance to between-sample community resemblance, to gain a better understanding of the relevant heterogeneity level in soil microbial communities. At the finest scale of our study, subsamples were clustered using a parsimony optimization of clustering solutions, and were regrouped nearly perfectly into their respective samples. An analysis of similarities test resulted in an extraordinarily high R-statistic ($R = 0.971$; $P < 0.001$), indicating an adequate fit, and corroborates the assumption that within-sample similarity is higher than between sample similarity, even between vertical layers within the same soil core, in the case of crust soil cores. After subsamples were regrouped appropriately, the remaining sample dissimilarity separated along both the vertical photic gradient (second axis of Fig. 3.8) and soil type (first axis

of Fig. 3.8), though the combination of the two variables results in the best fit for the whole eukaryotic community dataset. Diatom communities were considerably less discriminating across both photic depth and soil type, though fewer rarefied sequences were used for diatom analysis. β-diversity clearly depends on the quantity of data used; however, very few sequences per sample can effectively detect differences between communities (Kuczynski *et al.* 2010).

Results from an explicit test of pairwise community dissimilarity as a factor of spatial distance, were highly significant (Fig. 3.10 & Table 3.5). We generalized spatial distance along a log scale to account for inherent differences between sinter and crust replicate distances for the "Samples" set (i.e., single photic-depth particles for each sinter replicate were taken from within a $1m^2$ area while crust photic-depth samples were within replicate soil cores; we used the 1m scale to generalize spatial distance for this set). There was a strong increase in pairwise dissimilarity as spatial distance increased, though an asymptote is approached when subsamples are compared across soil types. It is important to note the dissimilarity values within subsamples: all pairwise values of non-independent subsamples were >50% dissimilar, pointing to the potential inference gained from a fine-scale microbial community survey. Current molecular methods certainly allow this targeted approach, in cases where the additional detail informs a given study.

For datasets derived from soil microbial communities, we tend to assume that homogenizing a soil core or composite of soil cores adequately represents an interacting soil microbial community. In fact, this sampling regime ignores the heterogeneity

of diverse habitats, as would taking a subsample of plant species from a homogenized sample across all vegetation types in North America. In cases where microbial communities are compared over regional or continental scales (Lauber *et al.* 2009), fine-scale heterogeneity can add unnecessary noise and erroneously focus on relatively unimportant localized conditions. On the other hand, fine-scale studies that seek to understand localized soil processes and their effects on soil microbial communities can certainly gain from an understanding of these differences at the soil aggregate scale. Our subsamples were approximately $3mm^3$, and were limited in size to attempt to capture the approximate heterogeneity in a single soil aggregate. The level of community dissimilarity among subsamples was predictably and substantially lower than across any other scale, indicating that, though non-independent subsamples contain distinctly different communities, their communities are significantly less dissimilar than across any other scale we investigated. This finding adds to the growing understanding of the information that is lost (or avoided) by sampling soil microbial communities in soil cores and composite samples versus the community that might be detected at the aggregate or soil particle scale.

CHAPTER 4.

PROKARYOTIC COMMUNITIES DIFFER ALONG A GEOTHERMAL SOIL PHOTIC GRADIENT

Contribution of Author and Co-Author

Manuscript in Chapter 4

Author: James F. Meadow

Contributions: Conceived the study, obtained funding, collected and analyzed output data, and wrote the manuscript.

Co-author: Catherine A. Zabinski

Contributions: Assisted in collecting data, and assisted with study design and discussed the results and implications and edited the manuscript at all stages.

Manuscript Information Page

James F. Meadow and Catherine A. Zabinski
Journal Name: Microbial Ecology
Status of Manuscript:
___ Prepared for submission to a peer-reviewed journal
_✓Officially submitted to a peer-reviewed journal
___ Accepted by a peer-reviewed journal
___ Published in a peer-reviewed journal

Prokaryotic communities differ along a geothermal soil photic gradient

James F. Meadow (corresponding author)
Montana State University
Department of Land Resources and Environmental Sciences
334 Leon Johnson Hall
Bozeman, MT 59717
jfmeadow@gmail.com
ph: (406) 370-7157
fax: (406) 994-3933

Catherine A. Zabinski
Montana State University
Department of Land Resources and Environmental Sciences
334 Leon Johnson Hall
Bozeman, MT 59717

Abstract

Geothermal-influenced soils exert unique physical and chemical limitations on resident microbial communities, but have received little attention in microbial ecology research. These environments offer a model system in which to investigate microbial community heterogeneity and a range of soil ecological concepts. We conducted a 16S bar-coded pyrosequencing survey of the prokaryotic communities in a diatomaceous geothermal soil system, and compared communities across soil types and along a conspicuous photic depth gradient. We found significant differences between the communities of the two different soils, and also predictable differences between samples taken at different depths. Additionally we targeted three ecologically-relevant bacterial phyla, Cyanobacteria, Planctomycetes, and Verrucomicrobia, for clade-wise comparisons with these variables, and found strong differences in their abundances, consistent with the autecology of these groups.

Key words: geothermal soil; biological soil crust; prokaryotic microbial community; fuzzy set ordination; adonis; Cyanobacteria; Planctomycetes; Verrucomicrobia

Introduction

Prokaryotic communities in soils are among the most taxon-rich of any microbial habitat (Madigan *et al.* 2008; Tringe *et al.* 2005; Fierer & Jackson 2006), and the abiotic heterogeneity in soils is a major contributor to their biological diversity (Fierer & Lennon 2011). For example, soil bacterial communities can differ on the millimeter scale due to localized pH changes, oxygen diffusion gradients, uneven distribution of soil nutrients, solar radiation, surface temperatures, substrate limitations and soil-water relations. Salinity and pH are major drivers of microbial communities in all habitats, and especially in soils (Lauber *et al.* 2009; Herlemann *et al.* 2011). Detecting and studying fine-scale differences in soil microbial communities has progressed quickly given recent advances in high-throughput DNA sequencing technology, and much of the analysis of these massive datasets has borrowed from macroecology analysis (Kuczynski *et al.* 2010). Community-wide phylogenetic relationships are especially useful in analyzing bacterial communities since high-level taxonomic groups show ecological coherence (Horner-Devine & Bohannan 2006; Philippot *et al.* 2011), and since phylogeny effectively incorporates evolutionary history into our understanding of community ecology.

Plants exert strong controls on the composition of bacterial communities in vegetated soils (Wardle 2002; Zak *et al.* 2003; Ehrenfeld *et al.* 2005), but in the absence of plants, abiotic conditions likely hold stronger influence over microbial primary productivity, and thus on dependent communities. Biological soil crusts (BSCs) are consortia of microorganisms that dominate soil cover in many arid soil systems where

environmental conditions limit plant cover (Belnap & Lange 2003). In some arid soil systems, BSC-associated cyanobacteria account for most C- and N-fixation and thus they are sometimes the largest contributors to primary productivity in these systems (Housman *et al.* 2006). BSCs are potentially useful for testing ecological theory in microbial communities because their constituents are well-studied, and because a breadth of trophic levels is present in a small space (Bowker *et al.* 2009).

Soils under geothermal influence often hold limited plant cover due to chemical and thermal conditions that are beyond the tolerance of most vascular plants (Bunn & Zabinski 2003); the eukaryotic component of an unstudied geothermal BSC was described in Chapter 3, in a soil with full solar radiation due to essentially no plant cover. This soil community is dominated by diatoms in both 18S sequences and in diatom-derived frustular residuum. In the present study, we analyzed prokaryotic community composition by contrasting two adjacent soil types across a photic depth gradient. Our first objective was to assess prokaryotic community differences between two adjacent soil types and along a visible soil photic gradient. Secondly, we investigate specific differences in three ecologically-relevant taxonomic groups of the prokaryotic community across the study system.

Methods

Site Description and Sample Collection

All soil samples were taken from Imperial Meadow (centered around 44°33′04″ N, 110°51′05″ W) in the Lower Geyser Basin, Yellowstone National Park, WY, USA.

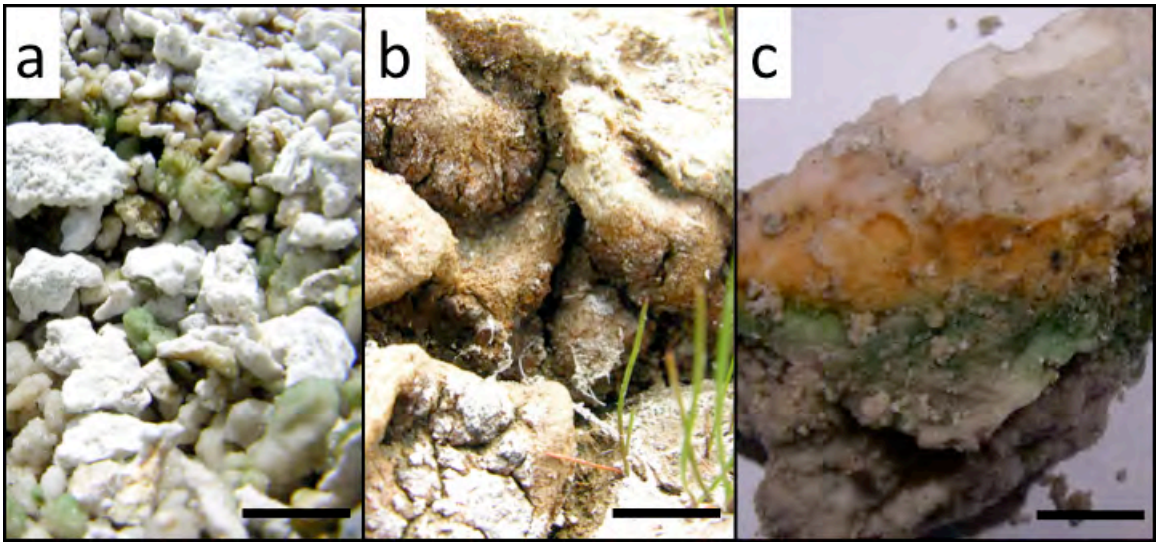

Figure 4.1. Diatomaceous geothermal soils: (a) Sinter particles (disturbed) with white on the surface above orange and green photic zones (scale bar = 1cm); (b) Fibrous, dense diatomaceous soil crust (scale bar = 3cm). This is the general appearance and surface texture of crust samples; (c) Vertical photic layering (scale bar = 1cm). This striation was observed in similar appearance in both sinter and crust soils, across the study area.

These biologically crusted soils represent a history of influence from alkaline-siliceous geothermal water from a single source spring in the Lower Geyser Basin of Yellowstone National Park, WY, USA (YNP Thermal Inventory ID LFMNN020; Rowe *et al.* 1973). Alkaline water from the source spring (pH = 9.7) contains little N or P, but is elevated in several biologically important elements, including As, B, Na, and Si (see Chapter 3 for a description of geothermal water chemistry at this site). The spring water occasionally inundates the adjacent soils, but these events appear to be infrequent. Sinter soils (Fig. 4.1a) are composed of coarse, white, taphonized siliceous particles adjacent to the spring, while crust soils (Fig. 4.1b) are dense, fibrous remnant wet meadow substrate largely composed of diatomaceous residuum and opaline silica. Sinter and crusted soils are within a 3 ha area surrounding the source spring. The

only vascular plant present across the site is *Triglochin maritima* L., and its coverage is sparse; this results in a soil surface exposed to full solar radiation, and thus a colored phototrophic microbial community along a depth gradient that is conspicuous in both soil types (Fig. 4.1c). There are several clear differences between the two soil types, primarily in soil texture and structure; sinter soil particles are relatively more coarse (mean particle diameter $\approx$ 3mm), while crust soils are primarily composed of taphonized *T. maritima* root material and diatomaceous frustular residuum, both a product of history under a geothermal wet meadow. This difference lends to a generally dry, aerated surface in sinter soils, and a moist, microbially active surface in crust soils. Both soil types are very alkaline, with pH $\approx$ 10.

All samples used in this study were taken in October, 2010. Three individual sinter particles ($\approx$ 3 mm diameter) were sampled from each of the representative photic depths using sterile forceps. All sinter particles were sampled from within 10m of one another, with one replicate set containing a particle from each photic depth taken from within a 1m area. Mid-depth sinter samples consist of an extra set of replicate particles, since sinter particles were more ambiguous in orange versus green coloration (Fig. 4.1). These were preliminarily tested for community differences, and were subsequently lumped into one mid-depth group for final analysis. Likewise, deep-depth crust samples were also separated in preparation and grouped for this analysis. Crust samples were taken from three different locations within the upland crusted soils, each within 10m of another and approximately 100m from sinter sampling locations. Crust soils were sampled with a sterilized soil core ($\approx$ 2 cm diameter), and were sampled to

5 cm depth. All samples were frozen on dry ice, on site, and archived at -80°C until DNA extraction.

DNA Extraction

Crust soil cores were cut along conspicuously colored photic boundaries with a sterile scalpel. Edges of the photic slices that touched the soil corer, as well as transition regions between conspicuous colors were removed so as to reduce cross-contamination from adjacent photic levels. These photic slices ($\approx$ 0.3 cm thick by 1cm wide) were cut into three equal pieces for subsampling. All subsampling was performed in a sterile hood. Whole genomic DNA was extracted from each sinter particle and crust subsample using the FastDNA Spin Kit for Soil from MPBio (www.mpbio.com) following manufacturer's instructions, with an additional ethanol precipitation step. Subsampling was performed to assess within-sample community heterogeneity, but these results are not presented here; subsamples were recombined *in silico* for these analyses.

Bar-coded Pyrosequencing.

The V4 region of the 16S small-subunit ribosomal gene was amplified using the 515F primer (5′-GTGCCAGCMGCCGCGGTAA-3′) containing a 10-bp bar-code sequence with a Roche 454-A pyrosequencing adapter (Titanium Lib-L adapters), and the 806R primer (5′–GGACTACVSGGGTATCTAA-3′) with a Roche 454-B adapter. The region targeted by this primer pair is suitable for β-diversity comparisons using short pyrosequenced reads (Liu *et al.* 2007), and has been adopted for standard

protocols by the Earth Microbiome Project (earthmicrobiome.org). Unique bar-code sequences were used to distinguish between all samples and subsamples. PCRs were conducted with 10μM of each forward and reverse primer, 3μl template DNA, 22μl GoTaq Green Master Mix (Promega, Madison, WI, USA), and an additional 1μl BSA. Three samples, all mid-depth crust subsamples, contained noticeably higher DNA concentrations than other samples, and were diluted 5$\times$ with DNA-free water to achieve consistent amplification results with 3μl template volume. PCR was conducted with the following thermocycler conditions: 3 minute denaturation at 94°C; 30 cycles of 45 seconds at 94°C, 30 seconds at 50°C, and 90 seconds at 72°C; and an additional 10 minute annealing step at 72°C. Each sample and subsample was amplified in triplicate, pooled and cleaned using a QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA, USA). Equal concentrations of cleaned PCR product were combined and sent to the Center for Integrated BioSystems at Utah State University to be run on a Roche FLX Ti 454 pyrosequencing machine (www.biosystems.usu.edu). Raw sff files from this project have been deposited in the NCBI Short Read Archive under accession SRA050273, and bar-code mapping files in QIIME format are included as supplemental material.

Sequence Processing.

All sequence processing was performed using QIIME v. 1.3 (Caporaso *et al.* 2010b). Quality-filtered sequences were clustered at the 97% similarity level, and taxonomy was assigned to representative sequences from each cluster using the RDP classifier; taxa that were represented by fewer than 10 high-quality sequences were

removed from analysis. Sequences were assigned to samples using their unique 10-bp bar-codes. The community dataset was rarefied to 1500 sequences per sample before proceeding with β-diversity analysis.

Phylogeny and Statistical Analysis

OTU sequences were aligned in QIIME using PyNAST (Caporaso *et al.* 2010a), and a subsequent phylogenetic tree was built with FastTree (Price *et al.* 2009). The tree and rarefied community datasets were used to calculate unweighted UniFrac community dissimilarity with the `picante` package in `R` (Lozupone *et al.* 2006; Kembel *et al.* 2010; R Development Core Team 2010).

UniFrac distance values were used for ordination and β-diversity assessments, all carried out in `R`. To visualize additional phylogenetic influence gained by using the UniFrac method, the UniFrac matrix was compared to distances from a geometric qualitative index, Ochiai, and a similar quantitative index, Canberra, both as implemented in the `vegan` package (Oksanen *et al.* 2011).

Non-metric multidimensional scaling (NMDS) was used for unconstrained ordination of UniFrac distances, on 2 axes, using iterative random starts with the `bestnmds` function in the `labdsv` package (Roberts 2010b). Fuzzy set ordination (FSO) was used as constrained ordination to visualize the associations between each individual variable, as well as their combination; FSO was performed with the `fso` package (Roberts 2010a). Soil type- and gradient-based groups were tested with permutational MANOVA using the `adonis` function in the `vegan` package.

All OTUs in the combined sample dataset were tested for association to either soil type or depth category by χ^2 test, and those with P-values < 0.1 for any group were used for phylum-level analysis. These sequences were aligned with ClustalW, and used to construct the ML tree ($GTR + \Gamma + I$; 1000 iterations), both in Mega5 (Tamura *et al.* 2011). False discovery rates were not used to select these taxa with χ^2, since the objective was simply to limit OTUs to those showing marginal association with a particular group, and not to assign significance to any particular taxon. The maximum likelihood tree was used to extract whole clades for phylum-level phylogenetic analysis. OTU abundances were used to test clade-wise associations with analysis of variance; multiple comparisons were performed with Tukey's Honestly Significant Difference tests, and P-values were adjusted for multiple comparisons. Both of these functions, `aov` & `TukeyHSD`, along with `chisq.test`, are in the `stats` package in `R`. Box-plots of differences were produced with `ggplot2` (Wickham 2009).

Results

Sequence Processing

After quality filtering with QIIME default settings, 156,213 sequences were retained (average sequence length = 300.4nt), resulting in 15,370 OTUs clustered by RDP classifier at the 97% similarity level. The final rarefied dataset (1500 sequences per sample; 24 samples) contained 1314 OTUs. Figure 4.2 is a breakdown of the distribution of prokaryotic phyla in each sample. The figure is simplified to include only those phyla that made up more than 1% of any given sample, and all others

are included in the "Other" category, which also contains OTUs not identified at the phylum level. The most striking difference among samples is that cyanobacteria were primary constituents of some samples, and they were predictably abundant along the photic gradient with a clear difference in their positional abundances between soil types.

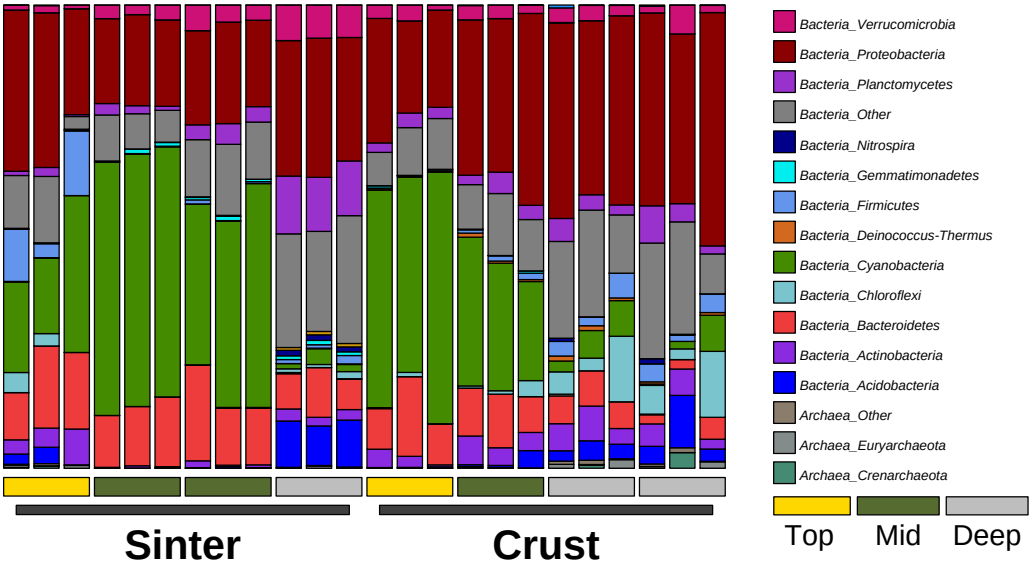

Figure 4.2. Breakdown of the relative abundance of prokaryotic phyla for OTUs found in this study, after rarefaction. For simplification, the legend includes only those phyla that comprised >1% of any sample.

Ordination and Discriminant Analysis

Before proceeding with ordination, we assessed phylogenetic influence when using the unweighted UniFrac distance rather than a more traditional non-phylogenetic index (Supplemental Fig. 1). We found surprising agreement between UniFrac and the Ochiai distances, and similar agreement when compared to the abundance-based

Canberra index. Though Canberra allows greater dissimilarity due to abundance differences, both are comparable; we used the UniFrac metric for all further multivariate analyses since the small amount of information gained by considering phylogeny is a step toward incorporating community ecology and evolution. Final stress of the 2-dimensional NMDS (Fig. 4.4) was 8, below the conventional problematic cutoff of >15. Points are coded by soil type (shape) and depth (grayscale), with replicate points joined by gray lines. There is a predictable grouping by both soil type and by depth, with only one clear exception: two replicate sinter surface particles are not grouped accordingly, but rather with mid-depth crust samples.

Four fuzzy set ordinations were produced, and are displayed in Fig. 4.5. The purpose of FSO is a single-dimensional comparison of community data (UniFrac phylogenetic distance, in this case) with a fuzzy set ordination value (y-axis); the ordering of the x-axis is categorical (a) and ordinal (b & c), and the degree of separation along the y-axis indicates community agreement with x-axis data (Roberts 1986, 2008). The three combinations tested were: (a) soil type; (b) depth; and (c) the combination of both categorical variables (FSO correlations and P-values are reported in Table 4.1). While the results from these FSOs are convincing on their own, we also formally tested the three sets of groups with permutational MANOVA using `adonis`. It is important to note that we ran the combination of the two variables as an interaction with `adonis`, while FSO was set up as two different sets for the "interaction" FSO (Fig. 4.5), since this produces a more straightforward ordination. All three combinations tested with `adonis` were significant (P-values < 0.001 in every case; Table 4.2). R^2 values from `adonis` are shown in each panel of Fig. 4.5, and the third

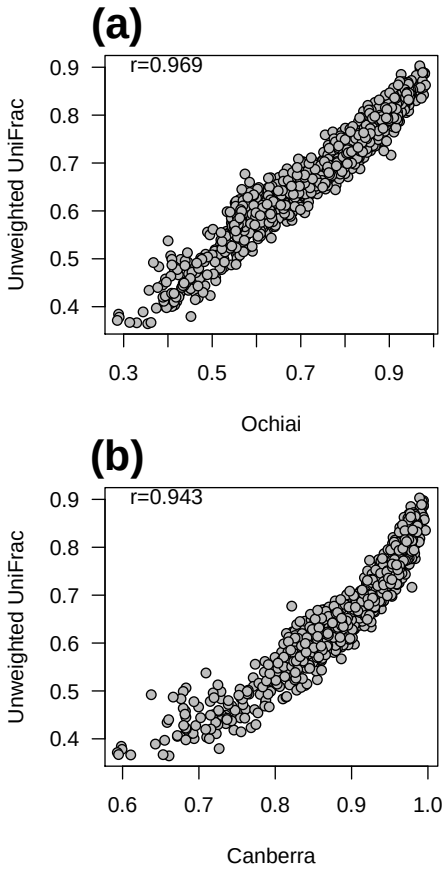

Figure 4.3. Supplemental Figure 1. Comparison of phylogenetic unweighted UniFrac distances to qualitative Ochiai (a) and quantitative Canberra (b). Correlation coefficients (r-values) indicate that limited additional information was gained by considering phylogenetic relationships.

combination is clearly the most explanatory of the variation in the data. Soil types show a remarkably consistent ordination pattern in Fig. 4.5c, with overlap between surface and mid-depth samples, and more separation of deep samples from all others.

Phylum-Level Comparisons

For comparisons of individual taxonomic groups, we limited the dataset to only those taxa that were more abundant in any combination of soil type and depth. This

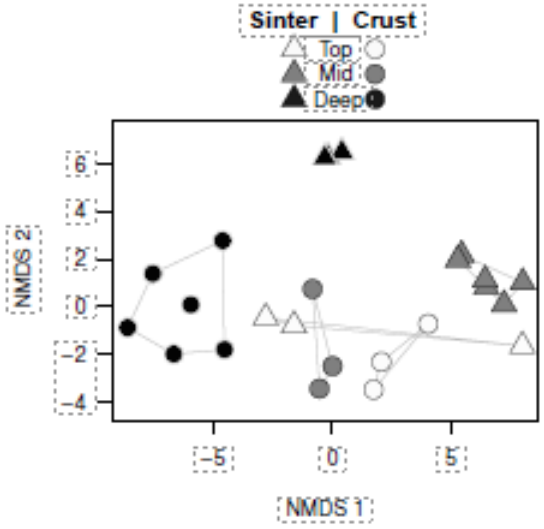

Figure 4.4. Non-metric multidimensional scaling (NMDS) of samples across soil types (shape) and depth (grayscale). Replicates in each group are joined by gray polygons. Final stress after iterative random starts was 8.0.

Table 4.1. Fuzzy set ordination (FSO) model results. d-values represent fuzzy set correlations between variables and community dissimilarity data. P-values estimate the probability of obtaining a correlation as large as observed.

Model	d	P-value
Soil Type	0.75	< 0.001
Depth	0.72	< 0.001
Sinter × Depth	0.59	0.001
Crust × Depth	0.72	0.002

was done to focus on OTUs that appeared to consistently prefer specific conditions, and also to reduce the size of the dataset used to construct an ML tree; the latter was necessary to group OTUs by clades, since many taxa were not identified below, or even to, the phylum level. Out of the total 1314 OTUs in the original rarefied dataset,

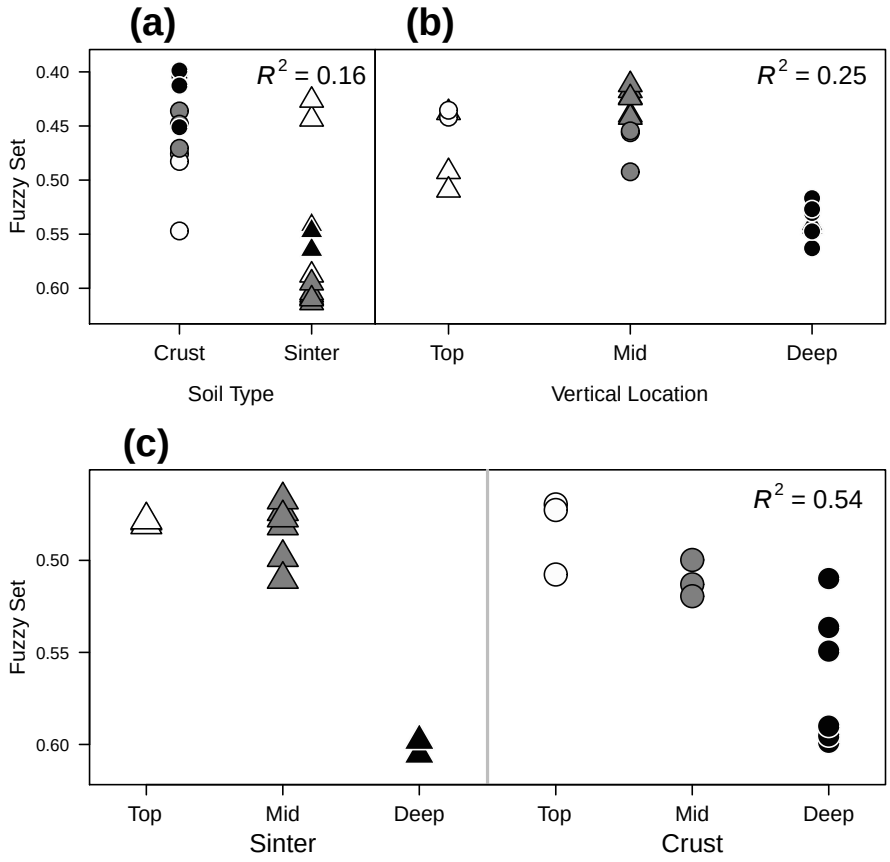

Figure 4.5. Fuzzy set ordination (FSO) of (a) soil type; (b) depth; and (c) both variables. On the y-axis are fuzzy set values that describe the agreement between UniFrac community dissimilarity and x-axis variables for each sample. R^2 values are from permutational MANOVA; the third value (c) is cumulative R^2 of an interaction model.

432 were retained for clade-wise comparisons. We selected three ecologically relevant bacterial phyla that were preferentially abundant in some combination of samples: Cyanobacteria, Planctomycetes and Verrumicrobia (Fig. 4.6). Results from Tukey's HSD tests for all combinations are shown in Table 4.3; P-values were adjusted for multiple comparisons. Cyanobacteria were most abundant in the mid-depth sinter particles and surface to mid-depth crust samples (Fig. 4.6a & b). This is perhaps

Table 4.2. Results from permutational MANOVA tests on both variables individually, as well as their interaction. Interaction R^2 is cumulative (full - residual R^2) from both factors and their interaction, while both single factor models were run separately.

Model	R^2	P-value
Soil Type	0.16	< 0.001
Depth	0.25	< 0.001
Soil Type $\times$ Depth	0.54	< 0.001

indicative of the surface environments in either soil type; sinter surface particles are coarse and porous with lower moisture retention potential and organic matter content than the surface of the crust soils, and thus surface crust soils present a more hospitable habitat for cyanobacterial growth. Planctomycetes and Verrucomicrobia (Fig. 4.6c–f) were both more abundant on subsurface sinter particles, with relatively low abundance in surface particles. Both groups were present in all crust samples, but no significant differences were found between depths.

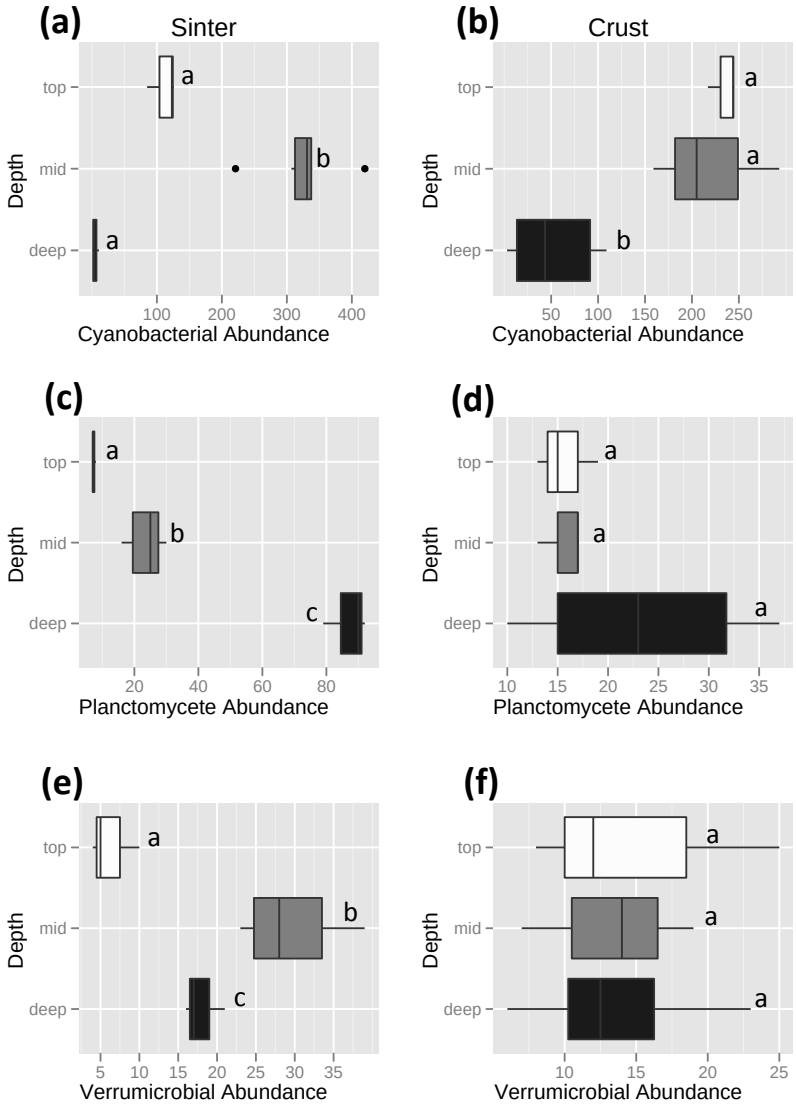

Figure 4.6. Clade-wise phylum comparisons. Plots on the left side (a, c & e) show abundance in sinter samples; crust samples are in right-hand plots (b, d & f). Cyanobacterial OTUs (a & b) were more common in photic zones, though surface sinter samples contained fewer than mid-depth samples. Planctomycete abundance increased with depth in sinter samples (c), but there was no significant difference between crust depths (d). Verrucomicrobial abundance was highest below the surface in sinter samples (e), and was consistent throughout crust depths (f). Letters beside boxes are results from Tukey's Honest Significant Difference tests after accounting for multiple comparisons. Full results from these tests are in Table 4.3.

Table 4.3. Multiple comparisons of three ecologically relevant phyla. Results are from Tukey's Honest Significant Difference tests, and P-values were adjusted for multiple comparisons. Differences are from averaged abundance. Stars indicate significant differences at the $\alpha = 0.05$ level, and these correspond to lettered groups in Fig. 4.6.

		Sinter			Crust		
	Comparison	Difference	P-value		Difference	P-value	
Cyabobacteria	Top - Mid	-213.5	< 0.001	*	16.0	0.912	
	Top - Deep	106.3	0.061		183.3	0.001	*
	Mid - Deep	319.8	< 0.001	*	167.3	0.002	*
Planctomycetes	Top - Mid	-16.3	0.004	*	0	–	
	Top - Deep	-79.7	< 0.001	*	-7.7	0.428	
	Mid - Deep	-63.3	< 0.001	*	-7.7	0.428	
Verrucomicrobia	Top - Mid	-23.2	< 0.001	*	1.7	0.951	
	Top - Deep	-11.7	0.049	*	1.5	0.947	
	Mid - Deep	11.5	0.027	*	-0.2	0.999	

Discussion

Microorganisms living in geothermal soils are faced with extreme environmental conditions that not only severely limit plant cover, but also exert strong controls over microbial communities. We were able to take advantage of this lack of plant cover in the present system which makes for an especially unique environment in which to study microbial communities across a phototrophic potential gradient. Both soils presented here exhibit a conspicuously-colored delineation of phototrophs within the top few centimeters of the soil surface. One reason that the colors are so visible in this system is that the soil parent material, essentially opaline silica and diatomaceous residuum, is white. Thus, an investigation of microbial β-diversity across a photic depth gradient is an obvious first step to understanding the drivers of heterogeneity

in this unique system. We found strong community partitioning as a factor of both soil type and depth, and this is consistent with findings of the eukaryotic component of the microbial communities (presented in Chapter 3). One primary physical difference between the sinter and crust soil types is particle size, and this largely dictates water relations at the soil surface. Sinter particles are generally >2mm, and are thus technically outside of the common definition of soil particles. That being said, the mixture of these particles and diatomaceous frustules creates a highly macro-porous, and well aerated, soil surface that increases in moisture content with depth. The crust soils, on the other hand are dense and fibrous up to the very surface, and hold more organic matter content, especially visible microbial exudates. This matrix holds more water at the surface, and also likely represents greater potential for anaerobicity throughout the epipedon.

We hypothesized that prokaryotic communities would vary with depth and soil type. The top samples receive full solar radiation and the mid-depth samples are phototrophically active, as evidenced by orange and green color bands at mid depth. Deep samples lack these colors, and were thus considered to be below the direct influence of solar radiation. Both unconstrained (NMDS; Fig. 4.4) and constrained ordinations (FSO; Fig. 4.5) reveal a pattern of samples clustered by replicates across both soil type and photic depth, and discriminant analysis results (Table 4.2) fully support this first hypothesis. Sinter surface particles (white triangles), however, stand out as a noticeable deviation. Additionally, surface samples in both soil types show a high degree of overlap with mid-depth samples in Fig. 4.5c. The major constituents at the phylum-level in the top sinter samples are in general agreement except for the

abundance of Fermicutes, Chloroflexi, and Acidobacteria (Fig. 4.2). The surface of the sinter soil is essentially composed of unconsolidated particles in an area often traversed by bison; multiple field sampling events were delayed to allow bison to travel on at least one well established game trail through the sinter area. Therefore it is possible that the disturbance regime of these samples, along with their surficial exposure, play a role in their differentiation relative to other replicated groups; it is also possible that this inconsistency reflects contamination during molecular processing, though all samples were accompanied by negative control reactions. The abundance of members of the Fermicutes phylum is notable in two of the three top sinter samples, since this group contains many anhydrobiotic endospore-producing bacteria, and these are the only samples that are likely to experience full desiccation for extended periods. The methods employed here would not have been able to differentiate between active Fermicute DNA and that from endospores, though all surface sinter samples used here were visibly desiccated.

Cyanobacteria, Planctomycetes, and Verrucomicrobia were selected for additional analysis since these are three potentially ecologically important phyla in this system (Fig. 4.6). Cyanobacteria are clearly visible, both as indicated by coloration of samples and through microscopy. Surface sinter particles are the most vulnerable to desiccation, so it is not surprising that Cyanobacteria were significantly more common in the mid-depth sinter samples compared to the surface samples. Crust soils, however, have greater water holding characteristics, and so phototrophs are more likely to proliferate at the surface, and were present at the same abundance level as at mid-depth crust samples. Planctomycetes and Verrucomicrobia, on the other hand,

were chosen for their important roles in anaerobic soil microbial processes. Members of the Planctomycetes are responsible for anarobic ammonia oxidation (anammox) in a variety of environments, and are common constituents of aquatic environments, and especially brackish waters (Madigan *et al.* 2008; Herlemann *et al.* 2011). The closely related phylum, Verrucomicrobia, has been more closely studied in soils in recent years since the relative importance and abundance of this group in soils has gone underreported given well-reported primer biases (Bergmann *et al.* 2011; Wagner & Horn 2006). We chose primers that effectively target both groups (Bergmann *et al.* 2011), and found differential abundances in the two soil types. Planctomycetes and Verrucomicrobia were both more abundant below the surface in sinter samples, but there was no difference in abundance with depth in crust samples (Fig. 4.6c–f). This is again indicative of the interrelated water relations and anaerobic potential in crust soils versus sinter soils, since the top crust samples are more likely to contain a consistently moist environment with a greater potential for the anaerobic microhabitats required by some members of both of these groups. These three groups lend support to our second hypothesis, that community differences can be traced to higher taxonomic groups, and thus it is clear that the differentiation of prokaryotic communities carries a strong phylogenetic signal.

In conclusion, we have observed clear differentiation of prokaryotic communities between two adjacent soil types, and along a photic depth gradient, consistent with soil physical differences and with visible coloration bands near the soil surface. We were able to single out three ecologically-important taxonomic groups with phylogenetic signals that trace to photic and aerobic differences, and thus we identified two

potential drivers of microbial community diversity. These geothermal diatomaceous soils harbor a surprising diversity of microorganisms and present a unique habitat that lends to microbial community investigations. Future studies of this system, and in other geothermal soils, might be able to identify soil-chemical drivers of some of the differences observed in these distinct communities.

Acknowledgements

We are grateful to the Yellowstone National Park permit personnel for help with permitting; to the Grand Teton Association for financial support through the Boyd Evison Graduate Fellowship to JM; to the Institute on Ecosystems for financial support through a Dissertation Fellowship to JM; and to D. Skorupa, S. Clingenpeel, and T. R. McDermott for help with primer design, barcoding, and amplification.

CHAPTER 5.

CONCLUSIONS

We know surprisingly little of the environmental factors that shape soil microbial communities; this is partly because of the staggering heterogeneity of the soil environment, but it is also because methods for studying these organisms are still continuing to be developed. Soil microbiology often revolves around the regression of soil chemical conditions with occurrence or abundance patterns of certain taxa or groups of microbes that are hypothesized to respond to such changes. This has been a very productive approach to understanding the relationship between the soil environment and its various constituents. The three major chapters of my dissertation (2, 3 & 4) were, in some ways, attempts at addressing this broad question using unconventional hypotheses and theoretical approaches in an unconventional environmental setting.

The question addressed in Chapter 2, what is the relative strength of influence on mycorrhizal fungal communities from either host plants or the soil environment, is an exceedingly difficult question to address *in situ*, and this is the major reason why there is still not a complete answer. This study, to my knowledge, has come the closest to being able to assign an answer to this question, and it had a lot to do with the selection of a model system in which contrasting plant community types were located in very close proximity, and contained a single common host plant, *M. guttatus*. The strength of this study, however, proved to be its major flaw; adjacent contrasting plant community types were growing in soils that had very different pH. While designing

this study, I did anticipate this being the case, but found in preliminary sampling that the differences in soil conditions between contrasting plant communities might not be as great as expected. By chance, it turns out, my preliminary sampling sites were the exception and not the rule, and this mistake resulted in plant communities confounded with soil conditions. I was able to use analytical finesse to deal with this problem, but not in a robust way. This study also perhaps loses some scope of inference given that it was designed around only two community types: those with only annual plants, and those with a mixture of annual and perennial plants for AMF hosts. This approach did achieve a very stark result in terms of fungal community, but a more informative study would have accounted for a gradient of plant community types and richness values. Weaknesses aside, I did manage to show that AMF communities were more closely associated with differences in host community than with soil conditions, and this has not previously been done in AMF literature. Perhaps the most interesting result from this study was that adjacent plant communities harbored distinct AMF communities with surprisingly little overlap; only three of the 28 AMF taxa observed were found in both community types. This interesting finding sets up a potential to address a similar hypothesis in environments where plants are similarly stressed, and thus present similarly contrasting community life-history types. This study, to me, also strongly indicates that some AMF taxa appear to exhibit strengths in some ecological settings over others, meaning that each has its own dispersal limitations and/or ability to survive in harsh soil conditions for nearly a year without a host plant, its own ability to compete for establishment in roots, and its own context with respect to the types of symbiotic pairing. These concepts have received little

attention in AMF ecology, and my study was not designed to answer the questions they pose, but it does present a step in the direction toward addressing this gap in our understanding.

In terms of methodology, the study presented in Chapter 2 was greatly informed by the use of ribosomal DNA sequence identification, since I was able to assign each AMF OTU with an approximate taxonomic distinction, and that was corroborated by existing LSU (28S) sequences from known isolates. While the goal of the study was not to isolate novel AMF species, it appears that at least some of the OTUs used in analysis were likely novel, since the closest Glomeromycotan sequence in GenBank was quite distant in some cases (<95%). Recent molecular work seems to corroborate this, since AMF taxa appear to exist on the continuum from metropolitan to localized and endemic; much more research is still to be done to more adequately describe *Glomeromycota*. The molecular approach that I took to characterizing AMF communities in plants was slow and laborious, but productive.

Chapters 3 & 4 were the result of barcoded-pyrosequencing of samples from across a microbially-active geothermal soil crust. I happened across this system while sampling nearby plants. The soil is visually striking, and had an appearance similar to arid biological soil crusts. The clear difference is that the geothermal soil stays moist because of subsurface water throughout the year and because of the tremendous water-holding capacity of the highly porous, dense, fibrous soil matrix. The visible coloration banding within the top few centimeters of the soil surface begs explanation, and seemed to be an obvious starting point to understanding the system. I collected samples with the intention of defining microbial communities in two soil types and

across a photic gradient evident in both soil types. This was initially done as a descriptive study, though the hypothesis-driven components revolved around corroborating visible microbial community differences with sequence data. While this produced a very interesting first look at an unstudied system, I have definitely learned, through manuscript peer-review, that this is an unconventional, and at times unwelcome, approach to study design in microbiology. My goal was to assess community structure in an interesting system, and to contrast the visible with the microscopic - it is rare that we are presented with such a visible system in which to ask this type of question. In-depth chemical analysis of particular groups seemed to me to be secondary, though important in future research. This certainly helps me perceive the oft bemoaned disconnect between microbiology and ecology, but I also understand much better how to approach a similarly descriptive situation in the future.

It is ironic that I spent the better part of two years generating, for research described in Chapter 2, a final dataset populated by fewer than 300 DNA sequences, and I spent a few weeks generating nearly half a million DNA sequences, of similar length, for the projects presented in Chapters 3 & 4, and for comparable price. This is a perfect illustration of the current trajectory and acceleration of DNA sequencing as of this writing; more than 4 million nucleotide bases can be sequenced for each dollar spent. Indeed, Moore's law of technology is playing out in impressive form in the world of molecular biology. This acceleration of sequencing technology and its accompanied generation of unprecedented volumes of data are also necessitating advances in software and methodology capable of processing such data. Sequence processing, bioinformatics, statistical programming, and ecological statistical analysis

applicable to metagenomic data are among the most valuable skills developed during this process.

One original objective when starting graduate school was to address questions in basic soil ecology that have application in ecological restoration. For instance, we have major gaps in our understanding of mycorrhizal interactions in harsh environments, and mycorrhizal interactions have become a desirable component of restoration projects, which are generally taking place in harsh environments. The first component of my research does imply that there is a functional link between diversity of plants, and their community structure, with that of an associated mycorrhizal community. Inocula in the restoration industry are often composed of single species cultures, and it is tempting to judge these on the basis of 'not complex enough to benefit.' Chapter 2 of this dissertation did not attempt to assign, and was not designed to measure, any sort of effect size in benefit for isolated plants, but it was apparent that single, isolated plants in a naturally harsh condition were able to establish and complete their life cycles with an average of slightly more than a single AMF taxon in their roots. This could imply that there might not be a problem with applying a single species mycorrhizal inoculum to the roots of an isolated plant being placed in, for instance, arid reclaimed strip mine overburden without the benefits of microbially active topsoil. There certainly remain unanswered questions regarding local adaptation to environmental conditions, as well as relational specificity, and these are potential areas of future research in this field. The other major components of my dissertation (Chapters 3 & 4) have no clear application in this regard, but these projects do help to advance bioinformatical and statistical methods in metagenomics

and high-throughput microbial ecology, and these methods will increasingly be used in restoration ecology.

While attending a recent soil metagenomics conference, the constant theme, both in presentations and in discussions, was the rapid change happening in the field of soil microbiology due to high-throughput molecular technology. Against this backdrop, an elder, well recognized soil microbiologist asked the audience to predict the title of a presentation they might give at the same meeting in 10 years. Answers were revealed later in the day; some were thoughtful, some were humorous, and others were forward-thinking in terms of what might be accomplished in just a few years. The title that stood out, however, was this: "We still know nothing about soil microbial communities." This title perfectly illustrates the conundrum presented by evolving technology in science. All the data in the world cannot answer a poorly framed question, and simple techniques might be the most appropriate way to answer even the most complex question. This, perhaps more than any other lesson, is what I take away from having concluded this dissertation research. With so many choices for producing mountains of microbial community data, it is easy, certainly for me, to overlook the possibility that a relatively small portion will adequately tell the story.

The work presented herein addressed different sets of ecological questions using data produced through very different methods, each with its own strengths and weaknesses, and each was able to answer the relevant questions. The questions addressed in this dissertation, however, are all somewhat preliminary in their scope of inference, and they clearly demonstrate the value of geothermal soils as model systems for ecological research. Future work into the inherent link in symbiotic community structures

can benefit from a focus on 1) additional plant species to investigate specificity issues not addressed in Chapter 2; 2) finer gradation between the isolated and communal settings investigated here, to separate influences of host life history strategy and host species richness; and 3) a study design incorporating more overlap in soil-chemical conditions as a way to further disentangle relative influences from soil conditions and hosts. The biological soil crust study presented in Chapters 3 & 4 was the first look at an incredibly interesting system; many questions remain to be asked of such a system, and this dissertation approached a few of the most obvious. Chemical conditions affecting microbial communities in these soils are of primary interest – something my study was not designed to address. Additionally, geothermal water inundation of these soils likely has a strong pulse-dynamic effect on microbial communities and resulting soil chemistry. Both geothermal soil areas in this dissertation can, in the future, act as useful model soil systems for further addressing ecological questions.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Amaral-Zettler, L. A., McCliment, E. A., .Ducklow, H. W, & Huse, S. M. 2009. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLoS ONE*, **4**, e6372.

Appoloni, S., Lekberg, Y., Tercek, M. T., Zabinski, C. A., & Redecker, D. 2008. Molecular community analysis of arbuscular mycorrhizal fungi in roots of geothermal soils in Yellowstone National Park (USA). *Microbial Ecology*, **56**, 649–659.

Bates, S. T., Gundlapally, S. R., & Garcia-Pichel, F. 2006. *Exophiala crusticola* anom. nov. (affinity *Herpotrichiellaceae*), a novel black yeast from biological soil crusts in the Western United States. *International Journal of Systematic and Evolutionary Microbiology*, **56**, 2697–2702.

Bates, S. T., Cropsey, G. W. G., Caporaso, J. G., Knight, R., & Fierer, N. 2011a. Bacterial communities associated with the lichen symbiosis. *Applied Environmental Microbiology*, **77**, 1309–1314.

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., & Fierer, N. 2011b. Examining the global distribution of dominant archaeal populations in soil. *ISME Journal*, **5**, 908–917.

Bates, S. T., Berg-Lyons, Donna, Lauber, C. L., Walters, W. A., Knight, R., & Fierer, N. 2012. A preliminary survey of lichen associated eukaryotes using pyrosequencing. *Lichenologist*, **44**, 137–146.

Becraft, E. D., Cohan, F. M., Kühl, M., Jensen, S. I., & Ward, D. M. 2011. Fine-scale distribution patterns of *Senechococcus* ecological diversity in microbial mats of Mushroom Spring, Yellowstone National Park. *Applied Environmental Microbiology*, **77**, 7689–7697.

Begon, Michael, Townsend, Colin R., & Harper, John L. 2006. *Ecology: From Individuals to Ecosystems*. Vol. Fourth. Oxford: Blackwell.

Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R. R., & Stoeck, T. 2011. Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypergariable SSU rRNA gene regions. *Environmental Microbiology*, **13**, 340–349.

Belnap, J. 2002. Nitrogen fixation in biological soil crusts from southeast Utah, USA. *Biology and Fertility of Soils*, **35**, 128–135.

Belnap, J. 2003. The world at your feet: desert biological soil crusts. *Frontiers in Ecology and Environment*, **1**, 181–189.

Belnap, J., & Lange, O. L. 2003. *Biological Soil Crusts: Structure, Function, and Management*. New York: Springer.

Berberán, A., Bates, S. T., Casamayor, E. O., & Fierer, N. 2011. Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME Journal*, **6**, 343–351.

Bergmann, G. T., Bates, S. T., Eilers, K. G., Lauber, C. L., Caporaso, J. G., Walters, W. A., Knight, R., & Fierer, N. 2011. The under-recongnized dominance of *Verrucomicrobia* in soil bacterial communities. *Soil Biology & Biochemistry*, **43**, 1450–1455.

Bever, J. D. 2002. Host-specificity of AM fungal population growth rates can generate feedback on plant growth. *Plant and Soil*, **244**, 281–290.

Bever, J. D., & Wang, M. 2005. Arbuscular mycorrhizal fungi: Hyphal fusion and multigenomic structure. *Nature*, **433**, E3–E4.

Bhatnagar, A., Makandar, M. B., Garg, M. K., & Bhatnagar, M. 2008. Community structure and diversity of cyanobacteria and green algae in the soils of Thar Desert (India). *Journal of Arid Environments*, **72**, 73–83.

Börstler, B., Renker, C., Kahmen, A., & Buscot, F. 2006. Species composition of arbuscular mycorrhizal fungi in two mountain meadows with differing management types and levels of plant biodiversity. *Biology and Fertility of Soils*, **42**, 286–298.

Botero, L. M., D'Imperio, S., Burr, M., McDermott, T. R., Young, M., & Hassett, D. J. 2005. Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Applied Environmental Microbiology*, **71**, 1267–1275.

Bowker, M. A., & Belnap, J. 2008. A simple classification of soil types as habitats of biological soil crusts on the Colorado Plateau, USA. *Journal of Vegetation Science*, **19**, 831–840.

Bowker, M.A., Maestre, F.T., & Escolar, C. 2009. Biological crusts as model system for examining the biodiversity-ecosystem function relationship in soils. *Soil Biology & Biochemistry*, **42**, 405–417.

Brate, Jon, Logares, Ramiro, Berney, Cedric, Ree, Dan Kristofer, Klaveness, Dag, Jakobsen, Kjetill S, & Shalchian-Tabrizi, Kamran. 2010. Freshwater Perkinsea and marine-freshwater colonizations revealed by pyrosequencing and phylogeny of environmental rDNA. *ISME Journal*, **4**, 1144–1153.

Büdel, B. 2005. *Microorganisms in soils: roles in genesis and functions*. Soil Biology, vol. 3. Springer. Chap. 15.

Bunn, R., Lekberg, Y., & Zabinski, C. 2009. Arbuscular mycorrhizal fungi ameliorate temperature stress in thermophilic plants. *Ecology*, **90**, 1378–1388.

Bunn, R. A., & Zabinski, C. A. 2003. Arbuscular mycorrhizae in thermal-influenced soils in yellowstone national park. *Western North American Naturalist*, **63**, 409–415.

Burr, M. D., Botero, L. M., Young, M. J., Inskeep, W. P., & McDermott, T. R. 2005. Observations concerning nitrogen cycling in a Yellowstone thermal soil environment. *Pages 171–182 of:* Inskeep, W. P., & McDermott, T. R. (eds), *Geothermal biology and geothemistry in Yellowstone National Park*. Bozeman, MT: Thermal Biology Institute.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., & Knight, R. 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266:267.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Tumbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., & Knight, R. 2010b. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Cardon, Z. G., Gray, D. W., & Lewis, L. A. 2008. The green algal underground: evolutionary secrets of desert cells. *BioScience*, **58**, 114–122.

Carlson, R. 2003. The pace and proliferation of biological technologies. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, **1**, 203–214.

Channing, A. 2001. *Processes and environments of vascular plant silicification*. Ph.D. thesis, Cardiff University.

Channing, A., & Edwards, D. 2004. Experimental taphonomy: silicification of plants in Yellowstone hot-spring environments. *Transactions of the Royal Society of Edinburgh-Earth*, **94**, 503–521.

Channing, A., Edwards, D., & Sturtevant, S. 2004. A geothermally influenced wetland containing unconsolidated geochemical sediments. *Canadian Journal of Earth Science*, **41**, 809–827.

Chariton, A. A., Court, L. N., Hartley, D. M., Colloff, M. J., & Hardy, C. M. 2010. Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, **8**, 233–238.

Chase, Jonathan M., Kraft, Nathan J. B., Smith, Kevin G., Vellend, Mark, & Inouye, Brian D. 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere*, **2:art24**.

Chaudhary, V. B., Bowker, M. A., O'Dell, T. E., Grace, J. B., Redman, A. E., Rillig, M. C., & Johnson, N. C. 2009. Untangling the biological contributions to soil stability in semiarid shrublands. *Ecological Applications*, **19**, 110–122.

Chu, H., Fierer, N., Lauber, C. L., Caporaso, J. G., Knight, R., & Grogan, P. 2010. Soil bacterial diversity in the arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology*, **12**, 2998–3006.

Ciniglia, C., Yoon, H. S., Pollio, A., Pinto, G., & Bhattacharya, D. 2004. Hidden biodiversity of the extremophilic Cyanidales red algae. *Molecular Ecology*, **13**, 1827–1838.

Collins, C. D., & Foster, B. L. 2009. Community-level consequences of mycorrhizae depend on phosphorus availability. *Ecology*, **90**, 2567–2576.

Davidson, J., Öpik, M., Daniell, T. J., Moora, M., & Zobel, M. 2011. Arbuscular mycorrhizal fungal communities in plant roots are not random assemblages. *FEMS Microbiology Ecology*, **78**, 103–115.

de Bruijn, F. J. (ed). 2011. *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats*. Hoboken: Wiley-Blackwell.

Dormann, C. F., Gruber, B., & Frund, J. 2008. Introducing the bipartite Package: Analysing Ecological Networks. *R news*, **8/2**.

Dorn, Robert D. 2001. *Vascular Plants of Wyoming*. Cheyenne, WY: Mountain West Publishing.

Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C., & Fitter, A. H. 2010. Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME Journal*, **4**, 337–345.

Dumbrell, A. J., Ashton, P. D., Aziz, N., Feng, G., Nelson, M., Dytham, C., Fitter, A. H., & Helgason, T. 2011. Distinct seasonal assemblages of arbuscular mycorrhizal fungi revealed by massively parallel pyrosequencing. *New Phytologist*, **190**, 794–804.

Edgcomb, Virginia, Orsi, William, Bunge, John, Jeon, Sunok, Christen, Richard, Leslin, Chesley, Holder, Mark, Taylor, Gordon T, Suarez, Paula, Varela, Ramon, & Epstein, Slava. 2011. Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME Journal*, **5**, 1344–1356.

Egge, J. K., & Aksnes, D. L. 1992. Silicate as regulating nutrient in phytoplankton competition. *Marine Ecology Progress Series*, **83**, 281–289.

Ehrenfeld, J. G., Ravit, B., & Elgersma, K. 2005. Feedback in the plant-soil system. *Annual Review of Environment and Resources*, **30**, 75–115.

Evans, R. D., & Lange, O. L. 2003. *Biological soil crusts and ecosystem nitrogen and carbon dynamics*. Springer. Chap. 20.

Fierer, N., & Lennon, J. 2011. The generation and maintenance of diversity in microbial communities. *American Journal of Botony*, **98**, 439–448.

Fierer, N., Breitbart, M., Nulton, J., Salomon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R. A., Felts, B., Rayhawk, S., Knight, R., Rohwer, F., & Jackson, R. B. 2007a. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Applied Environmental Microbiology*, **73**, 7059–7066.

Fierer, N., Bradford, M. A., & Jackson, R. B. 2007b. Toward an ecological classification of soil bacteria. *Ecology*, **88**, 1354–1364.

Fierer, N., Grandy, A. S., Six, J., & Paul, E. A. 2009a. Searching for unifying principles in soil ecology. *Soil Biology & Biochemistry*, **41**, 2249–2256.

Fierer, N., McCain, C. M., Meir, P., Zimmermann, M., Rapp, J. M., Silman, M. R., & Knight, R. 2011. Microbies do not follow the elevational diversity patterns of plants and animals. *Ecology*, **92**, 797–804.

Fierer, Noah, & Jackson, Robert B. 2006. The diversity and biogeography of soil bacterial communities. *PNAS*, **103**, 626–631.

Fierer, Noah, Strickland, Michael S., Liptzin, Daniel, Bradford, Mark A., & Cleveland, Cory C. 2009b. Global patterns in belowground communities. *Ecology Letters*, **12**, 1238–1249.

Fitzsimons, M. S., & Miller, R. M. 2010. Serpentine soil has little influence on the root-associated microbial community composition of the serpentine tolerant grass species *Avenula sulcata*. *Plant and Soil*, **330**, 393–405.

Fulthorpe, R. R., Roesch, L. F., Riva, A., & Triplett, E. W. 2008. Distantly sampled soils carry few species in common. *ISME Journal*, **2**, 901–910.

Garcia-Pichel, F., López-Córtes, A., & Nübel, U. 2001. Phylogenetic and morphological diversity of Cyanobaceteria in soil desert crusts from the Colorado Plateau. *Applied Environmental Microbiology*, **67**, 1902–1910.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shaw, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J., & Nekrutenko, A. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, **15**, 1451–1455.

Gilbert, J. A., Meyer, F., & Bailey, M. J. 2011. The future of microbial metagenomics (or is ignorance bliss?). *ISME Journal*, **5**, 777–779.

Gollotte, A., van Tuinen, D., & Atkinson, D. 2004. Diversity of arbuscular mycorrhizal fungi colonizing roots of the grass species *Agrostis capillaries* and *Lolium Perenne* in a field experiment. *Mycorrhiza*, **14**, 111–117.

Guidon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321.

Gundlapally, S. R., & Garcia-Pichel, F. 2005. *Dyadobacter crusticola* sp. nov., from biological soil crusts in the Colorado Plateau, USA, and an emended description of the genus *Dyadobacter* Chelius and Triplett 2000. *International Journal of Systematic and Evolutionary Microbiology*, **55**, 1295–1299.

Gundlapally, S. R., & Garcia-Pichel, F. 2006. The community and phylogenetic diversity of biological soil crusts in the Colorado Plateau studied by molecular fingerprinting and intensive cultivation. *Microbial Ecology*, **52**, 345–357.

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., & Knight, R. 2008. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods*, **5**, 235–237.

Harper, K. T., & Belnap, J. 2001. The influence of biological soil crusts on mineral uptake by associated vascular plants. *Journal of Arid Environments*, **47**, 347–357.

Hausmann, N. T., & Hawkes, C. V. 2009. Plant neighborhood control of arbuscular mycorrhizal community composition. *New Phytologist*, **183**, 1188–1200.

Hausmann, Natasha Teutsch, & Hawkes, Christine V. 2010. Order of plant host establishment alters the composition of arbuscular mycorrhizal communities. *Ecology*, **91**, 2333–2343.

Hawkes, C. V. 2003. Nitrogen cycling mediated by biological soil crusts and arbuscular mycorrhizal fungi. *Ecology*, **84**, 1553–1562.

Helgason, T., Merryweather, J. W., Young, J. P. W., & Fitter, A. H. 2007. Specificity and resilience in the arbuscular mycorrhizal fungi of a natural woodland community. *Journal of Ecology*, **95**, 623–630.

Hendershot, W. H., Lalande, H., & Duquette, M. 2008. Soil Reaction and Exchangeable Acidity. *In:* Carter, M. R., & Gregorich, E. G. (eds), *Soil Sampling and Methods of Analysis*, vol. 2nd. Boca Raton: CRC Press.

Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S, Waniek, J. J., & Andersson, A. F. 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME Journal*, **5**, 1571–1579.

Hobbs, W. O., Wolfe, A. P., Inskeep, W. P., Amskold, L., & Konhauser, K. O. 2009. Epipelic diatoms from an extreme acid environment: Beowulf Spring, Yellowstone National Park, USA. *Nova Hedwigia*, **135**, 71–83.

Hollister, E. B., Engledow, A. S., Hammett, A. J. M., Provin, T. L., Wilkinson, H. H., & Gentry, T. J. 2010. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. *ISME Journal*, **4**, 829–838.

Horner-Devine, M. C., & Bohannan, B. J. M. 2006. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology*, **87**, 100–108.

Housman, D. C., Powers, H. H., Collins, A. D., & Belnap, J. 2006. Carbon and nitrogen fixation differ between successional stages of biological soil crusts in the Colorado Plateau and Chihuahuan Desert. *Journal of Arid Environments*, **66**, 620–634.

Housman, D. C., Yeager, C. M., Darby, B. J., Sanford, R. L., Kuske, C. R., Neher, D. A., & Belnap, J. 2007. Heterogeneity of soil nutrients and subsurface biota in a dryland ecosystem. *Soil Biology & Biochemistry*, **39**, 2138–2149.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, **54**, 187–182.

Hutchinson, C. A. 2007. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Research*, **35**, 6227–6237.

Johnson, D., Vandenkoornhuyse, P. J., Leake, J. R., Gilbert, L., Booth, R. E., Grime, J. P., Young, J. P. W., & Read, D. J. 2003. Plant communities affect arbuscular mycorrhizal fungal diversity and community composition in grassland microcosms. *New Phytologist*, **161**, 503–515.

Johnson, N. C., Hoeksema, J. D., Bever, J. D., Chaudhary, V. B., Gehring, C., Klironomos, J., Koide, R., Miller, R. M., Moore, J., Moutoglis, P., Schwartz, M., Simard, S., Swenson, W., Umbanhowar, J., Wilson, G., & Zabinski, C. A. 2006.

From Lilliput to Brobdingnag: Extending models of mycorrhizal function across scales. *BioScience*, **56**, 889–900.

Kaufman, L., & Rousseeuw, P. J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, **26**, 1463–1464.

Kivlin, S. N., Hawkes, C. V., & Treseder, K. K. 2011. Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil Biology & Biochemistry*, **43**, 2294–2303.

Klironomos, J. N. 2003. Variation in plant response to native and exotic arubuscular mycorrhizal fungi. *Ecology*, **84**, 2292–2301.

Klironomos, J. N., Hart, M. M., Gurney, J. E., & Moutoglis, P. 2001. Interspecific differences in the tolerance of arbuscular mycorrhizal fungi to freezing and drying. *Canadian Journal of Botany*, **79**, 1161–1166.

Koske, R. E., & Gemma, J. N. 1989. A modified procedure for staining roots to detect VA-mycorrhizas. *Mycology Resources*, **92**, 486–505.

Krüger, M., Krüger, C., Walker, C., Stockinger, H., & Schüßler, A. 2011. Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhyzal fungi from phylum to species level. *New Phytologist*, **doi: 10.1111/j.1469-8137.2011.03962.x**.

Kuczynski, J., Liu, Z. Z., Lozupone, C., McDonald, D., Fierer, N., & Knight, R. 2010. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods*, **7**, 813–821.

Kyle, J. E., & Schroeder, P. A. 2007. Role of smectite in siliceous-sinter formation and microbial-texture presercation: Octopus Spring, Yellowstone National Park, Wyoming, USA. *Clay and Clay Minerals*, **55**, 189–199.

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. 2009. Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. *Applied Environmental Microbiology*, **75**, 5111–5120.

Legendre, P, & Legendre, L. 1998. *Numerical Ecology, Volume 20, Second Edition (Developments in Environmental Modelling)*. San Fransisco: Elsevier.

Lekberg, Y., Koide, R. T., Rohr, J. R., Aldrich-Wolfe, L., & Morton, J. B. 2007. Role of niche restrictions and dispersal in the composition of arbuscular mycorrhizal fungal communities. *Journal of Ecology*, **95**, 95–105.

Lekberg, Y., Meadow, J., Rohr, J. R., Redecker, Dirk, & Zabinski, C. 2011. Importance of dispersal and thermal environment for mycorrhizal communities: Lessons from Yellowstone National Park. *Ecology*, **92**, 1292–1302.

Lemos, L. N., Fulthorpe, R. R., Triplett, E. W., & Roesch, L. F. W. 2011. Rethinking microbial diversity analysis in the high throughput sequencing era. *Journal of Microbiological Methods*, **86**, 42–51.

Lennon, J. T. 2011. Replication, lies and lesser-known truths regarding experimental design in environmental microbiology. *Environmental Microbiology*, **13**, 1383–1386.

Lim, Y. W., Kim, B. K., Kim, C., Jung, H. S., Kim, B. S., Lee, J. H., & Chun, J. 2010. Assessment of soil fungal communities using pyrosequencing. *Microbiology*, **48**, 284–289.

Liu, Z., Lozupone, C. A., Hamady, M., Bushman, F. D., & Knight, R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*, **35**, e120.

Lozupone, C. A., Hamady, M., & Knight, R. 2006. UniFrac – an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics*, **7**, 371.

Madigan, M. T., Martinko, J. M., Dunlap, P. V., & Clark, D. V. 2008. *Brock Biology of Microorganisms*. 12th edn. New York, USA: Pearson Higher Education.

Maechler, Martin, Rousseeuw, Peter, Struyf, Anja, & Hubert, Mia. 2005. *Cluster Analysis Basics and Extensions*. R package version 1.14.0.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C. R., Morin, P. J., Naeem, S., Øvreås, L., Reysenbach, A. L., Smith, V. H., & Staley, J. T. 2006. Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, **4**, 102–112.

Medinger, R., Nolte, V., Pandey, R. V., Jost, S., Ottenwälder, B., Schlötterer, C., & Boenigk, J. 2010. Diversity in a hidden world: potential and limiation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology*, **19**, 32–40.

Meyer, M., Stenzel, U., & Hofreiter, M. 2008. Parallel tagged sequencing on the 454 platform. *Nature Protocols*, **3**, 267–278.

Miller, S. R., Strong, A. L., Jones, K. L., & Ungerer, M. C. 2009. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Applied Environmental Microbiology*, **75**, 4565–4572.

Miller, W. P., & Miller, D. M. 1987. A micro-pipette method for soil mechanical analysis. *Communications in Soil Science and Plant Analysis*, **19**, 1–15.

Miransari, M. 2011. Interactions between arbuscular mycorrhizal fungi and soil bacteria. *Applied Microbiology and Biotechnology*, **89**, 917–930.

Morgan, L. A. (ed). 2007. *Integrated geoscience studies in the Greater Yellowstone Area—Volcanic, tectonic, and hydrothermal processes in the Yellowstone geoecosystem.* Vol. 1717. Reston, VA: USGS.

Mummey, D. L., & Rillig, M. C. 2007. Evaluation of LSU rRNA-gene PCR primers for analysis of arbuscular mycorrhizal fungal communities via terminal restriction fragment length polymorphism analysis. *Journal fo Microbiological Methods*, **70**, 200–204.

Mummey, D. L., Rillig, M. C., & Holben, W. E. 2005. Neighboring plant influences on arbusclar mycorrhizal fungal community composition as assessed by T-RFLP analysis. *Plant and Soil*, **271**, 83–90.

Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C. A., Jiang, L., Schmidt, S. K., Fierer, N., Townsend, A. R., Cleveland, C. C., Stanish, L., & Knight, R. 2010. Global patterns in biogeography of bacterial taxa. *Environmental Microbiology*, **13**, 135–144.

Newsham, K. K., Fitter, A. H., & Watkinson, A. K. 1995. Arbuscular mycorrhiza protect an annual grass from root pathogenic fungi in the field. *Journal of Ecology*, **83**, 991–1000.

Norris, T. B., & Castenholz, R. W. 2006. Endolithic photosynthetic communities within ancient and recent travertine deposits in Yellowstone National Park. *FEMS Microbiology Ecology*, **57**, 470–483.

Norris, T. B., Wraith, J. M., Castenholz, R. W., & McDermott, T. R. 2002. Soil microbial community structure across a thermal gradient following a geothermal heating event. *Applied Environmental Microbiology*, **68**, 6300–6309.

Oksanen, Jari, Blanchet, F. Guillaume, Kindt, Roeland, Legendre, Pierre, O'Hara, R. B., Simpson, Gavin L., Solymos, Peter, Stevens, M. Henry H., & Wagner, Helene. 2011. *vegan: Community Ecology Package.* R package version 1.17-6.

Öpik, M., Moora, M., Liira, J., & Zobel, M. 2006. Composition of root-colonizing arbuscular mycorrhizal fungal communities in different ecosystems around the globe. *Journal of Ecology*, **94**, 778–790.

Öpik, M., Metsis, M., Daniell, T. J., Zobel, M., & Moora, M. 2009. Large-scale parallel 454 sequencing reveals host ecological group specificity of arubuscular mycorrhizal fungi in a boreonemoral forest. *New Phytologist*, **184**, 424–437.

Öpik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davidson, J., Kalwij, J. M., Reier, Ü., & Zobel, M. 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytologist*, **188**, 223–241.

Owen, R. B., Renaut, R. W., & Jones, B. 2008. Geothermal diatoms: a comparative study of floras in hot springs systems of Iceland, New Zealand, and Kenya. *Hydrobiologia*, **610**, 175–192.

Paradis, E., Claude, J., & Strimmer, K. 2004. APE: analysis of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Parniske, M. 2008. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature Reviews Microbiology*, **6**, 763–775.

Paul, E. A. 2007. *Soil Microbiology, Ecology and Biochemistry*. 3rd edn. San Fransisco: Academic Press.

Pawlowska, T. E., & Taylor, J. W. 2004. Organization of genetic variation in individuals of arbuscular mycorrhizal fungi. *Nature*, **427**, 733–737.

Philippot, L., Andersson, S. V. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., & Hallin, S. 2011. The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology*, **8**, 523–529.

Price, M. N., Dehal, P. S., & Arkin, A. P. 2009. FastTree: Computing large minimum-evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution*, **26**, 1641–1650.

Prosser, J. I. 2010. Replicate or lie. *Environmental Microbiology*, **12**, 1806–1810.

Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckelton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, A. M., Solan, M., van der Gast, C. J., & Young, J. P. W. 2007. The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, **5**, 384–392.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., & Gloöckner, F. O. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188–7196.

R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramette, A., Tiedje, J. M., & Boetius, A. 2009. Impact of space, time and complex environments on microbial communities. *Clinical Microbiology & Infection*, **15**, 60–62.

Ramirez, K. S., Lauber, C. L., Knight, R., Bradford, M. A., & Fierer, N. 2010. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology*, **91**, 3463–3470.

Raup, David M., & Crick, Rex E. 1979. Measurement of Faunal Similarity in Paleontology. *Journal of Paleontology*, **53**, 1213–1227.

Redecker, Dirk. 2006. Molecular ecology of arbuscular mycorrhizal fungi: a review of PCR-based techniques. *In:* Cooper, J E, & Rao, J R (eds), *Molecular Approaches to Soil, Rhizosphere and Plant Microorganism Analysis*. CABI.

Redman, R. S., Litvintseva, A., Sheehan, K. B., Henson, J. M., & Rodriguez, R. J. 1999. Fungi from geothermal soils in Yellowstone National Park. *Applied Environmental Microbiology*, **65**, 5193–5197.

Roberts, D. W. 1986. Ordination on the basis of fuzzy set theory. *Vegetatio*, **66**, 123–131.

Roberts, D. W. 2008. Statistical analysis of multidimensional fuzzy set ordinations. *Ecology*, **89**, 1246–1260.

Roberts, David W. 2010a. *fso: Fuzzy Set Ordination*. R package version 1.2-0.

Roberts, David W. 2010b. *labdsv: Ordination and Multivariate Analysis for Ecology*. R package version 1.4-1.

Roberts, David W. 2010c. *optpart: Optimal partitioning of similarity relations*. R package version 2.0-1.

Rodman, A, Shovic, H F, & Thoma, D. 1996. *Soils of Yellowstone National Park*. Tech. rept. Yellowstone National Park.

Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M., Kent, A. D., Daroub, S. H., Camargo, F. A. O., Farmerie, W. G., & Triplett, E. W. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME Journal*, **1**, 283–290.

Rosendahl, S. 2008. Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist*, **178**, 253–266.

Rosendahl, Søren, McGee, Peter, & Morton, Joseph B. 2009. Lack of global population genetic differentiation in the mycorrhizal fungus Glomus mosseae suggests a recent range expansion which may have coincided with the spread of agriculture. *Molecular Ecology*, **18**, 4316–4329.

Rosentreter, R., Bowker, M., & Belnap, J. 2007. *A Field Guide to Biological Soil Crusts of Western U.S. Drylands*. Denver: U. S. Government Printing Office.

Rousk, J., Brookes, P. C., & Bååth, E. 2010a. Investigating the mechanisms for the opposing pH-relationships of fungal and bacterial growth in soil. *Soil Biology & Biochemistry*, **42**, 926–934.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R., & Fierer, N. 2010b. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME Journal*, **4**, 1340–1351.

Rowe, J. J., Fournier, R. O., & Morey, G. W. 1973. *Chemical analysis of thermal waters in Yellowstone National Park, Wyoming, 1960-65*. Geological Survey Bulletin 1303. Department of the Interior.

Ruiz-Luzano, J. M. 2003. Arbuscular mycorrhizal symbiosis and alleviation of osmotic stress. New perspectives for molecular studies. *Mycorrhiza*, **13**, 309–317.

Sanders, I. R. 2003. Preference, specificity and cheating in the arubuscular mycorrhizal symbiosis. *TRENDS in Plant Science*, **8**, 143–145.

Sanger, F., Nicklen, S., & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *PNAS*, **74**, 5463–5467.

Schechter, S. P., & Bruns, T. D. 2008. Serpentine and non-serpentine ecotypes of *Collinsia sparsiflora* associate with distinct arbuscular mycorrhizal fungal assemblages. *Molecular Ecology*, **17**, 3198–3210.

Schlesinger, W. H., Pippen, J. S., Wallenstein, M. D., Hofmockel, K. S., Klepeis, D. M., & Mahall, B. E. 2003. Community composition and photosynthesis by photoautotrophs under quartz pebbles, southern Mojave Desert. *Ecology*, **84**, 3222–3231.

Schloss, P. 2010. The effect of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*, **6**.

Schroeder, M. S., & Janos, D. P. 2004. Phosphorus and intraspecific density alter plant responses to arbuscular mycorrhizas. *Plant and Soil*, **264**, 335–348.

Schüßler, A., & Walker, C. 2010. The *Glomeromycota*: a species list with new families and new genera. Royal Botanic Garden, Edinburgh.

Schüßler, A., Schwarzott, D., & Walker, C. 2001. A new fungal phylum, the *Glomeromycota*: phylogeny and evolution. *Mycological Research*, **105**.

Schwartzman, D. W., & Volk, T. 1989. Biotic enhancement of weathering and the habitability of Earth. *Nature*, **340**, 457–460.

Seckbach, J., & Kocielek, J. P. (eds). 2011. *The Diatom World*. New York, USA: Springer.

Smith, Sally E., & Read, David J. 2008. *Mycorrhizal Symbiosis*. Vol. 3. London, UK: Academic Press.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., & Herndl, G. J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *PNAS*, **103**, 12115–12120.

Soule, T., Anderson, I. J., Johnson, S. L., Bates, S. T., & Garcia-Pichel, F. 2009. Archaeal populations in biological soil crusts from arid lands in North America. *Soil Biology & Biochemistry*, **41**, 2069–2074.

Stott, M. B., Crowe, M. A., Mountain, B. W., Smirnova, A. V., Hou, S., Alam, M., & Dunfield, P. F. 2008. Isolation of novel bacteria, including a candidate division, from geothermal soils in New Zealand. *Environmental Microbiology*, **10**, 2030–2041.

Stout, R. G., & Al-Niemi, T. S. 2002. Heat-tolerant flowering plants of active geothermal areas in Yellowstone National Park. *Annals of Botony*, **90**, 259–267.

Stout, R. G., Summers, M. L., Kerstetter, T., & McDermott, T. R. 1997. Heat- and acid-tolerance of a grass commonly found in geothermal areas within Yellowstone National Park. *Plant Science*, **130**, 1–9.

Sun, S., Chen, J., Li, W., Altanatas, I., Lin, A., Peltier, S., Stocks, K., Allen, E. E., Ellisman, M., Grethe, J., & Wooley, J. 2010. Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Research*, **doi: 10.1093/nar/gkq1102**.

Sylvia, D. M., Fuhrmann, J. J., Hartel, P. G., & Zuberer, D. A. 2004. *Principles and Application of Soil Microbiology*. 2 edn. Upper Saddle River, NJ, USA: Prentice Hall.

Tamura, K., Peterson, D., Peterson, N., Stencher, G., Nei, M., & Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using likelihood, distance, and parsomony methods. *Molecular Biology and Evolution*, **28**, 38–44.

Taylor, D. L., Booth, M. G., McFarland, J. W., Herriott, I. C., Lennon, N. J., Nusbaum, C., & Marr, T. G. 2008. Increasing ecological inference from high throughput sequencing of fungi in the environment through a tagging approach. *Molecular Ecology Resources*, **8**, 742–752.

Tercek, M. T., & Whitbeck, J. L. 2004. Heat avoidance life history strategy controls the distribution of geothermal *Agrostis* in Yellowstone. *Ecology*, **85**, 1955–1966.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P., & Rubin, E. M. 2005. Comparative metagenomics of microbial communities. *Science*, **308**, 554–557.

Troedsson, C., Lee, R. F., Stokes, V., Walters, T. L., Simonelli, P, & Frischer, M. E. 2008. Develpment of a denaturing high-performance method for detection of Protist parasites of Metazoans. *Applied Environmental Microbiology*, **74**, 4336–4345.

Unterseher, M., Jumpponen, A., Opik, M., Tedersoo, L., Moora, M., Dormann, C. F., & Schnittler, M. 2011. Species abundance distributions and richness estimations in fungal metagenomics - lessons learned from community ecology. *Molecular Ecology*, **20**, 275–285.

van de Voorde, Tess F. J., van der Putten, Wim H., Gamper, Hannes A., Hol, W. H. Gera, & Bezemer, T. Martijn. 2010. Comparing arbuscular mycorrhizal communities of individual plants in a grassland biodiversity experiment. *New Phytologist*, **186**, 746–754.

van der Heijden, M. G. A. 2004. Arbuscular mycorrhizal fungi as support systems for seedling establishment in grassland. *Ecology Letters*, **7**, 293–303.

van der Heijden, M. G. A., Klironomos, J. N., Ursic, M., Moutoglis, P., Streitwolf-Engel, R., Boller, T., Wiemken, A., & Sanders, I. R. 1998. Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature*, **396**, 69–72.

van der Heijden, M. G. A., Wiemken, A., & Sanders, I. R. 2003. Different arbuscular mycorrhizal fungi alter coexistance and resource distribution between co-existing plants. *New Phytologist*, **157**, 569–578.

van Tuinen, D., Jacquot, E., Zhao, B., Gollotte, A., & Gianinazzi-Pearson, V. 1998. Characterization of root colonization profiles by a microcosm community of arbuscular mycorrhizal fungi using 25S rDNA-targeted nested PCR. *Molecular Ecology*, **7**, 879–887.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y., & Smith, H. O. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66–74.

Wagner, M., & Horn, M. 2006. The *Planctomycetes, Verrucomicrobia, Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relavance. *Current Opinion in Biotechnology*, **17**, 241–249.

Walker, J. J., Spear, J. R., & Pace, N. R. 2005. Geobiology of a microbial endolithic community in the Yellowstone geothermal environment. *Nature*, **434**, 1011–1014.

Ward, D. 2009. *The Biology of Deserts*. Biology of Habitats. Oxford: Oxford University Press.

Wardle, D. A. 2002. *Communities and ecosystems: Linking the aboveground and belowground components*. Monographs in population biology, no. 34. Princeton University Press.

Whitford, W. 2002. *Ecology of Desert Systems*. San Fransisco: Academic Press.

Wickham, H. 2009. *ggplot2: elegant graphics for data analysis*. New York, USA: Springer.

Wu, B. Y., Hogetsu, T., Isobe, K., & Ishii, R. 2007. Community structure of arbuscular mycorrhizal fungi in a primary successional volcanic desert on the southeast slope of Mount Fuji. *Mycorrhiza*, **17**, 495–506.

Yeomans, J.C., & Bremner, J.M. 1991. Carbon and nitrogen analysis of soils by automated combustion techniques. *Communications in Soil Science and Plant Analysis*, **22**, 843–850.

Yergeau, E., Bokhorst, S., Kang, S., Zhou, J., Greer, C. W., Aerts, R., & Kowalchuk, G. A. 2011. Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. *ISME Journal*, **doi:10.1038/ismej.2011.124**.

Zak, D. R., Holmes, W. E., White, D. C., Peacock, A. D., & Tilman, D. 2003. Plant diversity, soil microbial communities, and ecosystem function: are there any links? *Ecology*, **84**, 2042–2050.

APPENDICES

APPENDIX A

SAMPLE TOTAL STANDARDIZATION EQUATIONS

$$\forall_{i,k}\ b_{ik} = \sum_{j \in c_i}^{10} a_{jk} \tag{5.1}$$

$$\forall_{i,k}\ f_{ik} = \frac{b_{ik}}{\sum_{k=1}^{28} b_{ik}} \tag{5.2}$$

Where

$$\begin{aligned} i &= site\ \{1, 10\} \\ j &= plant\ \{1, 95\} \\ c &= clustering\ of\ plants\ in\ site\ i\ \{1, C(c_i)\} \\ k &= fungal\ taxa\ \{1, 28\} \\ a_{jk} &= plant_j\ in\ site_i\ containing\ fungal\ taxon_k \\ f_{ik} &= sample\ total\ standardized\ fungal\ abundance \end{aligned}$$

Equations 1 & 2: Sample total standardization was imposed on fungal occurrence data (1) to represent occurrence within a single plant without erroneously applying abundance; and (2) to apply abundance based on occurrence within the larger context of the isolated or communal patch with a standardization to account for different numbers of plants sampled from each patch and different numbers of colonized root segments successfully sequence-identified

APPENDIX B

INFLUENCE OF RAREFACTION ON β-DIVERSITY

Microbial rDNA datasets generated through bar-coded pyrosequencing inevitably produce a different number of sequences from each sample. Prior to comparison of communities through β-diversity analysis, sample-sequence quantities must be normalized in number of sequences per sample, since an increase in observations will likely result in an increase in α-diversity; the common method of normalization is random selection of a uniform number of observations (sequences) per sample. This brings into question the optimal, or inversely, the minimum, number of sequences per sample to detect patterns and differentiate between communities. This has been addressed by Kuczynski *et al.* (2010) by comparing the effects of rarefaction from 10 to 10^4 16S sequences per sample. To ensure that the rarefaction rates applied to the 16S dataset in Chapter 4 were sufficient for robust analysis, and to informally test for normalization-level bias, I applied a series of rarefaction rates to the 16S dataset, from the 1500 sequences per sample used in analysis in Chapter 4, down to 25 sequences per sample. For this comparison, the Canberra dissimilarity index was used for community resemblance in each rarefied dataset. Plot identification letters in the first column of Table B.1 correspond to the plot letters in Figs B.1, B.2 & B.3. Fig. B.1 shows the correlation of pairwise dissimilarity values from each rarefaction level compared to that used in Chapter 4 (1500 sequences per sample). It appears quite clear that these values are highly correlated, and remains above r=0.9 until only 100 sequences were retained per sample (Fig. B.1n). Non-metric multidimensional scaling (NMDS) ordinations with 2 axes were created with each of these rarefied datasets to display community clustering. The effects of this treatment on ordination are displayed in Fig. B.2, though this is somewhat difficult to visually interpret since the NMDS is

rotated in a unique orientation in each plot. The salient piece of information in this figure is that the replicates (joined by gray lines) are consistently clustered together even when only 25 sequences per sample are used to construct the ordination; the exception throughout Fig. B.2 is the same exception seen in Fig. 4.4 from Chapter 4 - sinter surface particles are not well grouped in any ordination. The poor grouping in this set of sinter surface replicate samples was a motivating factor for producing this analysis; it is clear from this analysis that lack of community resolution does not play a role in the poor grouping of sinter surface particles. The next figure, Fig. B.3, shows an NMDS using the full 1500 sequence-per-sample dataset with the movement of each community point traced by thick gray lines. While somewhat redundant with Fig. B.2, this emphasizes the consistent grouping of replicates even when only 25 sequences per sample were used for ordination. The final part of this rarefaction analysis is presented in Fig. B.4; random selection of rarefied sequences was iterated 10 times for each rarefaction level, and the distributions of correlation coefficients from each comparison to that used in Chapter 4 are shown in a box-plot. Again, results are very consistent even when relatively few sequences are used to calculate dissimilarities. These results are consistent with findings from Kuczynski *et al.* (2010), and help to corroborate the rarefaction rates used in this study. It appears that the massive datasets produced by high-throughput sequencing more than adequately characterize the community, and rarefaction down to very low rates (25 sequences per sample, in this case; 10 sequences per sample in Kuczynski *et al.* (2010)) is sufficient to detect patterns across a fine spatial scale, and to differentiate between distinct communities.

Table B.1. Rarefaction comparisons for 16S data. Plot letters correspond to each individual plot in Figs B.1, B.2 & B.3. Dissimilarity correlation coefficient is derived from comparing Canberra dissimilarity values of each rarefied dataset to those from the dataset (1500 sequences per sample) used in Chapter 4 analysis.

Plot	Sequences Per Sample	Dissimilarity Correlation	Procrustes RMSE
a	1400	0.999	0.2660791
b	1300	0.998	0.5923888
c	1200	0.997	1.499809
d	1100	0.995	0.4574903
e	1000	0.991	0.5500876
f	900	0.990	1.273908
g	800	0.988	0.9714175
h	700	0.984	1.445766
i	600	0.981	1.747694
j	500	0.980	1.347681
k	400	0.967	1.741253
l	300	0.956	1.834211
m	200	0.941	1.916856
n	100	0.898	2.27627
o	50	0.818	3.061139
p	25	0.719	4.678412

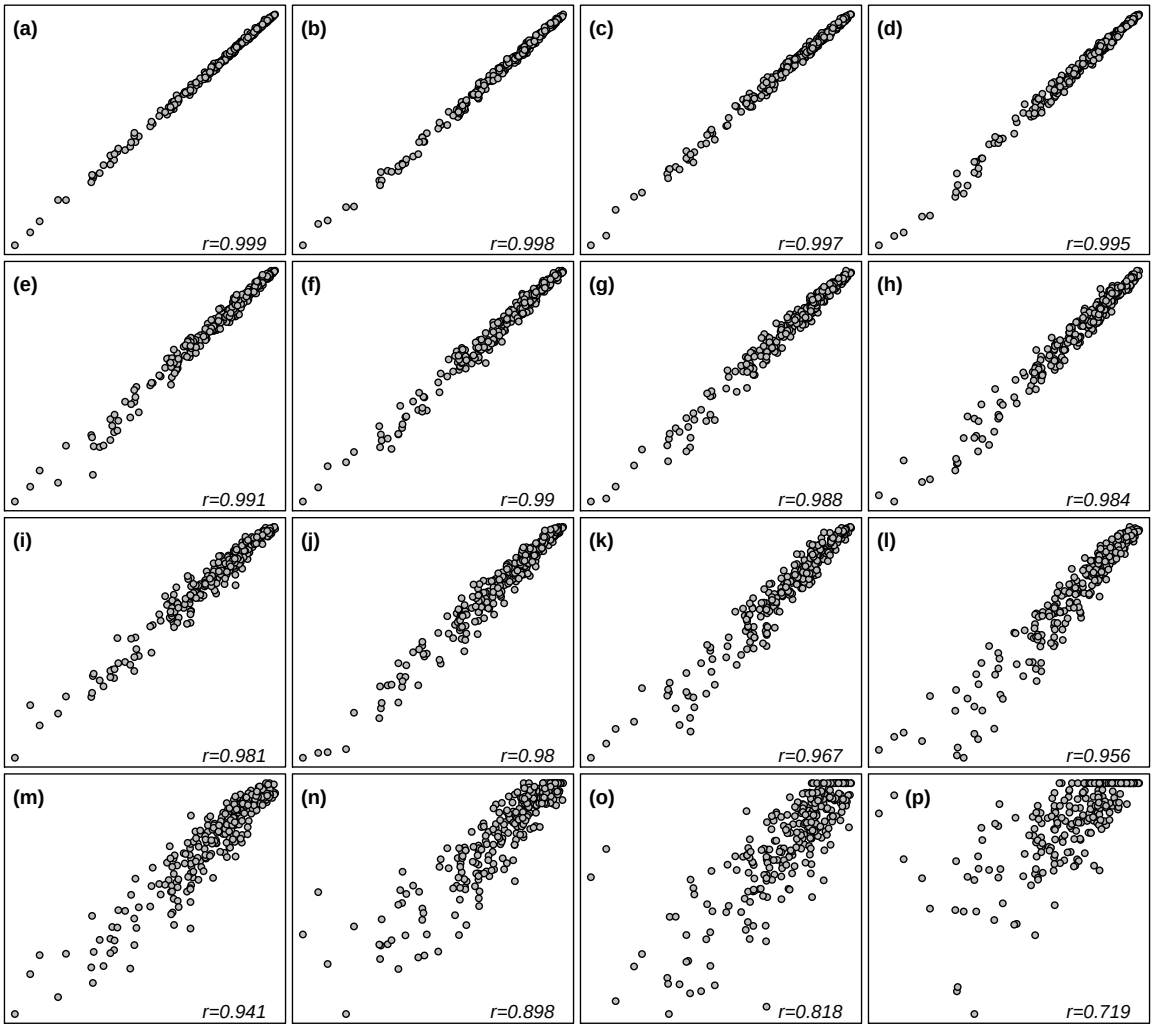

Figure B.1. Dissimilarity correlations. Canberra dissimilarities from prokaryotic datasets rarefied from 1400 sequences per sample (a) stepwise down to 25 sequences per sample (p) compared to those from the full dataset of 1500 sequences per sample. Plot ID letters correspond to Table B.1.

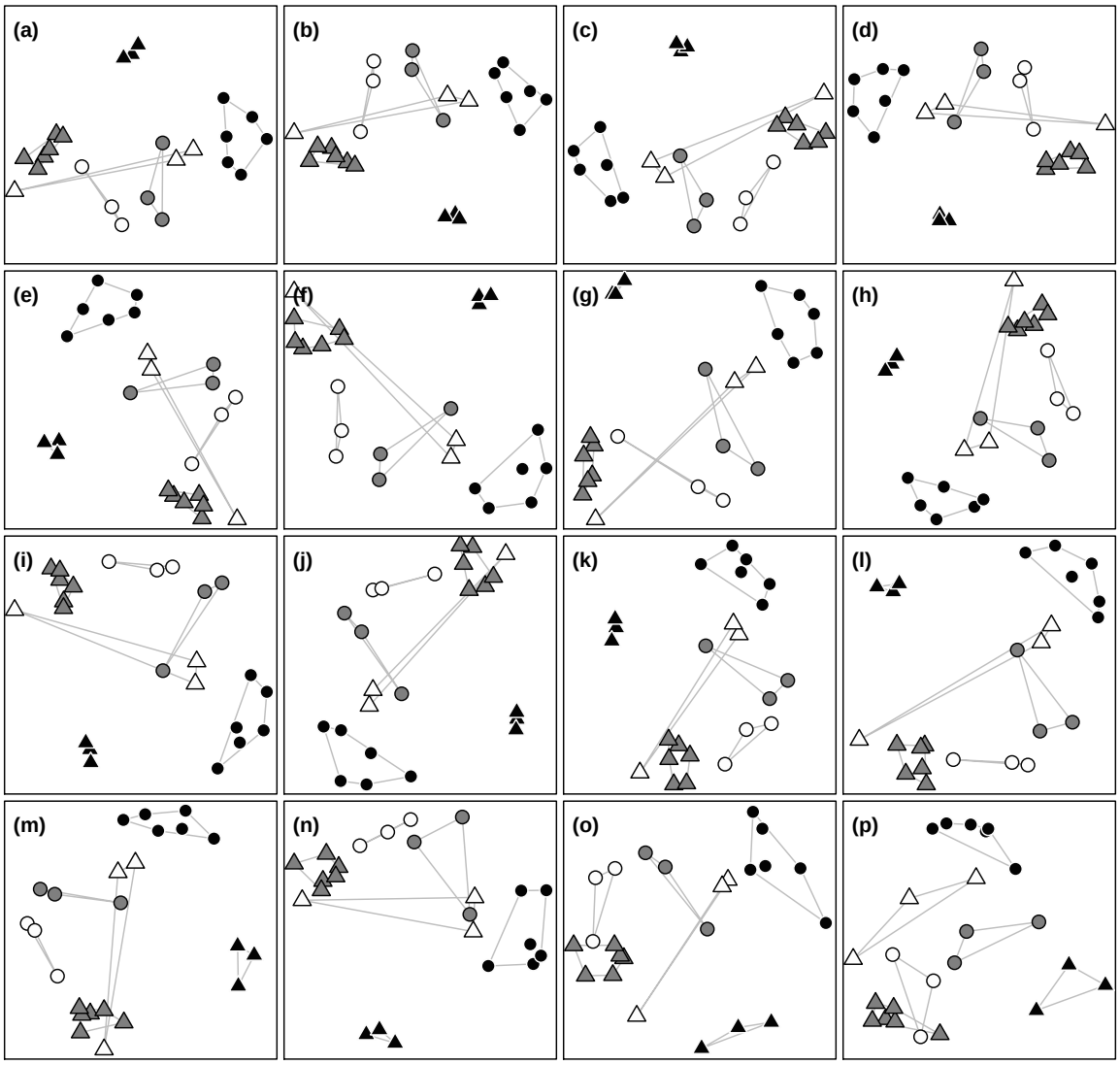

Figure B.2. NMDS comparisons. Each ordination was derived from rarefied datasets (Table B.1) with Canberra dissimilarities; each was produced with 2 axes. Replicate groups are connected by gray lines.

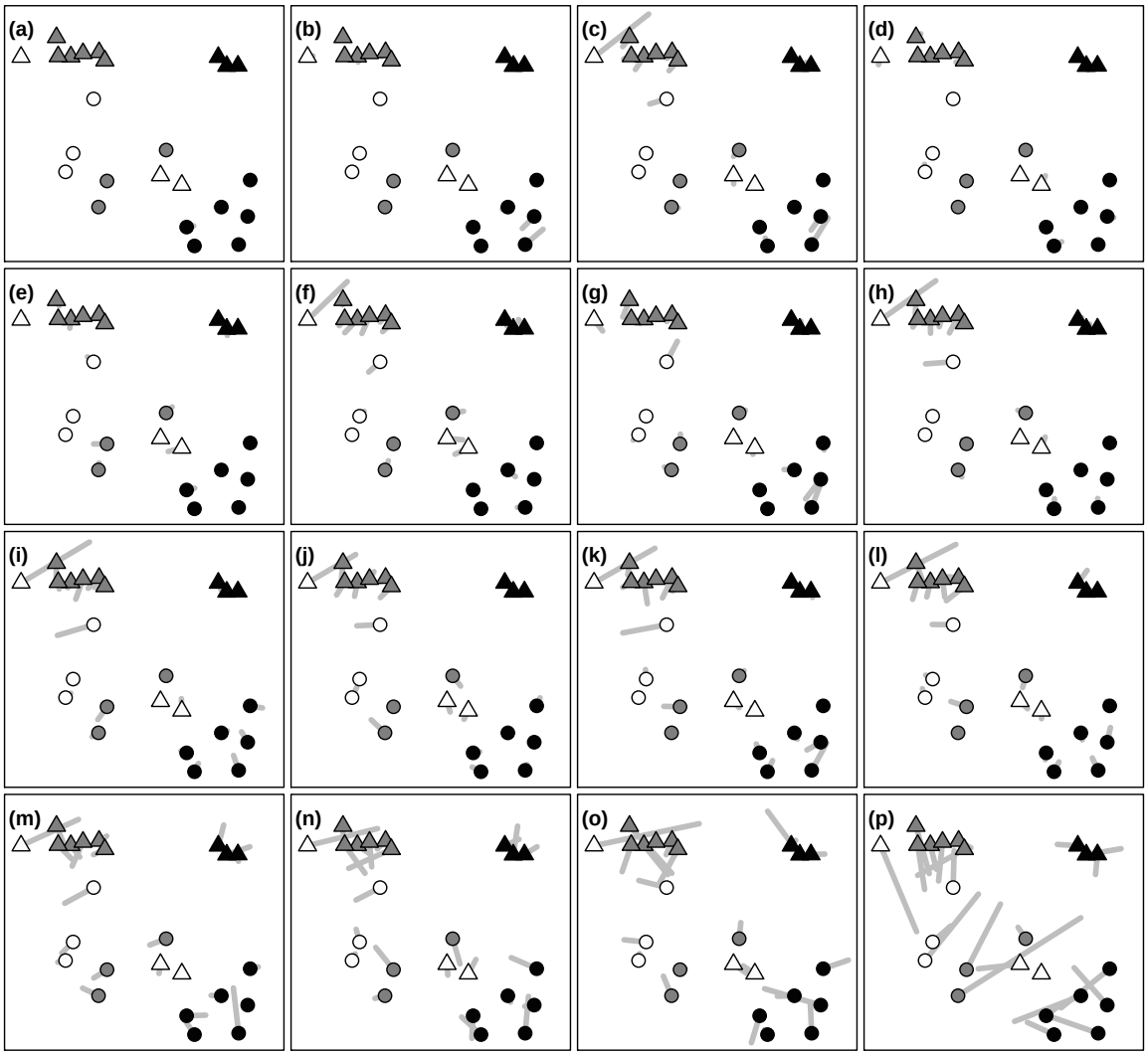

Figure B.3. Procrustes errors. Points in every ordination are the position of NMDS from the full dataset (1500 sequences per sample), and gray lines indicate movement of each point after sequentially reducing dataset size; each corresponds to Table B.1. Orientation has been changed in each and is essentially disregarded to visualize relative proximity of points.

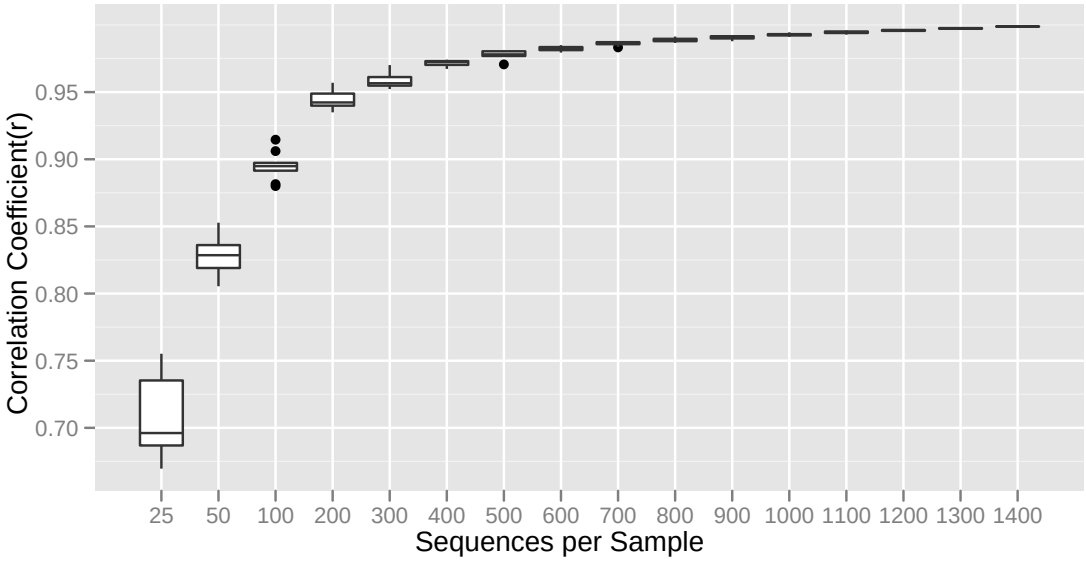

Figure B.4. Iterated rarefaction correlations. Each level or rarefaction was repeated 10 times, and Canberra dissimilarities were compared to the full dataset each time. Boxes show the distributions of correlation coefficients (r).

APPENDIX C

COMBINED 18S AND 16S ORDINATION

An original goal of the study presented in Chapters 3 & 4 was a combined analysis of the prokaryotic and eukaryotic components of the microbial community. The sequencing method I utilized did make this possible, with two problems: 1) the two datasets were difficult to combine given very different diversity characteristics, and thus attempts at normalization were somewhat arbitrarily designed to equalize influence from either community; and 2) little additional information was gleaned from this combined approach. This appendix describes two different ways that I attempted to normalize concatenated datasets, and their results.

The first, and most obvious method of normalization was to rarefy each dataset to exactly the same number of sequences per sample. The results of this first approach are presented in Fig. C.1. The problem that is apparent in Table C.1 is that the two datasets have very different diversity structures, in that the 18S dataset is dominated by a few very abundant bacillariophytes, with low abundance of anything else, while the 16S dataset is far more even, with many more OTUs. The top two subfigures in Fig. C.1 (a & b) are NMDS ordinations that resulted from only 18S and only 16S datasets, respectively, both rarefied to 800 sequences per sample. The central two subfigures of Fig. C.1 (c & d) are the same points seen in a & b, but showing procrustes errors (thick gray lines) indicating movement of points after combination of the datasets. It is important to note that ordination rotation and scaling are accounted for in the procrustes routine, so the errors shown are mapping positions relative to each other. The bottom two subfigures (Fig. C.1e & f) show correlative structure of pairwise Canberra dissimilarity values from 18S (e; x-axis) and 16S (f;

x-axis) with those from the combined dataset (y-axis in both subfigures). When rarefying each respective dataset to 800 sequences, the 18S dataset has only 161 OTUs while the 16S dataset retains 1139 OTUs. The dissimilarity indices that have been used throughout this dissertation, and essentially all others available for multivariate community dissimilarity calculation, depend heavily on species richness. This necessarily means that the 16S dataset, with its 10× higher OTU richness, will have far more influence in the concatenated dataset than that of the 18S. This is exactly the evident effect given the negligible procrustes errors seen in Fig. C.1d compared to Fig. C.1c. The NMDS that resulted from combination was essentially identical to the one from the 16S-only dataset. Clearly, the 16S dataset had more influence in the ordination than the 18S, and thus the equal rarefaction approach taken here is unsuitable when it is desirable to allow inference from both sets of organisms.

Table C.1. Normalization statistics for combined 18S and 16S datasets.

Dataset	Sequences per Sample	OTUs	Shannon Diversity (±sd)	Procrustes RMSE	Correlation
Combined	1600	1300	3.89 (±0.39)	–	–
Eukaryotes (18S)	800	161	2.17 (±0.62)	1.81	0.706
Prokaryotes (16S)	800	1139	4.21 (±0.52)	0.94	0.992

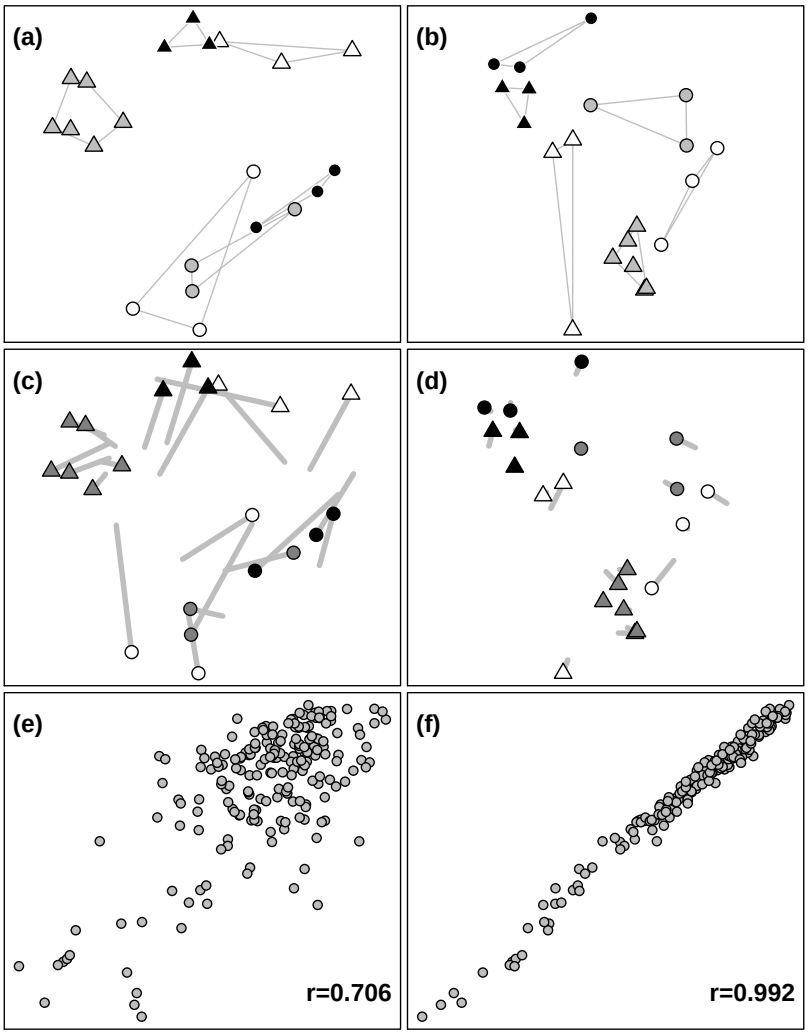

Figure C.1. Combined ordination comparisons of 18S and 16S communities. Both community datasets were rarefied to 800 sequences per sample, combined, and Canberra dissimilarities were calculated. The top two subfigures are 3-dimensional NMDS ordinations of only eukaryotic (a) and only prokaryotic data (b), with replicates joined by thin gray lines; (c & d) Procrustes comparison when considering both datasets combined. Thick gray lines indicate movement of each sample in ordination space (procrustes errors) when 18S (c) and 16S (d) point mappings are compared with combined dataset ordinations; (e & f) Correlations between pairwise Canberra dissimilarity values of 18S (e; x-axis) and 16S (f; x-axis) with those from the combined dataset (y-axis in both subfigures). Correlation coefficients (r) are shown in e & f..

The results from the second approach are displayed in Fig. C.2 and Table C.2. For this approach, the 150 most abundant OTUs were selected from the equally rarefied datasets, and thus the majority of rare prokaryotic OTUs were excluded from analysis. The same methods were used to assess influence; Canberra dissimilarity values were calculated, NMDS ordinations were constructed, procrustes errors were displayed, and pairwise dissimilarities were compared. This approach ameliorates some of the uneven 16S influence seen in Fig. C.1, since procrustes errors are more evenly split between Fig. C.2c & d, but Fig. C.2e & f indicate that 16S sequences were still far more influential in dissimilarity calculation. Additionally, procrustes RMSE values were both higher after reducing OTU numbers, and this is arguably not an intended, or desirable, result. All ordinations also begin to lose coherent replicate grouping after this strict normalization. One potential problem with this approach is summarized in Fig. C.3. Although 18S and 16S sequences are represented by the same OTU richness, and the number of observations is close to even (16,779 and 12,092 sequences per dataset, for 18S and 16S, respectively), the diversity structures of each component of the combined dataset are very different. 18S samples were less even and less diverse than 16S, and rare OTUs (inset of Fig. C.3) were uneven; rare 18S OTUs were represented by a minimum of 3 sequences per OTU, while 16S OTUs were represented by a minimum of 20 sequences, increasing the possibility that rare 16S OTUs were found in multiple samples, and thus feeding more information into dissimilarity calculation than 18S OTUs.

These results illustrate the potential pitfalls of combining two community datasets from organisms with very different population and community structures, and very

different distributional assumptions. I attempted two relatively simple normalizations in an attempt to include information from both groups, but one group (prokaryotes) consistently held uneven influence on the results, and thus the goal of equal inference from both groups was not accomplished. These two normalization schemes are certainly not the only options, but just the two most obvious, and two that might be attempted in the future for similar projects.

Table C.2. Normalization statistics for combined 18S and 16S datasets with 150 most abundant OTUs.

Dataset	Sequences per Dataset	OTUs	Shannon Diversity (±sd)	Procrustes RMSE	Correlation
Combined	28,871	300	3.29 (±0.33)	–	–
Eukaryotes (18S)	12,092	150	2.17 (±0.62)	2.68	0.736
Prokaryotes (16S)	16,779	150	3.27 (±0.32)	1.1	0.973

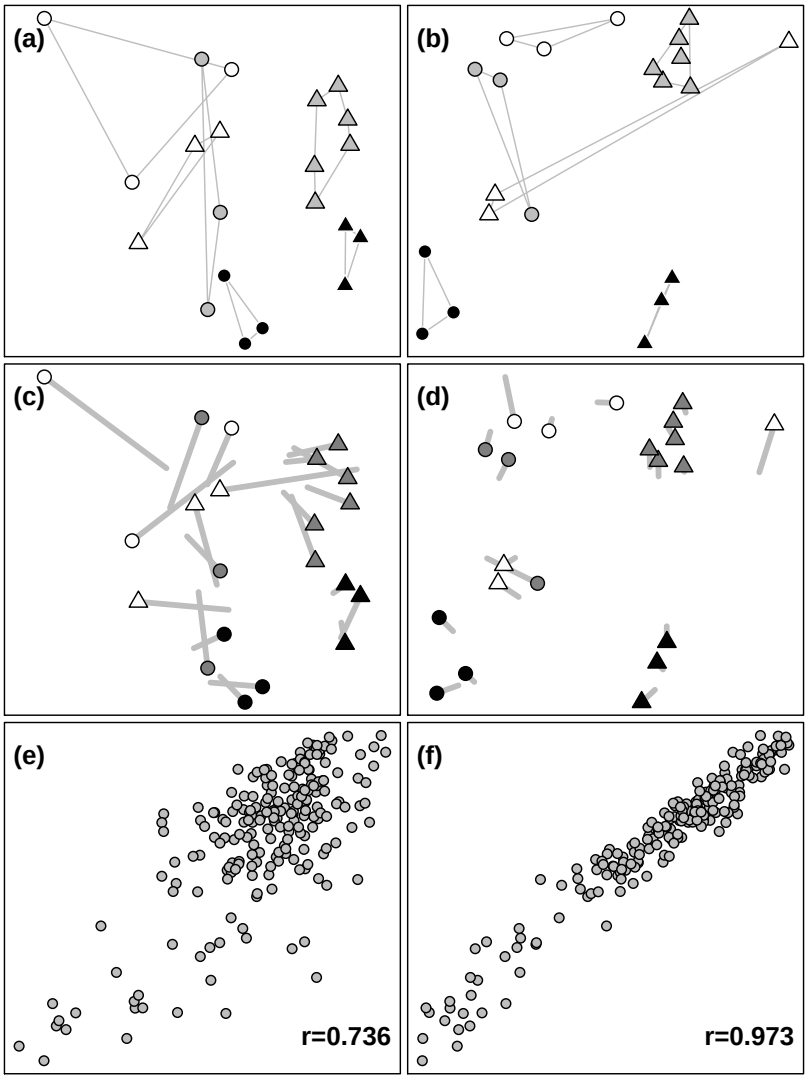

Figure C.2. Combined ordination comparisons of 18S and 16S communities, including only the 150 most abundant OTUs. The 150 most abundant OTUs were extracted from the rarefied datasets, and Canberra dissimilarities were calculated. The top two subfigures are 3-dimensional NMDS ordinations of only eukaryotic (a) and only prokaryotic data (b), with replicates joined by thin gray lines; (c & d) Procrustes comparison when considering both datasets combined. Thick gray lines indicate movement of each sample in ordination space (procrustes errors) when 18S (c) and 16S (d) point mappings are compared with combined dataset ordinations; (e & f) Correlations between pairwise Canberra dissimilarity values of 18S (e; x-axis) and 16S (f; x-axis) with those from the combined dataset (y-axis in both subfigures). Correlation coefficients (r) are shown in e & f. Note that NMDS stress in (a) was above 20 (21.4) for a 2-dimensional NMDS, but this was still used for consistency with other figures in this appendix.

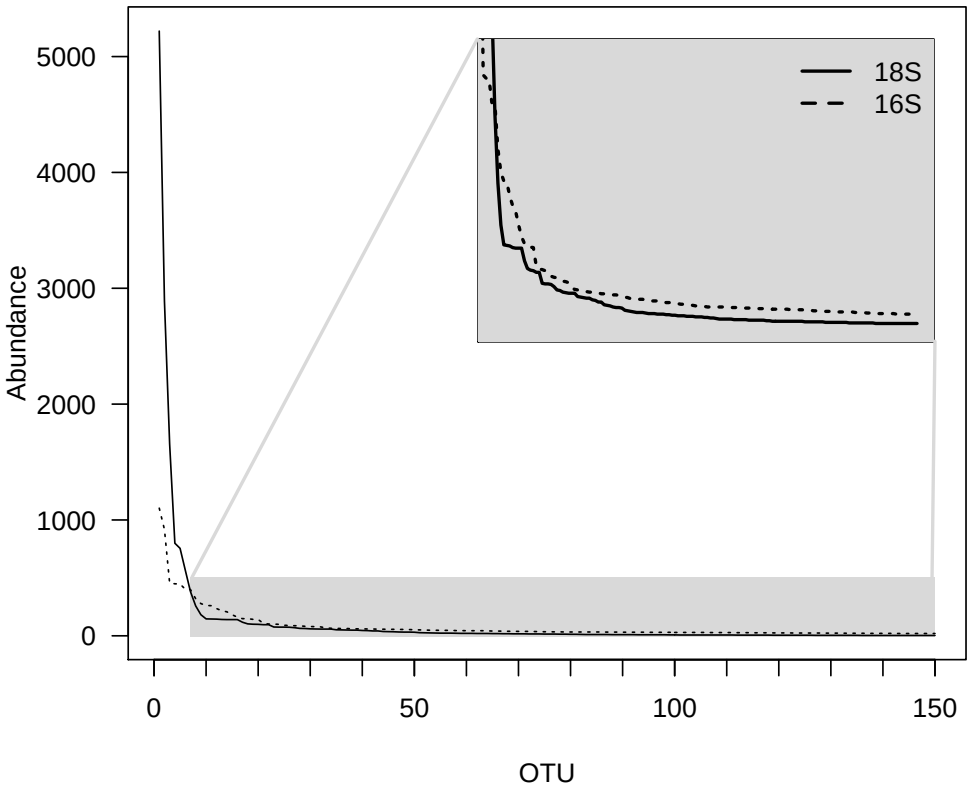

Figure C.3. OTU abundance distributions for normalized datasets. Abundance distributions for 18S and 16S datasets were very different after extracting the 150 most abundant OTUs from rarefied datasets. Tails of the distributions are shown in the inset to emphasize the difference in rare OTUs; these likely play some part in the uneven influence that 16S sequences have in combined ordination, since rare 16S OTUs are represented by a minimum of 20 sequences and the rarest of 18S OTUs are represented by only 3, increasing the likelihood that rare 16S OTUs were found in multiple samples and are thus influencing dissimilarity calculation.