PATTERNS, PROCESSES AND MODELS OF MICROBIAL RECOVERY IN A CHRONOSEQUENCE FOLLOWING REFORESTATION OF RECLAIMED MINE SOILS

Shan Sun

Dissertation submitted to the faculty of
the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
In
Crop and Soil Environmental Sciences

Brian Badgley, Chair
Song Li
Brian Strahm
Michael Strickland
Carl Zipper

Virginia Polytechnic Institute and State University
Blacksburg, Virginia
July 2017

Keywords: soil microorganism, recovery, chronosequence, reclaimed mine soils, amplicon sequencing, shotgun metagenomics, bacterial co-occurrence groups, artificial neural network

PATTERNS, PROCESSES AND MODELS OF MICROBIAL RECOVERY IN A CHRONOSEQUENCE FOLLOWING REFORESTATION OF RECLAIMED MINE SOILS

Shan Sun

Abstract

Soil microbial communities mediate important ecological processes and play essential roles in biogeochemical cycling. Ecosystem disturbances such as surface mining significantly alter soil microbial communities, which could lead to changes or impairment of ecosystem functions. Reforestation procedures were designed to accelerate the reestablishment of plant community and the recovery of the forest ecosystem after reclamation. However, the microbial recovery during reforestation has not been well studied even though this information is essential for evaluating ecosystem restoration success. In addition, the similar starting conditions of mining sites of different ages facilitate a chronosequence approach for studying decades-long microbial community change, which could help generalize theories about ecosystem succession. In this study, the recovery of microbial communities in a chronosequence of reclaimed mine sites spanning 30 years post reforestation along with unmined reference sites was analyzed using next-generation sequencing to characterize soil-microbial abundance, richness, taxonomic composition, interaction patterns and functional genes.

Generally, microbial succession followed a trajectory along the chronosequence age, with communities becoming more similar to reference sites with increasing age. However, two major branches of soil microbiota, bacteria and fungi, showed some contrasting dynamics during ecosystem recovery, which are likely related to the difference in their growth rates, tolerance to environmental change and relationships with plants. For example, bacterial communities displayed more intra-annual variability and more complex co-occurrence networks than did fungi. A transition from copiotrophs to oligotrophs during succession, suggested by taxonomic

composition shifts, indicated that the nutrient availability is one important factor driving microbial succession.

This theory was also supported by metagenomic analysis of the functional genes. For example, the increased abundance of genes involved in virulence, defense and stress response along ages indicated increased competition between microorganisms, which is likely related to a decrease of available nutrients. Metagenomic analysis also revealed that lower relative abundances of methanotrophs and methane monooxygenase at previously-mined sites compared with unmined sites, which supports previous observations that ecological function of methane sink provided by many forest soils has not recovered after 30 years.

Because of the difficulty identifying *in situ* functional mechanisms that link soil microorganisms with environmental change, modeling can be a valuable tool to infer those relationships of microbial communities. However, the extremely high richness of soil microbial communities can result in extremely complicated models that are difficult to interpret. Furthermore, uncertainty about the coherence of ecological function at high microbial taxonomic levels, grouping operational taxonomic units (OTUs) based on phylogenetic linkages can mask trends and relationships of some important OTUs. To investigate other ways to simplify soil microbiome data for modeling, I used co-occurrence patterns of bacterial OTUs to construct functional groups. The resulting groups performed better at characterizing age-related microbial community dynamics and predicted community structures and environmental factors with lower error.

PATTERNS, PROCESSES AND MODELS OF MICROBIAL RECOVERY IN A CHRONOSEQUENCE FOLLOWING REFORESTATION OF RECLAIMED MINE SOILS

Shan Sun

General Audience Abstract

Disturbances to ecosystems are known to largely impact important ecological functions such as soil carbon loss, decreased nutrient retention and increased greenhouse gas emission. As a result, surface mining, which totally removes the topsoil and original vegetation, has severe negative influences on forest ecosystem function. Reforestation is performed on reclaimed mined sites to accelerate the return of forest vegetation and ecosystem functions. Although considerable research has shown that the plant community can be well developed after 30 years, little is known on whether ecosystem functions are also recovered during a similar time period.

As direct mediators of many ecological processes in the environment, soil microorganisms are important for understanding the restoration progress of ecosystems. They could also provide early indications of restoration progress compared to plants. Historically, most soil microorganisms have been difficult to study because they are highly diverse and the majority cannot be cultured in lab, making it difficult to understand changes in the total soil microbiota. However, technological advances such as DNA sequencing have made it feasible to study soil microorganisms in detail. In this work, we studied soil microbial communities from reclaimed mined sites ranging from 5 to 30 years post-reforestation.

We found that overall the microbial community was recovering from the disturbances of surface mining, but many differences from unmined soils still remain after 30 years, such as the unrecovered function as methane sink. Two major groups of soil microorganisms, bacteria and fungi, showed different characteristics during recovery, which are likely due to differences

between the two groups with regard to growth rates, tolerance to environmental change and relationships with plants.

Mathematical modeling is a useful tool for simulating changes and impacts on microbial communities under different conditions, given that actual interactions between microorganisms and their environment can be difficult to measure. However, the high complexity of soil microbial communities becomes an obstacle for modeling that needs to be addressed by simplifying data describing soil microbial community. One approach is grouping organisms based on their natural evolutionary relationships, but this can mask the trends of some microorganisms since all organisms in these groups do not always respond the same to environmental change. Here we used a method of grouping microorganisms based on their co-occurrence patterns, which resulted in better predictions of changes in community structure and environmental factors when applied in modeling.

Acknowledgements

I would like to thank my advisor Dr. Brian Badgley for his guidance, help and support throughout my PhD study. His valuable and continuous guidance in labwork, data analysis, writing and presentation helped me gain the knowledge and skills for research and will benefit my future career. He also encouraged me and gave me the freedom to explore the research areas I am interested in. This work would not be possible without him.

I also would like to thank all my committee members: Dr. Song Li, Dr. Brian Strahm, Dr. Michael Strickland and Dr. Carl Zipper for their guidance and help during my study. I would also like to acknowledge and thank Dr. James Burger for his previous work in the Powell River Project that established the sites used in this study. Thanks to all my colleagues for their help during my study here: Lucas Waller, Romina Benitez, Catie Capellin, Bethany Avera, Chao Qin, Li Ma, Theresa Sosienski, Hanh Le, Fatmaalzhraa Awad, Stephanie Kulesza, Bee Khim Chim, Junxue Wu, Bishal Tamang, Jasper Alpuerto, Heyang Yuan and Xiaojin Li.

Finally I would like to thank my parents Jinqi Sun and Guilan Long for their love and support.

This dissertation was supported by Powell River Project and Virginia Agricultural Experiment Station.

Table of Contents

List of Abbreviations

ANN	Artificial Neural Network
ANOSIM	Analysis Of Similarity
ANOVA	Analysis Of Variance
AOA	Ammonia Oxidizing Archaea
AOB	Ammonia Oxidizing Bacteria
AOM	Ammonia Oxidizing Microorganisms
BCG	Bacteria Co-occurrence Group
CPM	Count Per Million
DAG	Directed Acyclic Graph
DGGE	Denaturing Gradient Gel Electrophoresis
FDR	False Discovery Rate
NOB	Nitrite Oxidizing Bacteria
NMDS	Non-metric Multidimensional Scaling
OTU	Operational Taxonomic Unit
rRMSE	relative Root-Mean-Square Error
SIR	Substrate Induced Respiration
SOM	Soil Organic Matter
SRC	Spearman Rank Correlations
TMB	Total Microbial Biomass
T-RFLP	Terminal Restriction Fragment Length Polymorphism
Tukey's HSD test	Tukey's honestly significant difference test
WCGNA	Weighted Correlation Network Analysis, also known as Weighted

Chapter 1. Introduction.

Soil microorganisms are important mediators of biogeochemical functions and play essential roles in the establishment of plant communities during ecosystem succession. Similar to plant communities, microbial communities can be significantly altered by disturbances, which can lead to impaired ecosystem functions (1, 2). However, the succession of microbial communities after disturbance has been less intensively studied until only recently because soil microbial communities are highly diverse and the majority of organisms are unculturable. Fortunately, recent advances in next-generation methods for DNA sequencing have largely improved the feasibility of characterizing microbial community structure and function from environmental samples, making it possible to examine microbial recovery in greater detail.

There are many advantages to studying the response of microorganisms specifically during ecosystem succession. For example, the rapid responses of microorganisms to environmental change, high turnover rates, and their involvement in many biogeochemical processes could potentially make them better indicators of ecosystem restoration progress compared to the plant community (3). Secondly, knowledge about post-disturbance microbial succession could greatly contribute to our understanding of the factors controlling the return of ecosystem structure and function during restoration. Finally, studying microbial succession provides an opportunity to compare observed processes to plant succession, which can help generalize the succession theories to a broader scale.

In previous studies of microbial succession in different environments, some trends have emerged. Microbial biomass generally increases during the process, while microbial diversity can be highly variable without a universal pattern (4-7). Changes in microbial diversity during

succession are highly variable, but generally microbial diversity may increase in the early stages when new communities are established, and fluctuate during shifts in community type. It has been hypothesized that the phylogenetic diversity would peak when resource supply is intermediate and multiple various carbon resources meet the demands of different groups (4). However, such predictions above are based on the assumption that nutrient limitation is the primary factor influencing microbial succession, while the dynamics of microbial communities are also highly susceptible to environmental factors such as pH, moisture, and vegetation, which make a succession trajectory more difficult to predict. Resource limitation and other environmental influences must be both considered to understand phylogenetic and functional alteration of microbial communities during succession.

Reclaimed mine soils represent a unique model to study soil development, including the roles of soil microorganisms (8), and have excellent potential for testing ecological theories related to microbial succession (9). This is primarily because reclamation and reforestation after mining follow similar procedures and create similar sites over time, which facilitates a chronosequence approach (10). Furthermore, surface mining causes severe disturbance of ecosystems (11, 12) as soils in mined sites are replaced by overburden and salvaged topsoil and original vegetation is removed. It also causes the displacement of the native soil microbial community and collapse of entire ecosystem, with negative influences remaining in soils even after 20 years of restoration (13). The study of microbial succession in reforested mine soils could provide information regarding factors that control the recovery of soil properties and reestablishment of soil ecological processes, which are essential for the recovery of ecologically sound system (8).

In this study, I examined the post-disturbance patterns of microbial taxonomic and functional

succession using amplicon sequencing of bacterial and fungal phylogenetic marker regions and shotgun metagenome sequencing. These approaches provide a comprehensive view of the changes in abundance, richness, taxonomic composition, potential interactions of bacterial and fungal communities, and changes in functional genes for the overall soil microbial community. With the resulting community structure data, I also proposed a way of simplifying microbial community structure for modeling and built artificial neural networks (ANN) to test its ability to predict changes in community structure and environmental factors.

References

1. Fichtner A, Von Oheimb G, Härdtle W, Wilken C, Gutknecht J. 2014. Effects of anthropogenic disturbances on soil microbial communities in oak forests persist for more than 100 years. Soil Biology and Biochemistry 70:79-87.
2. Ferrenberg S, O'Neill SP, Knelman JE, Todd B, Duggan S, Bradley D, Robinson T, Schmidt SK, Townsend AR, Williams MW. 2013. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. The ISME journal 7:1102-1111.
3. Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL, Murphy DV. 2011. Soil Microbial Community Successional Patterns during Forest Ecosystem Restoration. Appl Environ Microbiol 77:6158-6164.
4. Fierer N, Nemergut D, Knight R, Craine JM. 2010. Changes through time: integrating microorganisms into the study of succession. Res Microbiol 161:635-42.
5. Nemergut DR, Anderson SP, Cleveland CC, Martin AP, Miller AE, Seimon A, Schmidt SK. 2007. Microbial community succession in an unvegetated, recently deglaciated soil. Microb Ecol 53:110-22.
6. Schutte UM, Abdo Z, Bent SJ, Williams CJ, Schneider GM, Solheim B, Forney LJ. 2009. Bacterial succession in a glacier foreland of the High Arctic. ISME J 3:1258-68.
7. Zumsteg A, Luster J, Göransson H, Smittenberg R, Brunner I, Bernasconi S, Zeyer J, Frey B. 2012. Bacterial, Archaeal and Fungal Succession in the Forefield of a Receding Glacier. Microb Ecol 63:552-564.
8. Bradshaw A. 1997. Restoration of mined lands—using natural processes. Ecological Engineering 8:255-269.
9. Schipper LA, Degens BP, Sparling GP, Duncan LC. 2001. Changes in microbial heterotrophic diversity along five plant successional sequences. Soil Biology and Biochemistry 33:2093-2103.
10. Frouz J, Prach K, Pižl V, Háněl L, Starý J, Tajovský K, Materna J, Balík V, Kalčík J, Řehounková K. 2008. Interactions between soil development, vegetation and soil

fauna during spontaneous succession in post mining sites. Eur J Soil Biol 44:109-121.
11. Bradshaw A. 2000. The use of natural processes in reclamation — advantages and difficulties. Landscape and Urban Planning 51:89-100.
12. Frouz J, Jílková V, Cajthaml T, Pižl V, Tajovský K, Háněl L, Burešová A, Šimáčková H, Kolaříková K, Franklin J, Nawrot J, Groninger JW, Stahl PD. 2013. Soil biota in post-mining sites along a climatic gradient in the USA: Simple communities in shortgrass prairie recover faster than complex communities in tallgrass prairie and forest. Soil Biology and Biochemistry 67:212-225.
13. Lewis D, Chauhan A, White J, Overholt W, Green S, Jasrotia P, Wafula D, Jagoe C. 2012. Microbial and Geochemical Assessment of Bauxitic Un-mined and Post-mined Chronosequence Soils from Mocho Mountains, Jamaica. Microbial Ecology 64:738-749.

Chapter 2. Literature review.

The knowledge of microbial succession is not only important for the development of ecological theories but also could provide important information regarding the recovery of ecosystem processes after disturbances. Soil microbial communities are significant mediators of ecological functions and essential for the ecosystem function. Moreover, the high turnover rates and rapid response to environmental change of microorganisms make them potential indicators for the progress of ecosystem restoration. Microbial succession is under the influence of a variety of environmental factors such as nutrient availability and vegetation. The knowledge on the interactions between microorganisms and their environment is essential for a better understanding of the mechanisms controlling microbial succession. Here I reviewed studies of microbial succession in multiple environments and the influence of environmental factors on microbial succession.

2.1. Microbial communities in restored mine soils

In the restoration of mined sites, soils are replaced by overburden and salvaged topsoil, which differs substantially from developed soils (1, 2). Edaphic factors including texture, drainage and chemical characteristics are significantly altered, the original plant community has been removed, and the native soil microbial community has been significantly altered. In general, reclaimed mine soils are higher in pH but lower in organic matter, total C, and total N, and those effects can be detected even after 20 years of restoration, indicating long-standing influences of mining activities on soil biogeochemical processes and microbial communities (3).

The recovery of soil properties and reestablishment of soil ecological functions are of significant influence for the recovery of ecosystem, given that they support the continued reestablishment of

the plant community (4). The close relationships between microbial communities and their environments support the hypothesis that microbial community could work as integrated estimation of the soil environment, which may be better than any single soil attribute at evaluating success of ecosystem restoration (5, 6). In addition, rapid responses of microorganisms to environmental factors and significant roles in biogeochemical cycling further make them good potential indicators of restoration of ecosystem function. Clear relationships between increases in soil microbial biomass, restoration time, and taxonomic shifts indicate that the structure of soil microbial communities can become more similar to unmined forest soils with reforestation age, and may follow a predictable succession pattern in developing ecosystems (6). For example, Banning *et al*. reported increasing similarity of bacterial community structure to unmined forest soils during restoration of bauxite mined soils, with community similarity values higher than for similar comparisons of vegetation structure (6).

Some specific microbial taxa have been found to shift significantly during the restoration of post mining sites. Bacterial community shifts during restoration included a decrease in Firmicutes, Bacteroidetes, and Acidobacteria coupled with increases in Proteobacteria and Verrucomicrobia with restoration time in post-mining soils (3, 6). The finding that several proteobacterial genera, including *Bradyrhizobium*, *Rhodoplanes*, and *Methylobacterium*, had relatively higher abundance in un-mined soil than post-mining soils may suggest possibly unrecovered ecological processes associated with those organisms, as these species are considered metabolically versatile and involved in processes such as N fixation, nutrient cycling and pollutants degradation (7-9). Acidobacteria are poorly characterized, but sequenced acidobacterial genomes suggest that they might be able to tolerate long dry periods, low pH, and oligotrophic conditions (7, 10), which may aid their survival in newly restored soils.

In contrast to bacteria, previously studied fungal communities exhibited overall lower levels of similarity between unmined and reforested soils (6). It was also suggested that fungi, in particular mycorrhizal fungi, might have more association with vegetation structure and less efficient dispersal to newly restored sites compared to bacteria, which might lead to a slower recovery of fungal community (11). Overall, however, it remains unclear how environmental factors and/or the development of the microbial community itself drives ecosystem trends of succession during restoration, nor is it well known how changes of some specific microbial taxa are related to changes in environmental variables. Additional knowledge about which groups are mainly driving the shift of the soil microbial community and which environment factors most affect them will benefit the further development of successional theory and also the application of microbial indicators of ecosystem restoration.

2.2 Examples of soil microbial succession in other systems

2.2.1. Microbial succession in glacial forefields

Soils exposed after glacial retreat are other ideal environments for studying chronosequences of microbial succession, as microorganisms are the first colonizers contributing to soil development and biogeochemical cycling, prior to plant colonization. Reported changes of microbial diversity during succession in these systems are variable, however, as both constant diversity and increases in diversity with soil age have been detected (12-14). These different results may suggest that soil age is not the only important factor driving changes of microbial diversity, which is possibly also related to other factors including climate, pH and soil texture.

With regard to different soil microbiota, distinct shifts in bacterial, fungal and archaeal community structures can occur along glacial chronosequences (14). In early succession, limited

C and N select for specific microbial communities, which mainly comprise of C and N fixers and chemolithotrophic groups utilizing rock minerals (15-17). Once soils are vegetated, decomposers of plant residue can thrive and compete with microorganisms that are successful in the earlier stages. Overall, Proteobacteria has been reported as the most abundant phylum in all stages of succession (14) and the functional versatility of Proteobacteria, which consists of numerous phototrophs, photoheterotrophs and chemolithotrophs, gives this phylum advantages in environments with limited resources. Other abundant groups in early succession can include Cyanobacteria, Acidobacteria and Actinobacteria. As first colonizers, photoautotrophic cyanobacteria can survive with low nutrient inputs and they provide large inputs of C and N to soils, which are important for the further succession of heterotrophs (18, 19). The abundance of autotrophs would likely decrease with increasing plant input for heterotrophs and decreased light availability in the soil following canopy establishment.

Patterns in archaeal diversity have been more difficult to generalize, as both increased and decreased diversity along a successional gradient have been reported (14, 20). Many archaea are known for the ability to survive in extreme conditions, so the fluctuations of temperature and moisture and the lack of nutrients in new bare soil may provide various microhabitats that can accommodate different archaeal species. In one study archaeal community structure differed between young and old soils, with a shift from Euryarchaeota-dominated to Crenarchaeota-dominated (14), whereas in another study Crenarchaeota were dominant over the whole glacial forefield (20). Euryarchaeota and Crenarchaeota can adapt to soils with limited nutrients, and Crenarchaeota can be resistant to freeze-thaw cycles (21, 22).

Fungal communities have been found to be very different in bare soils and vegetated soils, which is likely related to the close relationships between fungi and plants. The most abundant phylum of fungal community shifted from Ascomycota, which could live on rocks (23), to Basidiomycota, which primarily colonize vegetated soils, during plant succession in deglaciated soils (14, 24). The dominance of Ascomycota can be partly explained by the appearance of lichen in early succession given that most lichen-forming fungal species belong to Lecanoromycetes, a class belonging to Ascomycota (14). The dominance of Basidiomycota in vegetated soils may be explained by their mycorrhizal associations with early colonizing plants and other litter decomposers in this phylum (25, 26).

Soil microbial taxonomic shifts in glacial forefields were also found to be associated with changes in soil properties. Carbon and nitrogen content, base saturation, pH, and vegetation were all reported to explain 25–30% of the microbial community change along a glacial forefield chronosequence (14), and changes in soil C and N across deglaciated chronosequences were shown to influence bacterial community composition and function (27, 28). Besides nutrients, pH was also an important variable significantly related to microbial community structure (29), and water limitation also had effects on the microbial communities in dry Antarctic glacier soils (30). Plant colonization also influences the successional trajectory of microbial communities in deglaciated soils (31, 32), and microbial communities have been shown to have higher fungi:bacteria ratios in late-stage vegetated soils than in recently deglaciated bare soils (29, 33).

In addition to overall community structure, soil parameters can also impact specific dominant taxa in glacial habitats. For example, the abundance of Bacteroidetes was correlated with cold (34-36), while Cyanobacteria, Deltaproteobacteria and Gemmatimonadetes have been shown to

increase with soil moisture and pH (37). Trace elements (magnesium, calcium, potassium) can influence the abundance of Actinobacteria (37), while nutrient availability and soil moisture have been shown to impact Acidobacteria (38). Overall, however, the controlling mechanisms of soil over dominant microbial taxa remain unclear because of the lack of enough cultured representatives from many microbial groups.

2.2.2. Microbial succession in abandoned arable fields

During secondary vegetative succession in chalk grasslands after abandonment of agriculture, microbial succession also shows clear patterns under more influence of soil nutrient change than vegetation (39). Different from the increases in plant diversity in these systems, microbial richness decreased with field age (39), which is congruent with the finding that agricultural soils can harbor higher microbial richness compared with forest soils (40). Firmicutes were abundant in all stages of grassland succession, and Actinobacteria was the second most abundant phylum (39). This microbial community composition is distinct from those in some other sites in the grassland system, where Bacteroidetes and Acidobacteria tend to have the highest abundance (40, 41). Soils of early successional stage had high relative abundance of Bacteroidetes, Cyanobacteria and Euryarchaeota, and low relative abundance of Acidobacteria, Actinobacteria, Crenarchaeota, Deltaproteobacteria and Verrucomicrobia, while soils of late stages had high relative abundances of Actinobacteria, Alphaproteobacteria, Betaproteobacteria and Planctomycetes and low relative abundances of Firmicutes and Crenarchaeota (39).

One study of soil microbial succession in abandoned arable fields in dry ecosystems also revealed clear successional patterns regarding changes in microbial biomass, activity and community composition, along with increases of fungal and bacterial biomass and microbial

activity over time (42). Different from the study on chalk grasslands with the dominance of Firmicutes, Actinobacteria was the phylum of the highest abundance across all stages in this system. The higher abundances of Actinobacteria in the initial stages (42) agrees with a previous finding that Actinobacteridae are more abundant in soils without vegetation than under shrubs (43). In contrast, abundances of Acidobacteria and Gemmatimonadetes increased with soil age (42), which may be affected by soil particle size (44). In contrast to the study on chalk grasslands, microbial succession in this system was reported to be closely related to plant succession, which may be caused by the different climate and dominant plants of these two ecosystems.

Overall, reports of microbial succession in different systems have followed different patterns as the starting conditions are different and the succession process is highly susceptible to the surrounding environment. But in general, succession in bare soils is usually characterized by the presence of autotrophic groups in early stages and the thriving of heterotrophic groups in later stages. Biomass increases during succession and diversity increases first and then decreases. Microbial succession in soils with ample nutrients, like abandoned arable fields, is less predictable. In these systems, the trajectory of succession seems to be more dependent on the changes of local nutrient availability and vegetation. Furthermore, it is still unclear how much different environmental factors contribute to the effects and which microbial groups are primarily influenced by those factors.

2.3. Environmental regulation of soil microbial succession

2.3.1. pH

Soil pH greatly influences the composition and structure of soil microbial communities across diverse biomes, which may be the best predictor for microbial community composition at a relatively large scale (45-49). Fierer et al. found that bacterial diversity across different ecosystems in a continental scale could be mostly explained by soil pH (49). At a smaller scale, Rousk et al. demonstrated that pH changes across 180 m in a liming experiment drove changes in microbial community structure that were approximately equal to the change in natural soils with a similar pH gradient across North and South America (45).

Specific responses of soil microorganisms to pH include a general increase in diversity from pH 4 to 8.5 (45). The relative abundance of Acidobacteria has been shown to negatively correlate with pH, while the relative abundance of Actinobacteria has not shown clear relationships with soil pH (45, 50, 51). The relative abundances of Bacteroidetes, Nitrospira and Proteobacteria have been shown to increase with pH, but the correlations are not always significant (13). The fungal community structure was also related to soil pH, but the observed relationship is not as strong as that for bacterial community (45). This could be explained by the different ecological physiologies of bacteria and fungi, specifically that fungi typically have wide pH optima (52, 53). It is also likely that competitive interactions between bacteria and fungi could influence the shift of fungal community along a pH gradient, as Rousk et al. reported that fungi were stimulated by inhibition of bacterial community (54).

2.3.2. Nutrients

Resource availability is believed to be one of the fundamental drivers of microbial communities, so it is not surprising that it plays an important role during succession as well (55). For example, soil bacterial community structures in three-year old soils following one year of application of N and P fertilizer became more similar to unamended 85-year old soils, even though there were still differences in other soil properties (56). Adding organic waste has also been shown to largely increase gross N mineralization, nitrification and immobilization in coal mined sites, with an effect magnitude similar to comparison with the reference ecosystems (57). Besides direct effects from nutrient availability, microbial communities could also be indirectly mediated by nutrients through changes in other aspects of abiotic factors and vegetation. In addition to total nutrients, specific influences of carbon, nitrogen and phosphorus on microbial succession are discussed below.

With regard to soil carbon, Eilers et al. reported that adding low molecular weight carbon compounds to soils caused the bacterial community to shift, and the bacterial communities had different compositions after different compounds were added (58). Added compounds included glucose, glycine and citric acid, which are common components of root exudate, and had large effects on Betaproteobacteria, Actinobacteria and Bacteroidetes, modest effects on Alphaproteobacteria and Gammaproteobacteria and little effects on Acidobacteria and Gemmatimonadates (58). Within Betaproteobacteria and Actinobacteria, the relative abundances of Burkholderiales, Actinobacteridae and Rubrobacteridae increased largely with addition of glucose (58). Increases in the relative abundance of Betaproteobacteria and Gammaproteobacteria were also found to be related to the addition of C in other studies (59, 60). The abundance of Acidobacteria decreased with addition of C compounds, which agrees with the

finding that Acidobacteria have a tendency of being more abundant in soils with low C availability (60, 61). This phenomenon may be explained by competitive interactions between copiotrophic and oligotrophic microorganisms. Copiotrophic bacteria can reproduce rapidly with high nutrient availability while oligotrophic bacteria are relatively abundant in resource-limited conditions. In natural ecosystems, microbial communities encounter much more complex combinations of C resources and have longer response times, so the microbial community shifts may be different from those induced by addition of simple carbon compounds. However, it is believed that in general, the competition between copiotrophs and oligotrophs will play an important role in regulating microbial community shifts.

The significance of nitrogen (N) limitation in early stages of microbial succession was supported by several studies (18, 62, 63), and N fertilization can induce a shift in microbial community structure (64). Microbial communities in soils with higher N additions had higher relative abundances of Proteobacteria, Bacteroidetes and Actinobacteria and a lower relative abundance of Acidobacteria (64), and many of the dominant groups have been identified as copiotrophic taxa, while groups with lower relative abundance were identified as oligotrophic. Metagenomic analysis and catabolic assays have also demonstrated that microbial communities with high N inputs are significantly different from those with no or low N inputs, and microbial communities became more copiotrophic with increasing N inputs (64). However, microbial communities in soils with low N inputs were not significantly different from those without N input, even though the low level of N addition caused shifts in plant communities and increases in plant productivity (64). It may be because plant communities are more sensitive to N additions and have advantages in competition with soil microbial communities for N. Using metagenomics, genes related to DNA/RNA replication, electron transport and protein metabolism have been shown to increase

across N gradients (64), but because of the shallow sequencing depths in this study, genes more directly associated with C and N cycles were not abundant enough to determine significant differences. Furthermore, this evidence is not sufficient to determine whether nitrogen effects on microbial communities are direct responses to N supply or indirect responses to increases in organic C caused by increased plant productivity.

While it is typically expected that phosphorus exert a larger influence on autotrophs than heterotrophs, in general the influence of phosphorus (P) on microbial succession is not well studied. The addition of P altered autotrophic succession in unvegetated deglaciated soils, with the relative abundance of Cyanobacteria increasing significantly and microbial community structure shifting compared to the control (56). In contrast, the P addition did not show significant effects on the heterotrophic community (56). Autotrophs, especially photoautotrophs, have advantages over heterotrophs in soils with relatively low organic C concentration, as the availability of C may be a much more important limitation for heterotrophs than P.

Overall, one of the clearest influences of nutrient availability on microbial community composition is the shift between different trophic groups, such as the shift between autotrophs and heterotrophs in the very early stages with low nutrient level and the shift between copiotrophs and oligotrophs with nutrient type change. Copiotrophs have relatively high metabolic activity, turnover rates and substrate use efficiency, while oligotrophs can metabolize more recalcitrant C substrates, survive with low nutrient availability and have lower growth rates (55). The copiotroph-oligotroph transition was driven by change of nutrient availability, and this transition could also lead to changes in soil biomass and respiration rates. Copiotrophic groups tend to comprise a small proportion of soil biomass but maintain high activity, so microbial mass

is not always positively correlated with microbial activity (64). Copiotrophs dominating soils could have a relatively low total microbial respiration rate because of little utilization of recalcitrant C, which is a major component of soil C pool (55). These hypotheses are consistent with the wide observation that soil microbial respiration rates and biomass decrease with addition of N (65-68). In addition to competition between copiotrophs and oligotrophs, competition between autotrophs and heterotrophs may occur even earlier in the process, at the beginning of secondary succession. Autotrophs in particular have advantages in soils with low concentrations of total C, but once vegetation recovers and accumulation of organic C begins, heterotrophs thrive and dominate in later stages (55).

2.3.3. Temperature and moisture

Temperature and precipitation across various ecosystems did not show significant effects on bacterial community composition at a global scale (49), even though microbial activity has been shown to decrease with decreasing soil moisture and temperature across biomes (69, 70). In contrast, it was found that mean annual temperature and precipitation had significant influences on fungal community composition at a continental scale (71, 72). Fungal richness increased with precipitation gradient at the high end, which is likely explained by variability in tree diversity across ecosystems. Fungal diversity is usually more closely related to tree species than bacterial diversity and there are obligate associations between some ectomycorrhizal fungi and host trees (73-75). For example, *Russula* and *Cortinarius* are ectomycorrhizal fungi dominating mature forest soils in high latitudes, but they were found to be absent in tropical and southern temperate soils (71).

At smaller spatial scales, microbial community structure and activity have been observed to

change across a temperature and soil moisture gradient in deglaciated unvegetated sites (68). At these sites microbial activity was higher in cold and moist sites than hot and dry sites. This could be explained by the relatively high accumulation and availability of organic C and nutrients in cold and moist soils (76), plus soil moisture increases microorganism motility and access to nutrients (77), as diffusion limitation can limit microbial activity in dry soils (69). With regard to specific taxa, Acidobacteria became more abundant in drier and warmer soils, and Actinobacteria exhibited the opposite pattern (78). Gammaproteobacteria were sensitive to soil moisture and usually of lower abundance in drier soils (78, 79). As Gammaproteobacteria are often considered to be a copiotrophic group (61), their abundance may be driven by the increased C availability in moist soils. Bacteroidetes was generally more abundant in dry soils, while Verrucomicrobia displayed an opposite pattern (80).

2.3.4. Soil Texture

Soil texture can also have significant effects on bacterial biomass, richness and community structure across ecosystems (71, 81), although this may be a result of significant correlations between the soil texture index and other soil properties including pH, C, N and soil moisture (71, 76). In studies where bacterial richness increased with the coarseness of soil (81, 82), it was hypothesized that the more highly fragmented water phase present in coarser soils provides more isolated microhabitats, which foster a higher richness of microorganisms. Moreover, compared with fine-textured soils, microhabitats are more heterogeneous because of the hydrologic disconnectedness. In contrast, other studies have found higher microbial richness and diversity in finer soils (44, 83), but it is important to note that since none of the studies above are designed as single variable experiments, their results may be biased by other soil properties.

Finer soil texture was also observed to increase the resilience of microbial communities during deforestation, as microbial communities did not change as much in finer soils as they did in coarse soils after reforestation (71). Nutrients and moisture in fine-textured soils are better protected from decomposition and leaching due to relatively high aggregation and absorption and low aeration common in these soil types. The highly reactive surface areas and porosity of silt and clay particles are better at buffering changes in soil conditions such as moisture, pH and nutrient (84). In fact, the susceptibility of bacterial biomass and richness and fungal biomass to deforestation was successfully predicted by soil texture (71). However, the role of soil texture in shaping microhabitats, influencing connectivity between niches, regulating interactions between microbes and substrates, and thus controlling microbial community composition are poorly understood.

2.3.5. Vegetation

Differences in vegetation play an important role in modulating variability in soil microbial communities (71, 85, 86). Vegetation types (forest and grassland) were reported to have consistent effects on biomass and richness of both bacterial and fungal community at a continental scale (71, 85). The dominant trees also influenced the activity of extracellular enzymes (86). But this influence is a complicated interaction of both abiotic and biotic factors. For example, microorganisms are directly associated with vegetation through symbioses including mutualism, parasitism and commensalism, but vegetation can also exert indirect control by altering nutrient supply and soil properties.

The differences in microbial community composition among different vegetation cover types are partly driven by plant inputs (71). The diversity of compounds in leaf litter, dead roots and root

exudate all influence the resource heterogeneity for a microbial community, and the amount of organic matter added to soils also differs among plant species (87-89). However, some work with detailed chemical analysis of the plant litter has suggested that their chemical composition were more similar than expected taxonomically (72). Studies on quantity and characteristics of organic matter input from litter, root exudate and dead roots in a single ecosystem and across ecosystems are required to fully understand these influences on microbial communities. In addition, vegetation may influence soil properties through modifying pH, moisture, soil texture and chemical forms of nutrients (90, 91).

Specific associations of microorganisms with the rhizosphere of different plant species are likely to be even stronger mediators of the vegetation effects than litter chemistry (86). However, fungal richness and diversity were not significantly correlated to tree species in a single biome (neotropical rain forests) (72), potentially indicating that fungal communities are more strongly correlated with plant communities, which are primarily determined by climate. Also, the strength of this effect can change depending on whether dominant trees are ectomycorrhizal or non-ectomycorrhizal (92), so the interactions between various ectomycorrhizal fungi and their host trees are important for a better understanding of the influence of vegetation.

Tree species also have a large influence on microbial communities during afforestation and deforestation processes within a single ecosystem, as the development of soils subjected to revegetation was found to be largely influenced by the dominant tree species via changes in soil organic matter (86). Microbial community shifts observed during afforestation and deforestation further reveal the significant influence of vegetation on underground communities. For example, microbial biomass and enzyme activities increased and microbial community composition

changed during the development of soils in spoil heaps succession after brown coal mining (93), but enzyme activity did not reach values observed in undisturbed forest soils after 25 years (94). It was also suggested that afforestation altered the microbial functional gene abundance, which may lead to changes in soil biogeochemical cycles (95). There are also studies suggesting that deforestation can drive microbial community responses through the reduction of soil C and N concentration (96, 97), resulting in reductions in bacterial biomass and increases in bacterial diversity (71). Previous studies suggests that Acidobacteria and Basidiomycota, respectively, had the largest shifts in bacterial and fungal community composition in response to deforestation and the relative abundance of Acidobacteria and Basidiomycetes were consistently low in deforested soil (71). It is hypothesized that these responses were driven by increased recalcitrant organic matter in late-stage successional forest plant input and the acidic property of forest soils, because Acidobacteria usually thrive with relatively low pH (64) and Basidiomycetes are the major players in decomposing lignocellulose-rich plant input (98, 99).

Vegetation effects on soil fungi are dependent on whether fungal taxa are ectomycorrhizal, arbuscular mycorrhizal, or saprotrophic fungi according to their function (71). Ectomycorrhizal fungi are possibly experiencing the strongest effects because their diversity is closely related to tree species. For example, ectomycorrhizal fungi colonizing the roots of temperate and boreal forest trees were absent in grassland soils, which contributed largely to the differences of fungal communities between forest and grassland soils (71). However, the catabolic response profiles, which estimate the community function, did not change significantly between vegetation types even though microbial catabolic function is correlated with taxonomic composition (40). This phenomenon addresses an ongoing important question in microbial ecology of whether the ecological functions of microbial communities change with taxonomic shifts, and given that

dormancy and functional redundancy are common among soil microorganisms, the extent to which microbial function can be explained by phylogenetic profiles is still controversial.

When all of these mechanisms of interactions between soil microorganisms, the environment, and plant species are examined together, it is clear that the regulation of succession by environmental factors is complicated by the diversity of microorganisms and the heterogeneity of the environment. Soil environments are rarely defined by any single factor, but by a combination of factors to which microorganisms must adapt simultaneously. The mechanisms of microbial succession are sure to be more complicated than the sum of individual influences by the factors discussed so far, and in addition microbial dormancy, physiological diversity, and rapid growth make succession patterns difficult to predict. Species with advantages in local habitats may become dominant and modify the microhabitat over short time frames, and the relatively rapid adaptation rates may alleviate some environmental stresses. However, if environmental stresses are persistent or large enough to alter community dynamics, they will still exert significant influences on driving microbial succession and shaping local microbial communities.

2.4. Techniques in microbial ecology

One important aspect of any study in microbial ecology is the limitation of the methods used to detect and quantify microorganisms in the system. Most soil microorganisms cannot be isolated and cultured in the laboratory, which has historically hampered studies on structure and function of soil microbial communities (100). The recent development of culture-independent techniques has increased our knowledge about soil microorganisms, but most previous work has still relied on DNA fingerprinting methods including denaturing gradient gel electrophoresis (DGGE) (101) and terminal restriction fragment length polymorphism (T-RFLP) (49), which can only quantify

community shifts with limited phylogenetic resolution. The recent adaptation of next-generation sequencing approaches to microbial ecology has provided unprecedented sampling depth, resolution, and phylogenetic information compared to traditional approaches. As such, microbial diversity assessed by sequencing of phylogenetically conserved marker genes provides a detailed overview of microbial phylogenetic profiles. In addition, the development of high-throughput sequencing finally makes it feasible to analyze microbial communities from a large number of samples with great detail (102).

The conserved marker gene analysis, especially 16S rRNA gene analysis, is commonly utilized for studying microbial community composition, but it is more difficult to use for estimation of microbial function because the metabolic characteristics of a large number of taxa are not well known and the relation between phylogenetic profiles and functional attributes is controversial. Previous studies revealed consistent correlation between microbial taxonomic dissimilarity and catabolic profile dissimilarity at a continental scale (71), which supports that the functional gene diversity is related to the taxonomic diversity of the microbial community to some extent (45, 80, 103, 104). However, it was also found that changes in biogeochemical processes are not consistent with soil microbial community structure change at a temporal scale (105), given that a major part of microorganisms in soils are dormant and inactive in different parts of a year (78, 105). Therefore, changes in the relative abundance of active microbial community members may contribute more to the temporal variability of soil ecological processes (105). For example, Actinobacteria are known to be persistent survivors in soils and they were found abundant in soils with 16S rRNA sequencing (106, 107), but if they are dormant they may not contribute much to biogeochemical cycling. In contrast, functional redundancy, which is the ability of two or more microbial taxa to perform a same process under similar conditions with similar rates,

may further obscure the relationships between taxonomic profiles and functions (108). If many taxa can perform the same biogeochemical function, species loss can occur without changes in function, at least to some degree.

Metagenomic methods, which are approaches studying the genomes of all microorganisms in a sample, have made it possible to estimate the functional attributes of microbial communities more directly. Shotgun metagenome sequencing provides information about the functional potential of microbial communities by randomly sequencing a proportion of all genes from soil microorganisms (in contrast to the target gene sequencing describe above). It has also been considered as the most accurate method for estimating taxonomic composition (109) because metagenomic analysis can characterize the phylogenetic and functional gene composition of samples and avoid bias from gene amplification.

Metagenome data can also provide insight to metabolic pathways in soils based on the abundance of genes responsible for specific reactions. For example, this approach has been used to study the prevalence of methanogenesis and denitrification in permafrost metagenomes (106) as well as carbon and nitrogen cycling in warmed soils in climate change simulation experiments, which were found to be community-wide and not attributable to specific taxa (110). Metagenomic analysis of an endophyte community colonizing rice roots revealed the involvement of bacterial endophytes in nitrogen cycling (111) and discovered several new enzymes including a thermostable lipase from Brazilian Atlantic Forest soil (112) and cellulolytic and hemicellulolytic enzymes from German grassland soil (113). Application of metagenomic analysis during ecosystem restoration, therefore, may provide important information on abundance changes of key functional genes, which can increase the

understanding of ecosystem function recovery. With the increase of sequencing capacities, we will continue to gain a more comprehensive understanding of how soil microbial community structure and function change with environmental factors.

References

1. Bradshaw A. 2000. The use of natural processes in reclamation — advantages and difficulties. Landscape and Urban Planning 51:89-100.
2. Šourková M, Frouz J, Fettweis U, Bens O, Hüttl RF, Šantrůčková H. 2005. Soil development and properties of microbial biomass succession in reclaimed post mining sites near Sokolov (Czech Republic) and near Cottbus (Germany). Geoderma 129:73-80.
3. Lewis D, Chauhan A, White J, Overholt W, Green S, Jasrotia P, Wafula D, Jagoe C. 2012. Microbial and Geochemical Assessment of Bauxitic Un-mined and Post-mined Chronosequence Soils from Mocho Mountains, Jamaica. Microbial Ecology 64:738-749.
4. Bradshaw A. 1997. Restoration of mined lands—using natural processes. Ecological Engineering 8:255-269.
5. Harris JA. 2003. Measurements of the soil microbial community for estimating the success of restoration. European Journal of Soil Science 54:801-808.
6. Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL, Murphy DV. 2011. Soil Microbial Community Successional Patterns during Forest Ecosystem Restoration. Appl Environ Microbiol 77:6158-6164.
7. Castro HF, Classen AT, Austin EE, Norby RJ, Schadt CW. 2010. Soil microbial community responses to multiple experimental climate change drivers. Appl Environ Microbiol 76:999-1007.
8. Quirino BF, Pappas GJ, Tagliaferro AC, Collevatti RG, Neto EL, da Silva MR, Bustamante MM, Kruger RH. 2009. Molecular phylogenetic diversity of bacteria associated with soil of the savanna-like Cerrado vegetation. Microbiol Res 164:59-70.
9. Tarlera S, Jangid K, Ivester AH, Whitman WB, Williams MA. 2008. Microbial community succession and bacterial diversity in soils during 77,000 years of ecosystem development. FEMS Microbiol Ecol 64:129-40.
10. Sait M, Davis KE, Janssen PH. 2006. Effect of pH on isolation and distribution of members of subdivision 1 of the phylum Acidobacteria occurring in soil. Appl Environ Microbiol 72:1852-7.
11. Hedlund K, Griffiths B, Christensen S, Scheu S, Setälä H, Tscharntke T, Verhoef H. 2004. Trophic interactions in changing landscapes: responses of soil food webs. Basic and Applied Ecology 5:495-503.

12. Nemergut DR, Anderson SP, Cleveland CC, Martin AP, Miller AE, Seimon A, Schmidt SK. 2007. Microbial community succession in an unvegetated, recently deglaciated soil. Microb Ecol 53:110-22.
13. Schutte UM, Abdo Z, Bent SJ, Williams CJ, Schneider GM, Solheim B, Forney LJ. 2009. Bacterial succession in a glacier foreland of the High Arctic. ISME J 3:1258-68.
14. Zumsteg A, Luster J, Göransson H, Smittenberg R, Brunner I, Bernasconi S, Zeyer J, Frey B. 2012. Bacterial, Archaeal and Fungal Succession in the Forefield of a Receding Glacier. Microb Ecol 63:552-564.
15. Stibal M, Tranter M, Benning LG, Řehák J. 2008. Microbial primary production on an Arctic glacier is insignificant in comparison with allochthonous organic carbon input. Environmental Microbiology 10:2172-2178.
16. Frey B, Rieder SR, Brunner I, Plötze M, Koetzsch S, Lapanje A, Brandl H, Furrer G. 2010. Weathering-Associated Bacteria from the Damma Glacier Forefield: Physiological Capabilities and Impact on Granite Dissolution. Applied and Environmental Microbiology 76:4788-4796.
17. Crews T, Kurina L, Vitousek P. 2001. Organic matter and nitrogen accumulation and nitrogen fixation during early ecosystem development in Hawaii. Biogeochemistry 52:259-279.
18. Schmidt SK, Reed SC, Nemergut DR, Grandy AS, Cleveland CC, Weintraub MN, Hill AW, Costello EK, Meyer AF, Neff JC, Martin AM. 2008. The earliest stages of ecosystem succession in high-elevation (5000 metres above sea level), recently deglaciated soils. Proc Biol Sci 275:2793-802.
19. Freeman KR, Pescador MY, Reed SC, Costello EK, Robeson MS, Schmidt SK. 2009. Soil CO2 flux and photoautotrophic community composition in high-elevation, 'barren' soil. Environmental Microbiology 11:674-686.
20. Nicol GW, Tscherko D, Embley TM, Prosser JI. 2005. Primary succession of soil Crenarchaeota across a receding glacier foreland. Environ Microbiol 7:337-47.
21. Ayton J, Aislabie J, Barker GM, Saul D, Turner S. 2010. Crenarchaeota affiliated with group 1.1b are prevalent in coastal mineral soils of the Ross Sea region of Antarctica. Environmental Microbiology 12:689-703.
22. Pesaro M, Widmer F, Nicollier G, Zeyer J. 2003. Effects of freeze–thaw stress during soil storage on microbial communities and methidathion degradation. Soil Biology and Biochemistry 35:1049-1061.
23. Gleeson DB, Clipson N, Melville K, Gadd GM, McDermott FP. 2005. Characterization of fungal community structure on a weathered pegmatitic granite. Microb Ecol 50:360-8.
24. Jumpponen A. 2003. Soil fungal community assembly in a primary successional glacier forefront ecosystem as inferred from rDNA sequence analyses. New Phytologist 158:569-578.
25. Mühlmann O, Peintner U. 2008. Ectomycorrhiza of Kobresia myosuroides at a primary successional glacier forefront. Mycorrhiza 18:355-362.
26. Ludley KE, Robinson CH. 2008. 'Decomposer'basidiomycota in Arctic and Antarctic ecosystems. Soil Biology and Biochemistry 40:11-29.
27. Tscherko D, Hammesfahr U, Marx M-C, Kandeler E. 2004. Shifts in rhizosphere microbial communities and enzyme activity of Poa alpina across an alpine chronosequence. Soil Biology and Biochemistry 36:1685-1698.

28. Edwards IP, Bürgmann H, Miniaci C, Zeyer J. 2006. Variation in Microbial Community Composition and Culturability in the Rhizosphere of Leucanthemopsis alpina (L.) Heywood and Adjacent Bare Soil Along an Alpine Chronosequence. Microbial Ecology 52:679-692.
29. Knelman JE, Legg TM, O'Neill SP, Washenberger CL, González A, Cleveland CC, Nemergut DR. 2012. Bacterial community structure and function change in association with colonizer plants during early primary succession in a glacier forefield. Soil Biology and Biochemistry 46:172-180.
30. Yergeau E, Bokhorst S, Kang S, Zhou J, Greer CW, Aerts R, Kowalchuk GA. 2012. Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. ISME J 6:692-702.
31. Bardgett RD, Walker LR. 2004. Impact of coloniser plant species on the development of decomposer microbial communities following deglaciation. Soil Biology and Biochemistry 36:555-559.
32. Tscherko D, Hammesfahr U, Zeltner G, Kandeler E, Böcker R. 2005. Plant succession and rhizosphere microbial communities in a recently deglaciated alpine terrain. Basic and Applied Ecology 6:367-383.
33. Ohtonen R, Fritze H, Pennanen T, Jumpponen A, Trappe J. 1999. Ecosystem properties and microbial community changes in primary succession on a glacier forefront. Oecologia 119:239-246.
34. Buckley DH, Schmidt TM. 2001. The Structure of Microbial Communities in Soil and the Lasting Impact of Cultivation. Microb Ecol 42:11-21.
35. Ganzert L, Lipski A, Hubberten HW, Wagner D. 2011. The impact of different soil parameters on the community structure of dominant bacteria from nine different soils located on Livingston Island, South Shetland Archipelago, Antarctica. FEMS Microbiol Ecol 76:476-91.
36. Liebner S, Harder J, Wagner D. 2008. Bacterial diversity and community structure in polygonal tundra soils from Samoylov Island, Lena Delta, Siberia. Int Microbiol 11:195-202.
37. Bajerski F, Wagner D. 2013. Bacterial succession in Antarctic soils of two glacier forefields on Larsemann Hills, East Antarctica. FEMS Microbiol Ecol 85:128-42.
38. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J. 2009. Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol 75:2046-2056.
39. Kuramae EE, Gamper HA, Yergeau E, Piceno YM, Brodie EL, Desantis TZ, Andersen GL, van Veen JA, Kowalchuk GA. 2010. Microbial secondary succession in a chronosequence of chalk grasslands. ISME J 4:711-5.
40. Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, Daroub SH, Camargo FA, Farmerie WG, Triplett EW. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283-90.
41. Fulthorpe RR, Roesch LF, Riva A, Triplett EW. 2008. Distantly sampled soils carry few species in common. ISME J 2:901-10.
42. Lozano YM, Hortal S, Armas C, Pugnaire FI. 2014. Interactions among soil, plants, and microorganisms drive secondary succession in a dry environment. Soil Biology and Biochemistry 78:298-306.

43. Hortal S, Bastida F, Armas C, Lozano YM, Moreno JL, García C, Pugnaire FI. 2013. Soil microbial community under a nurse-plant species changes in composition, biomass and activity as the nurse grows. Soil Biology and Biochemistry 64:139-146.
44. Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E. 2001. Microbial Population Structures in Soil Particle Size Fractions of a Long-Term Fertilizer Field Experiment. Applied and Environmental Microbiology 67:4215-4224.
45. Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 4:1340-51.
46. Lauber CL, Hamady M, Knight R, Fierer N. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75:5111-20.
47. Lauber CL, Strickland MS, Bradford MA, Fierer N. 2008. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biology and Biochemistry 40:2407-2415.
48. Rousk J, Brookes PC, Bååth E. 2010. Investigating the mechanisms for the opposing pH relationships of fungal and bacterial growth in soil. Soil Biology and Biochemistry 42:926-934.
49. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 103:626-31.
50. Dimitriu PA, Grayston SJ. 2010. Relationship between soil properties and patterns of bacterial beta-diversity across reclaimed and natural boreal forest soils. Microb Ecol 59:563-73.
51. Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J 3:442-453.
52. Wheeler KA, Hurdman BF, Pitt JI. 1991. Influence of pH on the growth of some toxigenic species of Aspergillus, Penicillium and Fusarium. International Journal of Food Microbiology 12:141-149.
53. Nevarez L, Vasseur V, Le Madec A, Le Bras MA, Coroller L, Leguerinel I, Barbier G. 2009. Physiological traits of Penicillium glabrum strain LCP 08.5568, a filamentous fungus isolated from bottled aromatized mineral water. Int J Food Microbiol 130:166-71.
54. Rousk J, Demoling LA, Bahr A, Baath E. 2008. Examining the fungal and bacterial niche overlap using selective inhibitors in soil. FEMS Microbiol Ecol 63:350-8.
55. Fierer N, Nemergut D, Knight R, Craine JM. 2010. Changes through time: integrating microorganisms into the study of succession. Res Microbiol 161:635-42.
56. Knelman JE, Schmidt SK, Lynch RC, Darcy JL, Castle SC, Cleveland CC, Nemergut DR. 2014. Nutrient addition dramatically accelerates microbial community succession. PLoS One 9:e102609.
57. Coyne MS, Zhai Q, MacKown CT, Barnhisel RI. 1998. Gross nitrogen transformation rates in soil at a surface coal mine site reclaimed for prime farmland use. Soil Biology and Biochemistry 30:1099-1106.
58. Eilers KG, Lauber CL, Knight R, Fierer N. 2010. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. Soil Biology and Biochemistry 42:896-903.

59. Chow ML, Radomski CC, McDermott JM, Davies J, Axelrood PE. 2002. Molecular characterization of bacterial diversity in Lodgepole pine (Pinus contorta) rhizosphere soils from British Columbia forest soils differing in disturbance and geographic source. FEMS Microbiol Ecol 42:347-57.
60. Cleveland C, Nemergut D, Schmidt S, Townsend A. 2007. Increases in soil respiration following labile carbon additions linked to rapid shifts in soil microbial community composition. Biogeochemistry 82:229-240.
61. Fierer N, Bradford MA, Jackson RB. 2007. Toward an ecological classification of soil bacteria. Ecology 88:1354-64.
62. Kastovska K, Elster J, Stibal M, Santruckova H. 2005. Microbial assemblages in soil microbial succession after glacial retreat in Svalbard (high arctic). Microb Ecol 50:396-407.
63. Gomez-Alvarez V, King GM, Nusslein K. 2007. Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. FEMS Microbiol Ecol 60:60-73.
64. Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R. 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J 6:1007-17.
65. Janssens IA, Dieleman W, Luyssaert S, Subke JA, Reichstein M, Ceulemans R, Ciais P, Dolman AJ, Grace J, Matteucci G, Papale D, Piao SL, Schulze ED, Tang J, Law BE. 2010. Reduction of forest soil respiration in response to nitrogen deposition. Nature Geosci 3:315-322.
66. Liu L, Greaver TL. 2010. A global perspective on belowground carbon dynamics under nitrogen enrichment. Ecol Lett 13:819-28.
67. Treseder KK. 2008. Nitrogen additions and microbial biomass: a meta-analysis of ecosystem studies. Ecol Lett 11:1111-20.
68. Zumsteg A, Bååth E, Stierli B, Zeyer J, Frey B. 2013. Bacterial and fungal community responses to reciprocal soil transfer along a temperature and soil moisture gradient in a glacier forefield. Soil Biology and Biochemistry 61:121-132.
69. Manzoni S, Schimel JP, Porporato A. 2012. Responses of soil microbial communities to water stress: results from a meta-analysis. Ecology 93:930-8.
70. Serna-Chavez HM, Fierer N, van Bodegom PM. 2013. Global drivers and patterns of microbial abundance in soil. Global Ecology and Biogeography 22:1162-1172.
71. Crowther TW, Maynard DS, Leff JW, Oldfield EE, McCulley RL, Fierer N, Bradford MA. 2014. Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study. Glob Chang Biol 20:2983-94.
72. McGuire KL, Fierer N, Bateman C, Treseder KK, Turner BL. 2012. Fungal community composition in neotropical rain forests: the influence of tree diversity and precipitation. Microb Ecol 63:804-12.
73. Vandenkoornhuyse P, Ridgway KP, Watson IJ, Fitter AH, Young JPW. 2003. Co-existing grass species have distinctive arbuscular mycorrhizal communities. Mol Ecol 12:3085-3095.
74. Peay KG, Baraloto C, Fine PV. 2013. Strong coupling of plant and fungal community structure across western Amazonian rainforests. ISME J 7:1852-1861.
75. Hazard C, Gosling P, van der Gast CJ, Mitchell DT, Doohan FM, Bending GD. 2013. The role of local environment and geographical distance in determining community

composition of arbuscular mycorrhizal fungi at the landscape scale. ISME J 7:498-508.
76. Rousk J, Frey SD, Bååth E. 2012. Temperature adaptation of bacterial communities in experimentally warmed forest soils. Global Change Biology 18:3252-3258.
77. Long T, Or D. 2009. Dynamics of microbial growth and coexistence on variably saturated rough surfaces. Microb Ecol 58:262-75.
78. Lauber CL, Ramirez KS, Aanderud Z, Lennon J, Fierer N. 2013. Temporal variability in soil microbial communities across land-use types. ISME J 7:1641-50.
79. Lipson DA. 2007. Relationships between temperature responses and bacterial community structure along seasonal and altitudinal gradients. FEMS Microbiology Ecology 59:418-427.
80. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 109:21390-5.
81. Chau JF, Bagtzoglou AC, Willig MR. 2011. The Effect of Soil Texture on Richness and Diversity of Bacterial Communities. Environmental Forensics 12:333-341.
82. Carson JK, Gonzalez-Quinones V, Murphy DV, Hinz C, Shaw JA, Gleeson DB. 2010. Low pore connectivity increases bacterial diversity in soil. Appl Environ Microbiol 76:3936-42.
83. Zhang P, Zheng J, Pan G, Zhang X, Li L, Rolf T. 2007. Changes in microbial community structure and function within particle size fractions of a paddy soil under different long-term fertilization treatments from the Tai Lake region, China. Colloids and Surfaces B: Biointerfaces 58:264-270.
84. Grandy AS, Strickland MS, Lauber CL, Bradford MA, Fierer N. 2009. The influence of microbial communities, management, and soil texture on soil organic matter chemistry. Geoderma 150:278-286.
85. Schimel JP, Schaeffer SM. 2012. Microbial control over carbon cycling in soil. Front Microbiol 3:348.
86. Šnajdr J, Dobiášová P, Urbanová M, Petránková M, Cajthaml T, Frouz J, Baldrian P. 2013. Dominant trees affect microbial community composition and activity in post-mining afforested soils. Soil Biology and Biochemistry 56:105-115.
87. Finér L, Helmisaari HS, Lõhmus K, Majdi H, Brunner I, Børja I, Eldhuset T, Godbold D, Grebenc T, Konôpka B, Kraigher H, Möttönen MR, Ohashi M, Oleksyn J, Ostonen I, Uri V, Vanguelova E. 2007. Variation in fine root biomass of three European tree species: Beech (Fagus sylvatica L.), Norway spruce (Picea abies L. Karst.), and Scots pine (Pinus sylvestris L.). Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 141:394-405.
88. Hooper DU, Bignell DE, Brown VK, Brussard L, Mark Dangerfield J, Wall DH, Wardle DA, Coleman DC, Giller KE, Lavelle P, Van Der Putten WH, De Ruiter PC, Rusek J, Silver WL, Tiedje JM, Wolters V. 2000. Interactions between Aboveground and Belowground Biodiversity in Terrestrial Ecosystems: Patterns, Mechanisms, and Feedbacks. BioScience 50:1049-1061.
89. Ekblad ALF, Huss-Danell K. 1995. Nitrogen fixation by Alnus incana and nitrogen transfer from A. incana to Pinus sylvestris influenced by macronutrients and ectomycorrhiza. New Phytologist 131:453-459.

90. Frouz J, Pižl V, Cienciala E, Kalčík J. 2009. Carbon storage in post-mining forest soil, the role of tree biomass and soil bioturbation. Biogeochemistry 94:111-121.
91. Iovieno P, Alfani A, Bååth E. 2010. Soil microbial community structure and biomass as affected by Pinus pinea plantation in two Mediterranean areas. Applied Soil Ecology 45:56-63.
92. McGuire KL, Allison SD, Fierer N, Treseder KK. 2013. Ectomycorrhizal-Dominated Boreal and Tropical Forests Have Distinct Fungal Communities, but Analogous Spatial Patterns across Soil Horizons. PLoS ONE 8:e68278.
93. Baldrian P, Trögl J, Frouz J, Šnajdr J, Valášková V, Merhautová V, Cajthaml T, Herinková J. 2008. Enzyme activities and microbial biomass in topsoil layer during spontaneous succession in spoil heaps after brown coal mining. Soil Biology and Biochemistry 40:2107-2115.
94. Štursová M, Baldrian P. 2011. Effects of soil properties and management on the activity of soil organic matter transforming enzymes and the quantification of soil-bound and free activity. Plant and Soil 338:99-110.
95. Berthrong ST, Schadt CW, Piñeiro G, Jackson RB. 2009. Afforestation Alters the Composition of Functional Genes in Soil and Biogeochemical Processes in South American Grasslands. Applied and Environmental Microbiology 75:6240-6248.
96. da C Jesus E, Marsh TL, Tiedje JM, de S Moreira FM. 2009. Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J 3:1222-1222.
97. Rodrigues JL, Pellizari VH, Mueller R, Baek K, Jesus Eda C, Paula FS, Mirza B, Hamaoui GS, Jr., Tsai SM, Feigl B, Tiedje JM, Bohannan BJ, Nusslein K. 2013. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proc Natl Acad Sci U S A 110:988-93.
98. Hättenschwiler S, Tiunov AV, Scheu S. 2005. BIODIVERSITY AND LITTER DECOMPOSITION IN TERRESTRIAL ECOSYSTEMS. Annual Review of Ecology, Evolution, and Systematics 36:191-218.
99. Crowther TW, Stanton DW, Thomas SM, A'Bear AD, Hiscox J, Jones TH, Voriskova J, Baldrian P, Boddy L. 2013. Top-down control of soil fungal community composition by a globally distributed keystone consumer. Ecology 94:2518-28.
100. Vartoukian SR, Palmer RM, Wade WG. 2010. Strategies for culture of 'unculturable' bacteria. FEMS Microbiol Lett 309:1-7.
101. Muyzer G, de Waal EC, Uitterlinden AG. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol 59:695-700.
102. Knight R, Jansson J, Field D, Fierer N, Desai N, Fuhrman JA, Hugenholtz P, van der Lelie D, Meyer F, Stevens R, Bailey MJ, Gordon JI, Kowalchuk GA, Gilbert JA. 2012. Unlocking the potential of metagenomics through replicated experimental design. Nat Biotech 30:513-520.
103. Gilbert JA, Field D, Swift P, Thomas S, Cummings D, Temperton B, Weynberg K, Huse S, Hughes M, Joint I, Somerfield PJ, Mühling M. 2010. The Taxonomic and Functional Diversity of Microbes at a Temperate Coastal Site: A 'Multi-Omic' Study of Seasonal and Diel Temporal Variation. PLoS ONE 5:e15545.

104. Bryant JA, Stewart FJ, Eppley JM, DeLong EF. 2012. Microbial community phylogenetic and trait diversity declines with depth in a marine oxygen minimum zone. Ecology 93:1659-1673.
105. Baldrian P, Kolarik M, Stursova M, Kopecky J, Valaskova V, Vetrovsky T, Zifcakova L, Snajdr J, Ridl J, Vlcek C, Voriskova J. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J 6:248-58.
106. Mackelprang R, Waldrop MP, DeAngelis KM, David MM, Chavarria KL, Blazewicz SJ, Rubin EM, Jansson JK. 2011. Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. Nature 480:368-371.
107. Sorek R, Cossart P. 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet 11:9-16.
108. Allison SD, Martiny JB. 2008. Colloquium paper: resistance, resilience, and redundancy in microbial communities. Proc Natl Acad Sci U S A 105 Suppl 1:11512-9.
109. von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N, Bork P. 2007. Quantitative phylogenetic assessment of microbial communities in diverse environments. Science 315:1126-30.
110. Luo C, Rodriguez-R LM, Johnston ER, Wu L, Cheng L, Xue K, Tu Q, Deng Y, He Z, Shi JZ, Yuan MM, Sherry RA, Li D, Luo Y, Schuur EAG, Chain P, Tiedje JM, Zhou J, Konstantinidis KT. 2014. Soil Microbial Community Responses to a Decade of Warming as Revealed by Comparative Metagenomics. Applied and Environmental Microbiology 80:1777-1786.
111. Sessitsch A, Hardoim P, Döring J, Weilharter A, Krause A, Woyke T, Mitter B, Hauberg-Lotte L, Friedrich F, Rahalkar M, Hurek T, Sarkar A, Bodrossy L, van Overbeek L, Brar D, van Elsas JD, Reinhold-Hurek B. 2011. Functional Characteristics of an Endophyte Community Colonizing Rice Roots as Revealed by Metagenomic Analysis. Molecular Plant-Microbe Interactions 25:28-36.
112. Faoro H, Glogauer A, Couto GH, de Souza EM, Rigo LU, Cruz LM, Monteiro RA, de Oliveira Pedrosa F. 2012. Characterization of a new Acidobacteria-derived moderately thermostable lipase from a Brazilian Atlantic Forest soil metagenome. FEMS Microbiology Ecology 81:386-394.
113. Nacke H, Engelhaupt M, Brady S, Fischer C, Tautzt J, Daniel R. 2012. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnology Letters 34:663-675.

Chapter 3. Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration.

This chapter was published in Applied and Environmental Microbiology in 2017.

Citation: Sun, S., Li, S., Avera, B. N., Strahm, B. D., & Badgley, B. D. (2017). Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration. *Applied and Environmental Microbiology*, *83*(14), e00966-17.

Attribution

The published article was co-authored by Dr. Song Li, Bethany Avera, Dr. Brian Strahm and Dr. Brian Badgley.

Dr. Brian Badgley is the principal investigator of this project and helped in experiment design and preparation and revision of manuscript for publication.

Dr. Song Li provided guidance in the network analysis and assisted in revision of the manuscript.

Bethany Avera and Dr. Brian Strahm assisted in establishment of research sites, collection and analysis of environmental data, and soil sampling.

Abstract

Bacteria and fungi are important mediators of biogeochemical processes and play essential roles in the establishment of plant communities, which makes knowledge about their recovery after extreme disturbances valuable for understanding ecosystem development. However, broad ecological differences between bacterial and fungal organisms, such as growth rates, stress tolerance, and substrate utilization, suggest they could follow distinct trajectories and show contrasting dynamics during recovery. In this study, we analyzed both the intra-annual variability and decadal scale recovery of bacterial and fungal communities in a chronosequence of reclaimed mine soils using next-generation sequencing to quantify their abundance, richness, β-diversity, taxonomic composition and co-occurrence network properties. Bacterial communities shifted gradually with overlapping β-diversity patterns across chronosequence ages, while shifts in fungal communities were more distinct among different ages. In addition, the magnitude of intra-annual variability in bacterial β-diversity was comparable to the changes across decades of chronosequence age, while fungal communities changed minimally across months. Finally, the complexity of bacterial co-occurrence networks increased with chronosequence age, while fungal networks did not show clear age-related trends. We hypothesize that these contrasting dynamics of bacteria and fungi in the chronosequence result from: (1) faster growth rates for bacteria leading to higher intra-annual variability; (2) higher tolerance to environmental changes for fungi; and (3) stronger influence of vegetation on fungal communities.

3.1 Introduction

Ecologists have been studying the succession dynamics of plant communities for more than a century (1, 2) and the patterns of changes in plant community composition are used as indicators

for the restoration of ecosystems (3-5). However, patterns and mechanisms of soil microbiota recovery after disturbance has been much less intensively studied until the recent rise of next generation sequencing techniques. Similar to plants, microbial succession is likely to be regulated by resource limitation, environmental stress, biological interactions, and historical effects (6-9). However, key biological differences between plants and microorganisms such as phylogenetic diversity, metabolic versatility, and dormancy are likely to cause some disparities in mechanisms driving recovery of the soil microbiota (10, 11). The rapid responses of microorganisms to alterations in environmental conditions and their high turnover rates may provide additional information and possibly an early indication of the restoration progress following ecosystem disturbance (12).

Both bacteria and fungi are significant components of soil microbiota and perform critical ecological functions, but they also have many differences in phenotype, phylogeny, and life history that could cause different succession patterns during recovery. For example, typical soil fungal growth rates appear to be about tenfold slower than soil bacteria, and fungi tend to be more tolerant of low temperatures and dry soils (13, 14). From a simplified perspective, fungi are considered to be dominant degraders of refractory organic matter and mediators of slower carbon cycling, whereas bacteria are typically considered more important regulators of the fast carbon cycling pathways in soil (15). While there are exceptions to these characterizations, at a broad scale they highlight the potential for functional differences in how each group responds to ecosystem restoration. Furthermore, while both bacteria and fungi form symbiotic relationships with plants, fungal symbionts are more common, with occurrence in over 85% of all angiosperms (16), and tend to have more specific relationships with plant species (17-20). Fungi

may also be more dispersal limited than bacteria due to their larger sizes (21), and priority effects can have greater influence on fungal communities (22).

The analysis of ecosystem recovery after disturbance is challenging because it requires observation over long time periods. Chronosequences (i.e., space for time substitutions) are an established approach to this challenge and have been used to study soil microbial succession dynamics spanning hundreds of years (23-25). However, reported patterns in chronosequences are not easily generalized. For example, microbial diversity has not shown consistent patterns in glacial chronosequences (26-29), and abundance of bacterial phyla showed different trends along chronosequence ages in two glacier forelands within the same study (30). These differences have been attributed to heterogeneity of physical landforms, soil structure, environmental conditions, and vegetation across those gradients (31). Compared with natural chronosequences, reforested lands reclaimed from mining operations represent an opportunity to study replicated sites with similar starting conditions. Surface mining removes all overlying soil and rock, which are then replaced during reclamation and reforested to accelerate ecosystem recovery (32). The use of specific controlled reforestation practices (32, 33) aids in creating replicated sites over time with relatively well controlled ecosystem development in terms of soil properties, landscapes, and vegetation (34).

In this study, we characterized both the intra-annual variability and long-term recovery in bacterial and fungal components of soil microbiota in a chronosequence of reclaimed mine lands that spans 5 to 30 years post-reforestation and includes unmined reference plots. We used next-generation sequencing to describe shifts in diversity and taxonomic composition and constructed co-occurrence networks to reveal potential interaction patterns. Specific research objectives were

to determine: (a) how bacterial and fungal communities changed during recovery in terms of abundance, richness and taxonomic composition; (b) how the bacterial and fungal co-occurrence network properties changed throughout the recovery process; (c) how the observed intra- and interannual dynamics differed between the bacterial and fungal communities.

3.2 Materials and methods

3.2.1 Study site and sample collection

The reforested mined sites are located in Wise County, Virginia, USA in the central Appalachian Coal Basin (35). The samples for this study were collected in collaboration with work done by Avera et al. (36) to look at ecosystem recovery at the same sites. Details of the sampling sites (climate, aspect, slope, vegetation, etc.) and plot design are described in the "Soil dynamics" section of that publication. Briefly, four reclaimed sites were selected to establish a chronosequence of 5, 11, 21, and 30 years (yr) post reforestation, and an unmined site in a natural hardwood forest nearby was used as a reference (Figure S1). Triplicate study plots were randomly located within each site. Three soil samples were taken from 0-10cm depth in each plot, pooled, and homogenized, to more fully capture the diversity at the plot level, while using the replicated plots to characterize variation in soil microbiota at each site. Soil samples were collected in the third or fourth week of May, July, September and October 2013 to estimate intra-annual variation of microbial communities. All samples were transported to the lab on ice, sieved to 4 mm to remove coarse roots and stones, and stored at -80 ^{o}C until DNA extraction. As part of their research, Avera et al. (36) measured other ecosystem properties and some of their results are further analyzed here to aid in the understanding of the factors influencing microbiota recovery. These include soil properties such as temperature (Temp), moisture, available NH_4^+

and NO_3^-; gas fluxes of N_2O, CO_2 and CH_4; and total microbial biomass (TMB) and substrate-induced respiration (SIR).

3.2.2 DNA extraction, amplification, and sequencing

Total DNA was extracted from approximately 0.25 g soil using the MoBio PowerSoil Soil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Quantitative polymerase chain reaction (qPCR) was performed to estimate the abundance of bacteria and fungi in soil samples. Bacteria were targeted with the Eub338 (5'-ACT CCT ACG GGA GGC AGC AG-3') and Eub518 (5'-ATT ACC GCG GCT GCT GG-3') primer pair, and fungi were targeted with ITS1f (5'-TCC GTA GGT GAA CCT GCG G-3') and 5.8s (5'- CGC TGC GTT CTT CAT CG-3') primer pair (37). Each 25 µL reaction contained 12.5 µL iTaq^{TM} Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 5 µL PCR water (MoBio Laboratories, Carlsbad, CA, USA), 1.25 µL of each of the primers (10 µM), and 5 µL of isolated DNA as template. Conditions were 15 min at 95°C, followed by 40 cycles of 95°C for 1 min, 53°C for 30 s and 72°C for 1 min. Standards were made from 10-fold dilution series of plasmids containing cloned target sequences. Standard curves for all qPCR runs had efficiency values ranging from 0.97 to 1.07 and R^2 values >0.99.

The V4 region of the bacterial 16S rRNA gene and the fungal ITS1 region were amplified by PCR to characterize bacterial and fungal communities, respectively. The PCR reaction mixture (25 µL) contained 10 µL 2.5X 5 Prime HotMaster mix (5 Prime, Gaithersburg, MD), 13 µL PCR water (MoBio Laboratories, Carlsbad, CA, USA), 0.5 µL of each of the primers (10 µM), and 1 µL of isolated DNA as template. The V4 region was amplified with the 515F and 806R primer pair (38), and the ITS1 region was amplified with ITS1F and ITS2 primer pair (39). Primers were designed for Illumina sequencing with adapters, primer pads and 2-bp linker sequences,

and the reverse primer contained a 12-bp barcode sequence unique to each sample (38, 39). The thermal cycling protocol was: 94 °C for 5 min, 35 cycles of 94°C for 45 s, 50 °C for 45 s and 72°C for 90 s, followed by a final extension period at 72 °C for 10 min. All samples were amplified in triplicate on thermal cyclers (Bio-Rad, Hercules, CA, USA), pooled and visualized on agarose gels. PCR products were purified using the UltraClean PCR cleanup kit (MoBio Laboratories, Carlsbad, CA, USA). Amplicon concentrations were quantified using Qubit 2.0 fluorometer (Invitrogen, USA) according to the manufacturer's protocol and amplicons from all samples were combined in equimolar ratios. Bacterial and fungal amplicons were sequenced separately on the Illumina Miseq platform using 250 bp and paired ends. The resulting raw reads were deposited in the NCBI BioProject database (accession number PRJNA324696). Sequencing reads were quality-filtered and processed using the USEARCH pipeline (40). In brief, the forward and reverse reads were merged with at least 150 bp overlap. The merged reads were filtered with a minimum length of 200 bp and expected errors per read lower than 0.5, which are calculated from quality scores. Chimeric sequences were identified with UCHIME and removed (41). SILVA and UNITE databases were used as references for bacteria and fungi, respectively (42, 43), and taxonomy was assigned using the RDP classifier (44). The Chao1 index was generated with QIIME (Quantitative Insights Into Microbial Ecology) to estimate the richness of bacterial and fungal communities (45).

3.2.3 Statistical and co-occurrence network analyses

The R program (46) was primarily used for statistical analyses and plotting. Significant differences in bacterial and fungal abundance and richness were detected across chronosequence ages or sampling months using ANOVA and Tukey's HSD tests. Matrices of pairwise similarities for all samples were calculated independently for bacterial and fungal communities

using the Bray-Curtis distance metric. Patterns of similarity among samples were visualized with non-metric multidimensional scaling (NMDS) using the 'metaMDS' function in the R vegan package (47), and environmental variables (Temp, moisture, NH_4^+, NO_3^-, N_2O, CO_2, CH_4), TMB and SIR (36) were added to the ordination with the 'envfit' function in the same package. The ANOSIM function in vegan was used to test the statistical significance of all pairs of samples with 999 permutations. The multivariate dispersion of all months and chronosequence groups was calculated with function 'betadisper'. Microbial OTUs that showed significant differences in abundance across sampling months or chronosequence ages were selected using the edgeR program with a false discovery rate (FDR) <0.05 and count per million (CPM) $>2^6$ in order to choose highly abundant and significantly variable OTUs for downstream analysis (48). After filtering, 2,480 bacterial OTUs and 1,340 fungal OTUs remained for further analysis.

Co-occurrence analyses have recently been increasingly used in microbial ecology to better understand community structure, characterize potential intra-community interactions, and identify potentially shared and contrasting niches (49, 50). The co-occurrence network analysis in this study was performed with the 'igraph' R package (51). The 500 most abundant OTUs (accounting for 61.6% of bacterial sequences and 48.0% of fungal sequences) per chronosequence age were used to build individual networks for each age. The cutoff value of 500 was based upon a similar approach used by Dini-Andreote et al. (23), but we also constructed networks using the most abundant 800 and 1,000 OTUs to verify that the interpretation of the general trends of network properties did not change. Spearman Rank Correlations (SRC) were calculated for each pair of OTUs in the same chronosequence age. The edges in the networks represent statistically significant (FDR<0.05) correlations with an absolute SRC value >0.9. The nodes in the networks represent individual OTUs with at least one significant correlation with

another OTU. The numbers of nodes and edges, average degree, clustering coefficient, and closeness centrality were calculated using the 'igraph' R package. Degree is the number of edges of each node, representing how many other nodes (OTUs) in the network are connected with the given node. Clustering coefficient is a measure of how often a node's neighbors are also connected (52). Closeness centrality measures the inverse distance to all other nodes from one node (53). The overlapping nodes and edges of networks from two consecutive stages were identified, and the significance of the difference between the number of actual overlaps and randomized overlaps were analyzed with Fisher's exact test and permutation methods (999 permutations). The abundance, degree and closeness centrality of nodes in overlapping edges were compared with the average of the whole network with Welch two sample t-tests.

3.3 Results

3.3.1 Sequencing results

The processing of raw sequencing reads resulted in 16.8 million high quality V4 sequences averaging about 280,000 sequences per sample and 5.1 million high quality ITS1 sequences averaging about 86,000 sequences per sample. The rarefaction curves are shown in Figure S2. Of the initial 60 samples for ITS1 sequencing, 59 samples were used for downstream analyses. Sequences were then clustered to 17, 017 bacterial operational taxonomic units (OTUs) and 5,979 fungal OTUs with a 97% identity threshold (54). For all downstream analyses, 97,000 reads were randomly selected for bacteria, and 28,000 reads were randomly selected for fungi per sample to correct for differences in sequencing depth.

3.3.2 Abundance and richness of microbial communities

Abundances of bacterial 16S rRNA copies decreased significantly from May and July to September and October, while abundances of fungal ITS copies increased significantly over the same period (*P*-value<0.05, Tukey's honestly significant difference [HSD] test) (Figure 3.1e and g). However, along the chronosequence age, there were no obvious patterns in abundance of 16S rRNA or ITS copy numbers, except that bacterial abundance was significantly lower at 5 yr (*P*-value<0.05, Tukey's HSD) (Figure 3.1f and h). Overall, the monthly shifts in bacterial and fungal abundance were as great or greater than the shifts across chronosequence ages. Bacterial richness, estimated by the Chao1 index, was significantly higher in May and July than September but not October, while fungal community richness was significantly higher in May than September only (*P*-value<0.05, Tukey's HSD) (Figure 3.1a and c). Both bacterial and fungal richness increased generally with chronosequence age up to 30 yr, but were lower in the unmined sites (Figure 3.1b and d). There were no significant interactions between chronosequence age and sampling month for either richness or abundance of either bacteria or fungi.

3.3.3 β-diversity and taxonomic composition

The patterns of bacterial and fungal β-diversity were visualized with NMDS plots and tested for significant differences using ANOSIM. Bacterial β-diversity varied significantly within the study year, with May and July harboring significantly different communities from September and October, regardless of chronosequence age (*P*-value=0.001, ANOSIM with 999 permutations) (Figure 3.2a and Table S1a). The differences between May and July or between September and October were not significant (*P*-values range from 0.437 to 0.846, ANOSIM with 999 permutations) (Table S1a). All pairwise comparisons of bacterial community structure among

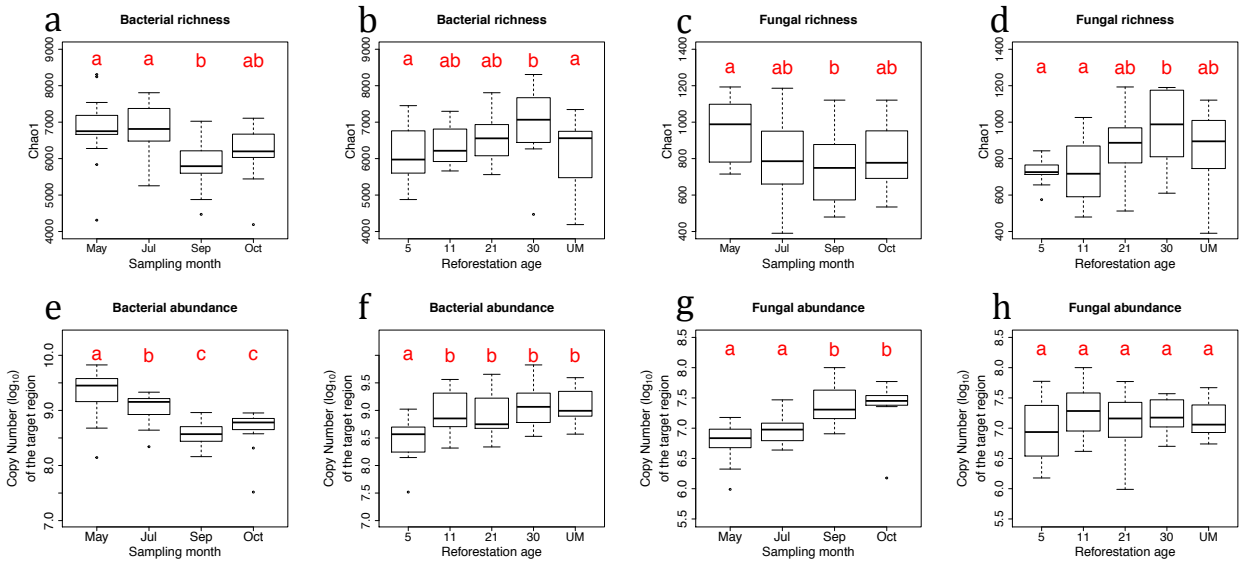


Figure 3.1. (a to d) Microbial richness was calculated using the Chao1 index. (e to h) Microbial abundance is shown as the copy numbers of the target regions, which were estimated by qPCR and normalized to gram of soil. The upper and lower edges of box represent the 75th and 25th percentiles, respectively. The upper and lower whiskers represent the maximum and minimum values with outliers excluded (> 1.5X interquartile range). Letters above the boxplots represent significant differences between groups (P < 0.05, Tukey's HSD test).

chronosequence ages were significantly different (*P*-values range from 0.001 to 0.037, ANOSIM with 999 permutations) (Table S1b). Interestingly, the overall pattern of β-diversity across the chronosequence age was maintained in each of the two clusters (Figure 3.2a), with bacterial community structure becoming more similar to reference sites with increasing chronosequence age. The magnitude of the shift in bacterial β-diversity within a year was comparable to that observed across decades of succession as evidenced by separation on the NMDS plot (Figure 3.2a) and average Bray-Curtis similarity values among samples (data not shown). In contrast, fungal communities from different chronosequence ages clustered more tightly without any overlap across ages (Figure 3.2b) and communities in each age were significantly different from

all other ages (*P*-value=0.001, ANOSIM with 999 permutations). In contrast, intra-annual variability across months in each chronosequence age was not significant (*P*-values range from 0.146 to 0.996, ANOSIM with 999 permutations) (Table S2). In addition, multivariate dispersion analysis showed that dispersion of bacterial communities by month and chronosequence age were not significantly different (Welch Two Sample t-test, t = 1.629, *P*-value = 0.147), while for fungi, the dispersion of fungal communities in each age class was significantly lower (i.e., tighter clustering) than when tested by month (Welch Two Sample t-test, t = 3.2691, *P*-value = 0.023).

Patterns of β-diversity were also related to soil and environmental data collected by Avera et al. (36) at each site. Changes in temperature, moisture, TMB, SIR and NH_4^+ were significantly correlated with bacterial community shifts (*P*-values range from 0.001 to 0.041) (Table S3). Moisture and NH_4^+ were mainly correlated with the intra-annual bacterial community shifts, with higher moisture in May and July and higher NH_4^+ in September and October (Figure 3.2a). In contrast, SIR was mainly correlated with bacterial community shifts that occurred with chronosequence age, with an increase along the chronosequence. For fungi, changes in temperature, moisture, TMB, SIR and NO_3^- were significantly correlated with community shifts (*P*-values range from 0.001 to 0.004) (Table S3). SIR, TMB and temperature were most strongly correlated with fungal shifts in β-diversity (Figure 3.2b), with higher SIR and TMB and lower soil temperature associated with fungal communities from older chronosequence ages. It is important to remember, however, that associated environmental factors could also be auto-correlated, and additional experimentation would be required to determine which might be directly driving or resulting from changes in soil microbiota.

The dominant bacterial phyla and proteobacterial classes across all samples were Acidobacteria and Alphaproteobacteria, respectively representing 21.5% and 18.0% of all sequences assigned to domain Bacteria (Figure 3.3a). Sampling month had significant influences on some taxa, with the relative abundances of Bacteroidetes, Deltaproteobacteria, Gemmatimonadetes, Betaproteobacteria and Gammaproteobacteria significantly higher in May and July and the relative abundances of Chloroflexi and Planctomycetes significantly higher in September and October (P-value<0.05, Tukey's HSD) (Figure 3.3b and Table S4). Chronosequence age also had significant influences on some taxa (P-value<0.05, ANOVA) (Figure 3.3c and Table S4). The relative abundances of Actinobacteria, Chloroflexi, Gemmatimonadetes were significantly higher in 5 yr sites than others, while the relative abundances of Nitrospirae, Alphaproteobacteria and Verrucomicrobia were significantly higher in later chronosequence ages. Nitrospirae were higher in 30 yr and UM sites than 5 yr and 11 yr sites, Alphaproteobacteria were higher in UM sites, and Verrucomicrobia were higher in UM sites than 5, 11, and 21 yr sites (P-value<0.05, Tukey's HSD). A summary of the most abundant OTUs that changed significantly across the chronosequence ages is presented in Figure S6a.

Ascomycota and Basidiomycota were the dominant fungal phyla at all sites, respectively representing 46% and 30% of classified OTUs, but their relative abundances were not clearly related to chronosequence age or sampling month. The dominant classes across all samples were Agaricomycetes (phylum Basidiomycota; 28.7%) and Sordariomycetes (phylum Ascomycota; 20.0%) (Figure 3.3d). Sampling month had significant effects on some fungal classes (P-value<0.05, ANOVA) (Figure 3.3e and Table S4). Dothideomycetes was significantly different between July and October, and Incertae sedis within Zygomycota was significantly different between May and September. Chronosequence ages had significant effects on more taxa,

including Dothideomycetes, Eurotiomycetes, Leotiomycetes, Sordariomycetes, Rozellomycota and Incertae sedis within Zygomycota (*P*-value<0.05, ANOVA) (Figure 3.3f and Table S4). The relative abundances of Dothideomycetes and Sordariomycetes were significantly higher in 5 yr sites, while the relative abundance of Agaricomycetes was significantly higher in UM site than 5 yr sites. The relative abundance of Incertae sedis in Zygomycota was significantly higher in 30 yr and UM sites than in 5 and 11 yr sites. All related *P*-values are shown in Table S4, and the most abundant fungal OTUs that changed significantly across the chronosequence ages are summarized in Figure S6b.

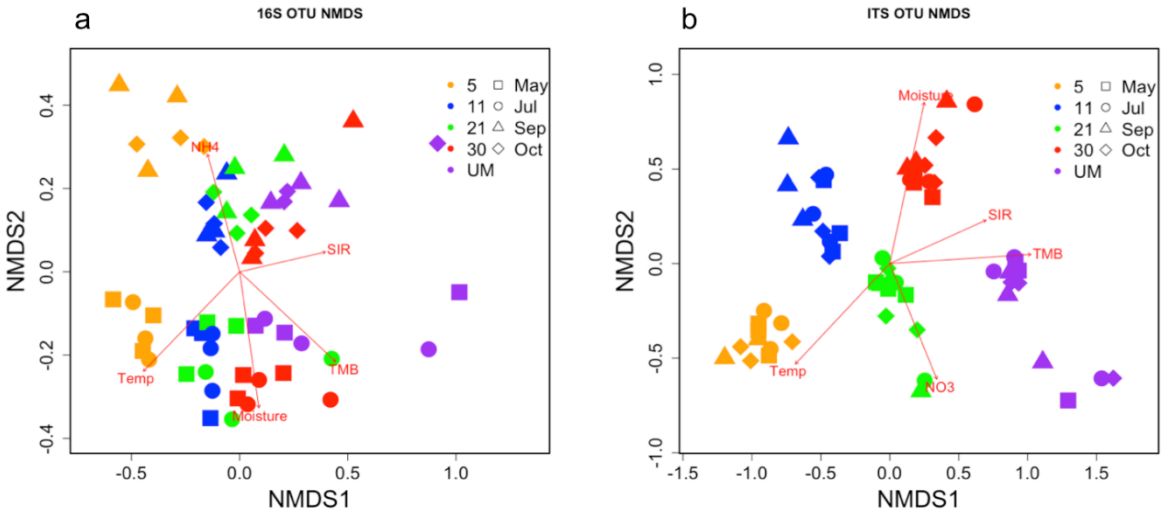


Figure 3.2. Bacterial (a) and fungal (b) β-diversity. Microbial β-diversity was visualized with NMDS based on OTU abundance data. Environmental factors (Temp, moisture, available NH_4^+, and NO_3^-; gas fluxes of N_2O, CO_2, and CH_4; and TMB and SIR) are fit to the ordination with function envfit in R vegan package. Only significant factors are shown in the figure.

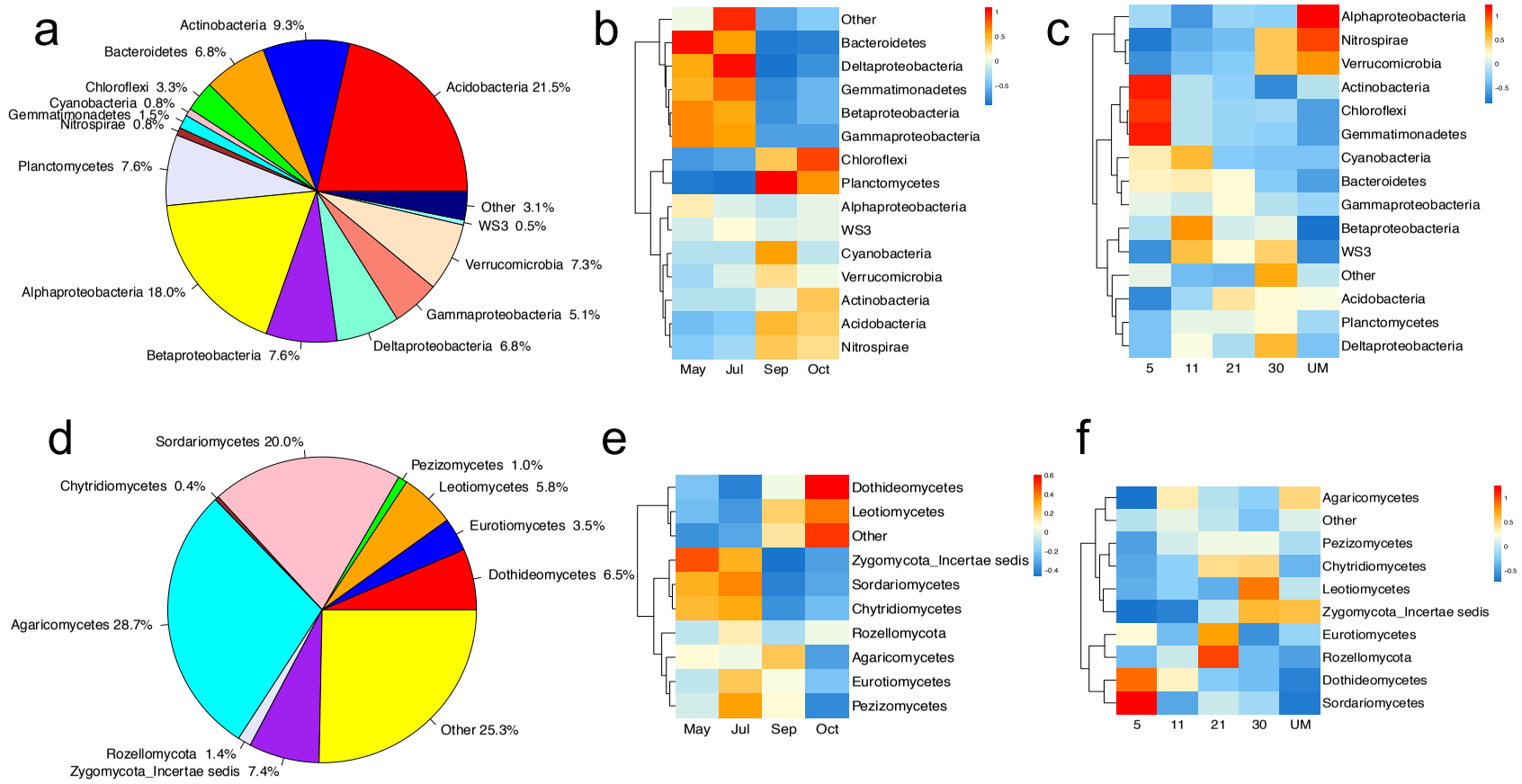


Figure 3.3. Bacterial and fungal taxonomic composition and their variations over sampling months and chronosequence ages. (a) Averaged relative abundance of bacterial phyla or classes (within Proteobacteria) over all samples. The relative abundance was calculated as the percentage of sequences. (b and c) The heatmaps show the bacterial taxonomic variations over sampling months and chronosequence ages. The key represents the z-scores of the relative abundances of the taxa. (d) Averaged relative abundance of fungal classes over all samples. (e and f) The fungal taxonomic variations over sampling months and chronosequence ages. The x axis values in panels b and e represent sampling times, and numbers on the x axis in panels c and f represent chronosequence ages in years. The dendrograms to the left of the heatmaps represent clustering of microbial taxa based upon similar patterns of variation.

3.3.4 Co-occurrence network analysis

Co-occurrence networks were generated separately for bacteria and fungi at each chronosequence age and topological properties were calculated to characterize changes over time. The numbers of nodes and edges in the bacterial networks typically increased with chronosequence age, with the unmined reference network having the highest number of nodes and edges (Table 3.1). Moreover, the average degree also increased from early chronosequence ages (5 and 11 yr) to later chronosequence ages (21 and 30 yr) (Table 3.1 and Figure 3.4). The average degree of the unmined reference bacterial network (with an average degree of 11.66) is about 1.5 times of those at 21 and 30 yr (7.53 and 7.16). Networks in fungal communities, however, were markedly different from bacterial communities (Table 3.1 and Figure 3.5). The numbers of fungal nodes and edges both peak at 21 yr (Table 3.1), with the fungal networks at 5 yr having the smallest number of nodes and edges and the lowest average degree. The average degree of the unmined reference network is larger than 5 yr but smaller than 11 yr and 21 yr. Overall, the nodes, edges and average degree of fungal networks typically peaked in the middle chronosequence ages.

Networks in consecutive chronosequence ages were compared to identify overlapping nodes and edges. On average, 76% of bacterial nodes and 8% of bacterial edges in an earlier age appeared in the next age, and overlaps of both bacterial nodes and edges are significantly higher than would be expected from random (Fisher's exact test for node, *P*-values range from 0.001 to 0.004; Permutation test for edge with 999 permutations, *P*-value=0.001) (Table S5). In contrast, a mean of 27% of fungal nodes and 0.5% of fungal edges in an earlier age appeared in the next chronosequence age, and the number of overlapping fungal nodes and edges are not significantly

greater than random (Fisher's exact test for node, *P*-values range from 0.872 to 0.999; Permutation test for edge with 999 permutations, *P*-values range from 0.614 to 0.990) (Table S5).

Among the overlapping bacterial nodes, about 30% belong to the phylum Acidobacteria, which is higher than the relative abundance of Acidobacteria among all OTUs used for network analysis (22%). Bacteroidetes and Planctomycetes were also overrepresented in overlapping nodes compared to their percentages in all OTUs, at 14% and 10% compared to 10% and 6%, respectively. In contrast, taxa that were underrepresented in overlaps included Actinobacteria (5% of overlapping nodes vs. 13% of OTUs) and Alphaproteobacteria (11% of overlapping nodes vs. 16% of OTUs). Among the 352 overlapping bacterial edges, 223 (63%) involve Acidobacteria and 146 of 352 (42%) are between two Acidobacteria nodes. More broadly, about 86% of overlapping edges involve Acidobacteria, Bacteroidetes or Planctomycetes, even though the combined relative abundance of these 3 taxa is only 38% of total OTUs used for network analysis. Nodes involved in the overlapping edges have significantly higher degrees and closeness centrality than average (*P*-values range from 0.001 to 0.006, Welch Two Sample t-test), but the relative abundance of OTUs represented by those nodes are not always significantly different from average (*P*-values range from 0.001 to 0.683, Welch Two Sample t-test) (Table S6). The overlapping fungal nodes mainly belong to Agaricomycetes, Sordariomycetes and Dothideomycetes, with Agaricomycetes being overrepresented (15% of overlapping nodes vs. 12% of OTUs). In contrast to bacteria, there were only 3 overlapping fungal edges among all possible pairs of consecutive chronosequence ages.

The bacterial and fungal networks show different patterns of co-occurrence, with typically higher numbers of nodes and edges in bacterial networks. The bacterial networks typically contained a

much higher number of interactions dominated by one large connected component that consisted of more than 65% of all OTUs (Figure 3.4). In contrast, fungal networks consisted of much fewer interactions scattered across multiple small clusters (i.e., only a few OTUs) that were discrete and densely clustered (higher clustering coefficients) (Figure 3.5, Table 3.1). It is important to note that the relatively low intra-annual turnover rates of fungi may result in less detectable co-occurrences. However, the stark differences between the two sets of networks, which were observed consistently even when the networks were re-analyzed with a range of alternative correlation cutoffs (Figure S3), suggest that there are real and significant differences between these sets of interactions. We also constructed bacteria-fungi co-occurrence networks with the 500 most abundant bacterial OTUs and the 500 most abundant fungal OTUs. In these networks, there are more bacterial nodes and edges than fungal (Figure S5 and Table S8), which is consistent with our observations in the separate networks. Also, the number of bacteria-bacteria edges increased with chronosequence age, while the number of fungi-fungi edges peaks at 21 yr, which are also consistent with the trends in the separate networks. The number of bacteria-fungi edges increased with the chronosequence age, but they comprised only 7% of all edges on average. Additional characteristics of bacterial and fungal networks including scale free indices, small world indices, and modularity are summarized in Table S7 and Figure S4.

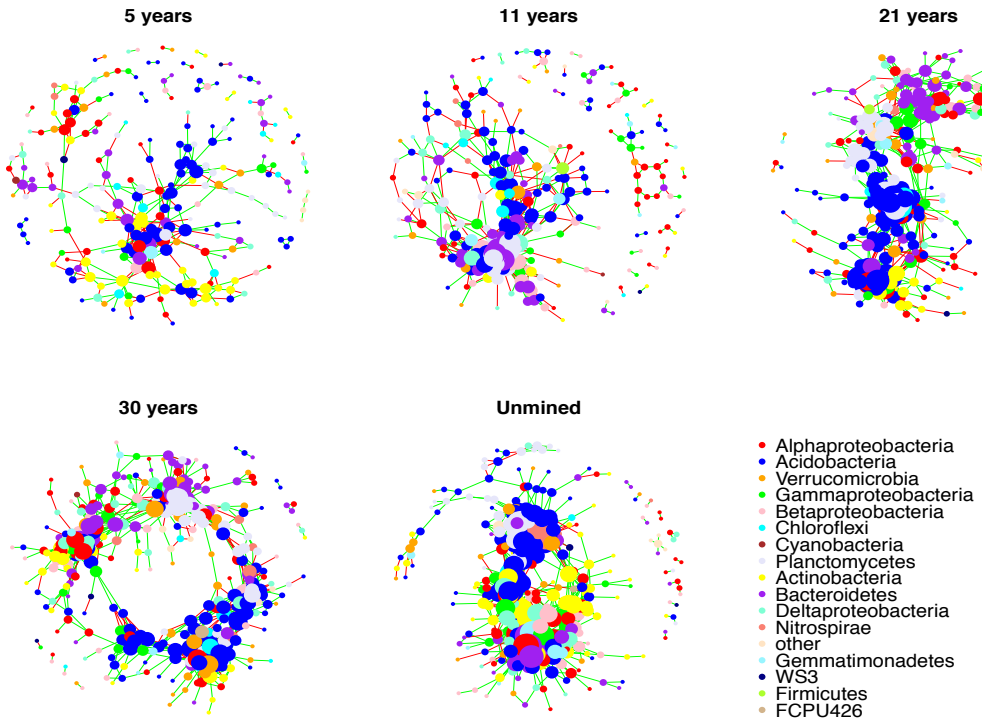


Figure 3.4. Co-occurrence network analysis of bacterial communities at different succession stages. Each dot represents a bacterial OTU. A connection stands for a spearman correlation with a magnitude > 0.9 (positive correlation—green edges) or < -0.9 (negative correlation—red edges) and statistically significant (FDR < 0.05). The size of each node is proportional to the square root of number of connections. The colors of nodes represent their classification at phylum level or class level for those belonging to phylum Proteobacteria.

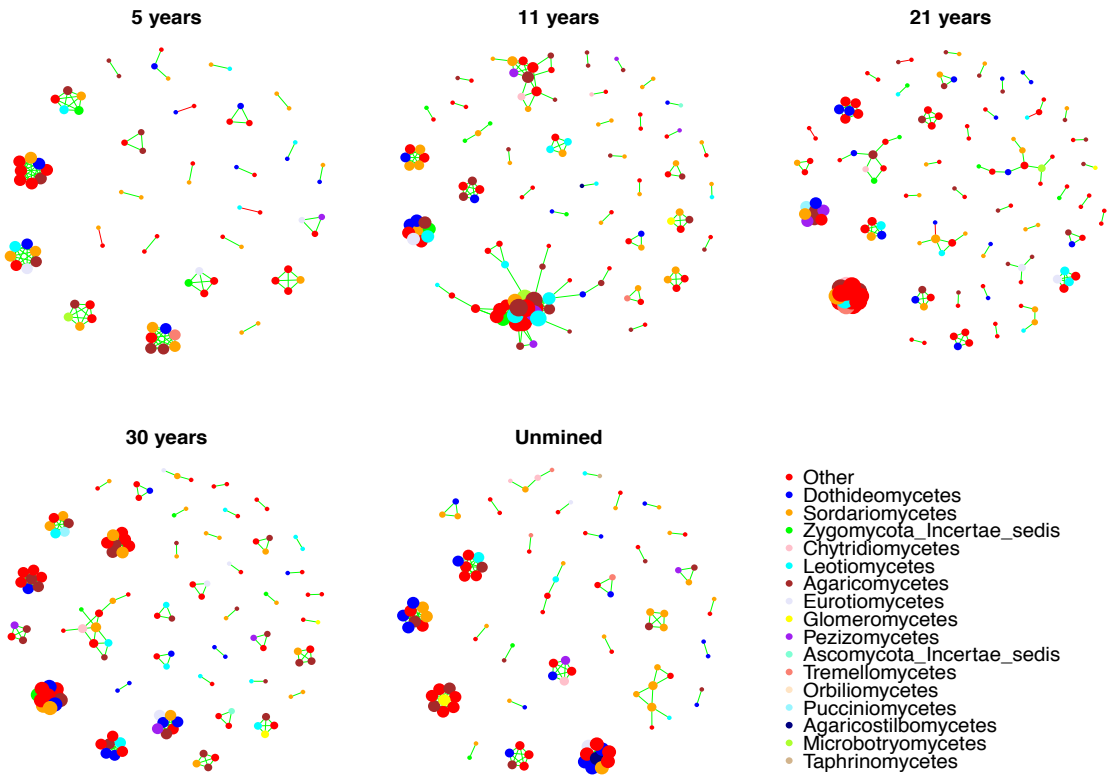


Figure 3.5. Co-occurrence network analysis of fungal communities at different succession stages. Each dot represents a fungal OTU. A connection stands for a spearman correlation with a magnitude > 0.9 (positive correlation—green edges) or < -0.9 (negative correlation—red edges) and statistically significant (FDR < 0.05). The size of each node is proportional to the square root of number of connections. The colors of nodes represent their classification at class level.

Table 3.1. Bacterial and fungal network properties at succession stages.

		Stage				
	Network metrics	*5*	*11*	*21*	*30*	*Un-mined*
Bacteria	Nodes	248	275	346	353	379
	Edges	384	733	1303	1263	2210
	Pos/Neg Edges	257/127	416/317	1044/259	892/371	1811/399
	Average Degree	3.10	5.33	7.53	7.16	11.66
	Cluster Coefficient	0.319	0.438	0.510	0.459	0.490
Fungi	Nodes	80	133	147	134	105
	Edges	127	256	266	240	188
	Pos/Neg Edges	124/3	256/0	262/4	240/0	188/0
	Average Degree	3.17	3.85	3.62	3.58	3.58
	Cluster Coefficient	0.998	0.744	0.984	0.983	0.986

3.4 Discussion

The results of this work highlight several key findings related to recovery of soil microbiota following extreme disturbance. Firstly, both bacterial and fungal fractions of the soil microbiota followed an observable taxonomic transition with chronosequence age, with community structure becoming more similar to unmined reference sites. Secondly, both bacterial and fungal community composition changes were apparent at high taxonomic levels (e.g., bacterial phylum and fungal class). Thirdly, the bacterial co-occurrence networks increased in complexity during succession while the fungal network complexity did not change monotonically with age. Overall, the bacterial and fungal fractions of the soil microbiota exhibited distinct patterns with regard to chronosequence age, intra-annual variability, and co-occurrence network properties.

3.4.1 Bacterial and fungal abundance, richness and β-diversity

Richness of both bacteria and fungi peaked at 30 yr and were lower in the unmined reference sites. We hypothesize that the short time since disturbance at the younger sites was limiting, such that richness was still increasing up to the 30 yr sites. This aspect of ecosystem recovery allows the coexistence of species adapted for both high growth rates in high resource availability (*r*-selected) and lower growth rates but higher competitive ability under low resource availability (*k*-selected) at the older sites, which could explain the peak richness at 30 yr site (55, 56). This general trend is in agreement with other observations of microorganisms in aquatic (57) and soil systems (58, 59)

The β-diversity also followed a clear pattern for both bacteria and fungi in that the community structure became more similar to the unmined sites as chronosequence age increased. The bacterial community shifts were more gradual and overlapping while the fungal communities at different ages were more distinct. This agrees with previous observations that plant symbioses with fungi are more common than with bacteria (16), and that plant community composition has strong influences on the fungal community (17-20), which has been shown to extend to bulk soil in the form of common mycelial networks (60, 61). The specificity with vegetation may explain the distinctness of fungal communities at different chronosequence ages, as the plant community is also experiencing taxonomic shifts from grass-dominated to tree-dominated during this timeframe (62). Furthermore, as dominant decomposers of litter (63), the variable litter properties with different vegetation (64, 65) could also influence fungal community composition. In contrast, most of bacteria inhabit small-scale niches in bulk soil and have less direct connection with plants than fungi (66). However, it is important to remember that in addition to external environmental factors, the fact that community changes along the chronosequence age

were correlated with TMB and SIR suggests that autogenic factors also play important roles in microbial recovery.

The patterns of intra-annual variability among bacterial and fungal communities were also markedly different (Figure 3.1 and 3.2), with bacterial abundance higher in May and July and fungal abundance higher in September and October. This shift was likely related to seasonal dynamics of plant inputs, temperature and precipitation. For example, fungi are often identified as the primary decomposers of recalcitrant organic matter (63), such as cellulose and lignin, which are increased in the later seasons due to plant litter and may favor higher fungal abundance in the later sampling months. Based on the ANOSIM test and NMDS visualization, bacterial community structure shifted significantly among sampling months, while fungal community structures did not. This difference in intra-annual variability may also be related to growth rates and highlights the ecological differences between these two communities. For example, soil fungal biomass turnover times have been estimated to be >100 days, which is about tenfold slower than typical soil bacteria (14, 67). In addition, bacteria are more sensitive to changes in soil moisture (68, 69), which correlated significantly with the intra-annual bacterial community shifts observed in this study (r^2=0.157, *P*-values=0.012). Short-term variability of bacterial communities has been observed previously, but the magnitude of change varies with soil or ecosystem types (70-72). In our study, β-diversity indicates that the shifts of soil bacterial communities within a 6-month span could be as large as those observed across decades of ecosystem recovery, although the effect of chronosequence age was still observable within each monthly shift (Figure 3.2a). In contrast, fungal community structures did not change significantly over 6 months, implying that the majority of change in soil fungal communities occurs over longer time scales than for bacteria.

3.4.2 Bacterial and fungal taxonomic shifts

Changes in the types of soil microbiota along chronosequence age were apparent at both low and high taxonomic levels, including bacterial phylum and fungal class (Figure 3.3c and f), which suggests that significant taxonomic changes were occurring across the entire community and supports the theory that at least some degree of ecological coherence exists among broad phylogenetic groups (73-75). In succession, bacterial communities appeared to shift from copiotrophic groups to oligotrophic groups, with decreased Bacteroidetes and increased Verrucomicrobia (Figure 3.3c) (76-79). This shift potentially resulted from the decrease of abundant labile C and N substrates and the increase of nutrient-poor and recalcitrant C compounds, which is supported by a measured increase in mineral soil C:N ratio with chronosequence age at these sites (36). Ecosystem changes are also apparent at lower taxonomic levels. Many of the OTUs that increase significantly over the chronosequence, such as *Bradyrhizobium*, Chthoniobacteraceae: DA101, *Mycobacterium*, and *Rhodoplanes*, indicate increased soil organic matter decomposition and nutrient cycling as conditions become more oligotrophic. Interestingly, we also saw an increase in Candidatus *Xiphinematobacter* in our data, suggesting an increase in nematode populations (80) and potential development of higher trophic levels in the soil ecosystem.

Overall, these composition changes also likely result from broader ecosystem succession and vegetation change, as plant inputs (leaf litter, root biomass, and exudates) vary with vegetation type (64, 65, 81) and have a strong influence on microbial community composition (82, 83). For example, we observed increased Alphaproteobacteria and decreased Actinobacteria (Figure 3.3c), which is a change consistent with a shift from grassland to forest soil (84). With regard to fungi, many of the Dothideomycetes are tolerant of extreme environmental conditions including

extreme heat and cold, drought and UV radiation, and some produce enzymes that help degrade rocks (85, 86). Given that the 5 yr sites lacked mature tree coverage that can help buffer rapidly fluctuating conditions and reduce direct UV radiation, Dothideomycetes may have competitive advantages in this relatively harsh environment (Figure 3.3f). Among more specific taxa, we saw increases in fungal genera like *Hygrophorus*, which includes tree-associated ectomycorhizzae (87, 88), and other common forest saprotrophs such *Entoloma* and *Mortierella*.

3.4.3 Changes in co-occurrence network properties

In larger organisms, the complexity of ecological networks has been shown to increase during ecosystem succession, leading to greater food web stability (89). Similarly, we observed increased bacterial co-occurrence network complexity with chronosequence age, with more co-occurrent OTUs retained and higher connectivity in later ages. Functionally, this development was also accompanied by previously reported increases in microbial activity at these sites (36). A. Zelezniak et al. (90) found that, across multiple habitat types, species co-occurrence tends to be mainly driven by metabolic dependencies. This supports the connection between network complexity and microbial activity at these sites, and illustrates the need to conduct more complex investigations of how interaction networks can be used to gain insight into the relationship between changes in microbial function and diversity and how they relate to ecosystem change.

To further explore changes in co-occurrence networks at different chronosequence ages, we quantified the overlaps of networks at consecutive stages, including the significance of overlapping OTUs and their taxonomic compositions. The significantly higher than average numbers of overlaps in both bacterial OTUs and co-occurrences indicates that a portion of the bacterial interaction network persists over relatively long periods of time during recovery of soil

microbiota. In contrast, fungal networks changed more substantially and displayed new interactions in each chronosequence age with no more overlaps between stages than would be expected from random chance, suggesting minimal to no influence from the previous stage. These network differences are consistent with the β-diversity results illustrating a gradual shift of bacterial communities between ages but relatively distinct shifts between fungal communities.

In a simulation of co-occurrence networks, D. Berry and S. Widder (91) identified keystone species as having higher values for mean degree and closeness centrality. Applying this rationale, we identified bacterial OTUs in overlapping edges as potential keystone species, which have degrees and closeness centrality significantly higher than average (Table S6). The potential keystone OTUs observed in our study were dominated mostly by Acidobacteria, but also by Bacteroidetes and Planctomycetes. It is important to note that these are not simply a result of more abundant taxa having a higher chance to co-occur with other taxa, because the relative abundance of some keystone OTUs identified in our study were not significantly different from average (Table S6). Identification of potentially important OTUs is critical to improving our understanding of the ecological roles played by specific bacterial taxa. For example, although Acidobacteria is one of the most abundant phyla in soils and represented many of the potential keystone OTUs identified in this study, little is known about them because they have been very difficult to isolate and culture (92, 93).

Overall, these results provide a comprehensive view of how the recovery of the bacterial and fungal components of the soil microbiota can be strikingly different during ecosystem recovery following extreme disturbance. This work highlights numerous ways in which different properties of soil bacteria and fungi can influence their variability during ecosystem recovery

within the same environment. Further investigations that focus on the ecological roles of specific taxa, microbial interactions, and potential keystone species are essential to obtaining a better understanding of the mechanisms that shape patterns of microbial diversity and function during ecosystem recovery after disturbances.

References

1. Clements FE. 1916. Plant succession: an analysis of the development of vegetation. Carnegie Institution of Washington.
2. Cowles HC. 1899. The ecological relations of the vegetation on the sand dunes of Lake Michigan (concluded). Bot Gaz:361-391.
3. McLachlan SM, Bazely DR. 2001. Recovery patterns of understory herbs and their use as indicators of deciduous forest regeneration. Conserv Biol 15:98-110.
4. Ruiz‐Jaen MC, Mitchell Aide T. 2005. Restoration success: how is it being measured? Restoration Ecol 13:569-577.
5. Matthews JW, Spyreas G, Endress AG. 2009. Trajectories of vegetation‐based indicators used to assess wetland restoration progress. Ecol Appl 19:2093-2107.
6. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, Yang Y, He Z, Wu L, Stahl DA, Hazen TC, Tiedje JM, Arkin AP. 2014. Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. Proc Natl Acad Sci U S A 111:E836-E845.
7. Fukami T, Morin PJ. 2003. Productivity–biodiversity relationships depend on the history of community assembly. Nature 424:423-426.
8. Knelman JE, Schmidt SK, Lynch RC, Darcy JL, Castle SC, Cleveland CC, Nemergut DR. 2014. Nutrient addition dramatically accelerates microbial community succession. PLoS One 9:e102609.
9. Paver SF, Hayek KR, Gano KA, Fagen JR, Brown CT, Davis‐Richardson AG, Crabb DB, Rosario‐Passapera R, Giongo A, Triplett EW. 2013. Interactions between specific phytoplankton and bacteria affect lake bacterial community succession. Environ Microbiol 15:2489-2504.
10. Fierer N, Nemergut D, Knight R, Craine JM. 2010. Changes through time: integrating microorganisms into the study of succession. Res Microbiol 161:635-42.
11. Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF, Knelman JE, Darcy JL, Lynch RC, Wickey P. 2013. Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev 77:342-356.
12. Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL, Murphy DV. 2011. Soil Microbial Community Successional Patterns during Forest Ecosystem Restoration. Appl Environ Microbiol 77:6158-6164.
13. Kirchman DL. 2012. Processes in microbial ecology. Oxford University Press.
14. Rousk J, Bååth E. 2007. Fungal biomass production and turnover in soil estimated using the acetate-in-ergosterol technique. Soil Biol Biochem 39:2173-2177.

15. Rinnan R, Bååth E. 2009. Differential utilization of carbon substrates by bacteria and fungi in tundra soil. Appl Environ Microbiol 75:3611-3620.
16. Bonfante P, Anca I-A. 2009. Plants, Mycorrhizal Fungi, and Bacteria: A Network of Interactions. Annu Rev Microbiol 63:363-383.
17. Vandenkoornhuyse P, Ridgway KP, Watson IJ, Fitter AH, Young JPW. 2003. Co-existing grass species have distinctive arbuscular mycorrhizal communities. Mol Ecol 12:3085-3095.
18. Mueller RC, Paula FS, Mirza BS, Rodrigues J, Nüsslein K, Bohannan B. 2014. Links between plant and fungal communities across a deforestation chronosequence in the Amazon rainforest. ISME J 8:1548-1550.
19. Hazard C, Gosling P, van der Gast CJ, Mitchell DT, Doohan FM, Bending GD. 2013. The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. ISME J 7:498-508.
20. Peay KG, Baraloto C, Fine PV. 2013. Strong coupling of plant and fungal community structure across western Amazonian rainforests. ISME J 7:1852-1861.
21. Schmidt S, Nemergut D, Darcy J, Lynch R. 2014. Do bacterial and fungal communities assemble differently during primary succession? Mol Ecol 23:254-258.
22. Peay KG, Belisle M, Fukami T. 2011. Phylogenetic relatedness predicts priority effects in nectar yeast communities. Proc Natl Acad Sci U S A:rspb20111230.
23. Dini-Andreote F, de Cassia Pereira e Silva M, Triado-Margarit X, Casamayor EO, van Elsas JD, Salles JF. 2014. Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. ISME J 8:1989-2001.
24. Kitayama K, Mueller‐Dombois D, Vitousek PM. 1995. Primary succession of Hawaiian montane rain forest on a chronosequence of eight lava flows. J Veg Sci 6:211-222.
25. Schutte UM, Abdo Z, Bent SJ, Williams CJ, Schneider GM, Solheim B, Forney LJ. 2009. Bacterial succession in a glacier foreland of the High Arctic. ISME J 3:1258-68.
26. Nemergut DR, Anderson SP, Cleveland CC, Martin AP, Miller AE, Seimon A, Schmidt SK. 2007. Microbial community succession in an unvegetated, recently deglaciated soil. Microb Ecol 53:110-22.
27. Zumsteg A, Luster J, Göransson H, Smittenberg R, Brunner I, Bernasconi S, Zeyer J, Frey B. 2012. Bacterial, Archaeal and Fungal Succession in the Forefield of a Receding Glacier. Microb Ecol 63:552-564.
28. Sigler W, Zeyer J. 2002. Microbial diversity and activity along the forefields of two receding glaciers. Microb Ecol 43:397-407.
29. Schutte UM, Abdo Z, Foster J, Ravel J, Bunge J, Solheim B, Forney LJ. 2010. Bacterial diversity in a glacier foreland of the high Arctic. Mol Ecol 19 Suppl 1:54-66.
30. Philippot L, Tscherko D, Bru D, Kandeler E. 2011. Distribution of high bacterial taxa across the chronosequence of two alpine glacier forelands. Microb Ecol 61:303-312.
31. Bradley JA, Singarayer JS, Anesio AM. 2014. Microbial community dynamics in the forefield of glaciers. Proc Biol Sci 281.
32. Fields-Johnson C, Zipper C, Burger J, Evans D. 2012. Forest restoration on steep slopes after coal surface mining in Appalachian USA: soil grading and seeding effects. For Ecol Manage 270:126-134.

33. Burger J, Graves D, Angel P, Davis V, Zipper C. 2005. The Forestry Reclamation Approach. US Office of Surface Mining, Forest Reclamation Advisory 2 http://arriosmregov/FRA/Advisories/FRA_No27-18-07Revisedpdf.
34. Frouz J, Prach K, Pižl V, Háněl L, Starý J, Tajovský K, Materna J, Balík V, Kalčík J, Řehounková K. 2008. Interactions between soil development, vegetation and soil fauna during spontaneous succession in post mining sites. Eur J Soil Biol 44:109-121.
35. Ruppert L. 2000. Chapter A–Executive summary–Coal resource assessment of selected coal beds and zones in the northern and central Appalachian Basin coal regions. Northern and central Appalachian basin coal regions assessment team.
36. Avera BN, Strahm BD, Burger JA, Zipper CE. 2015. Development of ecosystem structure and function on reforested surface-mined lands in the Central Appalachian Coal Basin of the United States. New For 46:683-702.
37. Fierer N, Jackson JA, Vilgalys R, Jackson RB. 2005. Assessment of soil microbial community structure by use of taxon-specific quantitative PCR assays. Appl Environ Microbiol 71:4117-20.
38. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621-4.
39. Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol 10:189.
40. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-1.
41. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194-2200.
42. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590-6.
43. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson K-H. 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol 22:5271-5277.
44. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol 73:5261-5267.
45. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-6.

46. R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
47. Jari Oksanen FGB, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens and Helene Wagner. 2015. vegan: Community Ecology Package. R package version 22-1 http://CRANR-projectorg/package=vegan.
48. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140.
49. Faust K, Raes J. 2012. Microbial interactions: from networks to models. Nat Rev Microbiol 10:538-550.
50. Ma B, Wang H, Dsouza M, Lou J, He Y, Dai Z, Brookes PC, Xu J, Gilbert JA. 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. ISME J.
51. Csardi G, Nepusz T. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695:1-9.
52. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. 2004. The architecture of complex weighted networks. Proc Natl Acad Sci U S A 101:3747-3752.
53. Freeman LC. 1978. Centrality in social networks conceptual clarification. Soc Networks 1:215-239.
54. Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996-8.
55. Mouquet N, Moore JL, Loreau M. 2002. Plant species richness and community productivity: why the mechanism that promotes coexistence matters. Ecol Lett 5:56-65.
56. Roxburgh SH, Shea K, Wilson JB. 2004. The intermediate disturbance hypothesis: patch dynamics and mechanisms of species coexistence. Ecology 85:359-371.
57. Flöder S, Sommer U. 1999. Diversity in planktonic communities: an experimental test of the intermediate disturbance hypothesis. Limnol Oceanogr 44:1114-1119.
58. Wakelin SA, Chu G, Lardner R, Liang Y, McLaughlin M. 2010. A single application of Cu to field soil has long-term effects on bacterial community structure, diversity, and soil processes. Pedobiologia 53:149-158.
59. Degens BP, Schipper LA, Sparling GP, Duncan LC. 2001. Is the microbial community in a soil with reduced catabolic diversity less resistant to stress or disturbance? Soil Biol Biochem 33:1143-1153.
60. Urbanová M, Šnajdr J, Baldrian P. 2015. Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. Soil Biology and Biochemistry 84:53-64.
61. Babikova Z, Gilbert L, Bruce TJ, Birkett M, Caulfield JC, Woodcock C, Pickett JA, Johnson D. 2013. Underground signals carried through common mycelial networks warn neighbouring plants of aphid attack. Ecology Letters 16:835-843.
62. Zipper CE, Burger JA, Skousen JG, Angel PN, Barton CD, Davis V, Franklin JA. 2011. Restoring forests and associated ecosystem services on Appalachian coal surface mines. Environ Manage 47:751-765.

63. Boer W, Folman LB, Summerbell RC, Boddy L. 2005. Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol Rev 29:795-811.
64. Hättenschwiler S, Aeschlimann B, Coûteaux MM, Roy J, Bonal D. 2008. High variation in foliage and leaf litter chemistry among 45 tree species of a neotropical rainforest community. New Phytol 179:165-175.
65. Cornwell WK, Cornelissen JH, Amatangelo K, Dorrepaal E, Eviner VT, Godoy O, Hobbie SE, Hoorens B, Kurokawa H, Pérez‐Harguindeguy N. 2008. Plant species traits are the predominant control on litter decomposition rates within biomes worldwide. Ecol Lett 11:1065-1071.
66. Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. 2013. Micro-scale determinants of bacterial diversity in soil. FEMS microbiology reviews 37:936-954.
67. Bååth E. 1998. Growth rates of bacterial communities in soils at varying pH: a comparison of the thymidine and leucine incorporation techniques. Microb Ecol 36:316-327.
68. Manzoni S, Schimel JP, Porporato A. 2012. Responses of soil microbial communities to water stress: results from a meta-analysis. Ecology 93:930-8.
69. Bell C, McIntyre N, Cox S, Tissue D, Zak J. 2008. Soil Microbial Responses to Temporal Variations of Moisture and Temperature in a Chihuahuan Desert Grassland. Microb Ecol 56:153-167.
70. Lauber CL, Ramirez KS, Aanderud Z, Lennon J, Fierer N. 2013. Temporal variability in soil microbial communities across land-use types. ISME J 7:1641-50.
71. Matulich KL, Weihe C, Allison SD, Amend AS, Berlemont R, Goulden ML, Kimball S, Martiny AC, Martiny JB. 2015. Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. ISME J doi:10.1038/ismej.2015.58.
72. Slaughter LC, Weintraub MN, McCulley RL. 2015. Seasonal Effects Stronger than Three-Year Climate Manipulation on Grassland Soil Microbial Community. Soil Sci Soc Am J 79:1352-1365.
73. Philippot L, Andersson SG, Battin TJ, Prosser JI, Schimel JP, Whitman WB, Hallin S. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol 8:523-529.
74. Lennon JT, Aanderud ZT, Lehmkuhl B, Schoolmaster Jr DR. 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology 93:1867-1879.
75. Martiny JB, Jones SE, Lennon JT, Martiny AC. 2015. Microbiomes in light of traits: A phylogenetic perspective. Science 350:aac9323.
76. Fierer N, Bradford MA, Jackson RB. 2007. Toward an ecological classification of soil bacteria. Ecology 88:1354-64.
77. Davis KER, Sangwan P, Janssen PH. 2011. Acidobacteria, Rubrobacteridae and Chloroflexi are abundant among very slow-growing and mini-colony-forming soil bacteria. Environ Microbiol 13:798-805.
78. Barnard RL, Osborne CA, Firestone MK. 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. ISME J 7:2229-41.
79. Mueller RC, Belnap J, Kuske CR. 2015. Soil bacterial and fungal community responses to nitrogen addition across soil depths and microhabitat in an arid shrubland. Front Microbiol 6.

80. Vandekerckhove TT, Coomans A, Cornelis K, Baert P, Gillis M. 2002. Use of the Verrucomicrobia-specific probe EUB338-III and fluorescent in situ hybridization for detection of "Candidatus Xiphinematobacter" cells in nematode hosts. Appl Environ Microbiol 68:3121-3125.
81. Finér L, Helmisaari HS, Lõhmus K, Majdi H, Brunner I, Børja I, Eldhuset T, Godbold D, Grebenc T, Konôpka B, Kraigher H, Möttönen MR, Ohashi M, Oleksyn J, Ostonen I, Uri V, Vanguelova E. 2007. Variation in fine root biomass of three European tree species: Beech (Fagus sylvatica L.), Norway spruce (Picea abies L. Karst.), and Scots pine (Pinus sylvestris L.). Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology 141:394-405.
82. Crowther TW, Maynard DS, Leff JW, Oldfield EE, McCulley RL, Fierer N, Bradford MA. 2014. Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study. Glob Chang Biol 20:2983-94.
83. Schimel JP, Schaeffer SM. 2012. Microbial control over carbon cycling in soil. Front Microbiol 3:348.
84. Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Daniel R. 2011. Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils. PLoS One 6:e17000.
85. Ohm RA, Feau N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, Condon BJ, Copeland AC, Dhillon B, Glaser F. 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog 8:e1003037.
86. Treseder KK, Lennon JT. 2015. Fungal traits that drive ecosystem dynamics on land. Microbiol Mol Biol Rev 79:243-262.
87. Kernaghan G. 2001. Ectomycorrhizal fungi at tree line in the Canadian Rockies. Mycorrhiza 10:217-229.
88. Rineau F, Garbaye J. 2009. Effects of liming on ectomycorrhizal community structure in relation to soil horizons and tree hosts. Fungal Ecol 2:103-109.
89. Neutel A-M, Heesterbeek JA, van de Koppel J, Hoenderboom G, Vos A, Kaldeway C, Berendse F, de Ruiter PC. 2007. Reconciling complexity with stability in naturally assembling food webs. Nature 449:599-602.
90. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. 2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci U S A 112:6449-6454.
91. Berry D, Widder S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol 5.
92. Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J 3:442-453.
93. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J. 2009. Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol 75:2046-2056.

Chapter 4. Metagenomic analysis reveals potential changes in microbial function during forest ecosystem restoration.

Abstract

The soil microbial community mediates numerous important ecological processes and plays essential roles in the recovery of an ecosystem after disturbance. However, given that relationships between microbial phylogeny and ecological function can be difficult to predict, data on microbial community structure alone can make it difficult to identify and interpret processes and relationships between microorganisms and ecosystem processes during recovery. In contrast, direct examination of microbial functional genes within the community via shotgun metagenome sequencing can provide a more comprehensive picture of the functional interactions driving recovery of the microbial community and the two-way regulation between microorganisms and their environment. In this study, we analyzed 15 metagenomes from a chronosequence of reforested mine soils spanning 6 to 31 years and from unmined reference soils. Patterns of phylogenetic change within the soil microbiota as characterized by metagenome analysis are highly similar to those revealed with amplicon sequencing, and the changes further reinforce a shift from copiotrophic to oligotrophic taxa. With regard to functional shifts, changes in relative abundances of genes involved in virulence, defense, and stress response indicated increased competition between microorganisms with increasing chronosequence age, which is likely related to decreases in available nutrients. Low relative abundances of methanotrophs and methane monooxygenase genes in all reforested soils compared to reference sites indicate that the ecosystem function of forest soil as methane sink is not recovered even 31 years after

reforestation. This work provides a detailed look at the important information regarding the possible mechanisms influencing microbial succession and the recovery of ecosystem functions.

4.1 Introduction

Forest ecosystems provide important ecological services; they play essential roles in global biogeochemical cycling and regulate atmospheric greenhouse gases (1). The nitrogen cycle is of particular importance in forest ecosystems, where N is usually a major limiting nutrient for plants and microorganisms (2). Forest clearing and disturbance can disrupt the nitrogen cycle (3), and the return of nitrogen cycle during reforestation is essential for the recovery of ecosystem functions. Within forest ecosystems, soil supports most ecosystem processes (4), and soil microbial communities mediate many of the ecological processes essential for ecosystem function and biogeochemical cycling (5, 6). In particular, soil microorganisms play important roles in C and N cycling, which influence the ecosystem services such as C sequestration and N retention (7, 8); methanotrophs in forest soils, for example, are the largest biological sinks for methane in atmosphere (9). Therefore, knowledge about soil microbial communities and their interactions with the environment is important for understanding the change and recovery of ecosystem function after disturbance.

Microbial communities are exposed to varying environmental factors and are likely under the regulation of available resources, vegetation impacts, environmental stress and biological interactions (10-12). For example, the influence of plant litter deposition and root exudates on microbial community suggested that vegetation could be an important driver of microbial succession post disturbances (13, 14), and changes in resource availability can also influence microbial succession (10). Previous studies have observed increases in microbial biomass,

respiration, N mineralization and nitrification during succession (15-17). Soil pH, accumulated organic matter, plant community structure and litter chemistry influenced microbial community function during succession (17). Studies of microbial taxonomic changes during ecosystem succession suggest that microbial communities follow predictable patterns, which likely reflect corresponding changes in environmental conditions and also likely play an important role in the recovery of ecosystem functions (18-20). However, taxonomic changes alone can only provide limited information regarding functions based on the assumption of ecological coherence within taxonomic groups (21). In contrast, changes in the abundances of specific functional genes within the soil microbial community during recovery can provide more direct insight into how environmental changes impact microbial processes that relate to ecosystem functions.

Because of the importance of nitrogen cycle, many studies have focused on functional genes involved in nitrogen cycling, most commonly using qPCR to quantify specific genes directly from mixed microbial communities in soil samples. With regard to ecosystem recovery, one study in a salt marsh system found that abundances of genes involved in N fixation (*nifH*) and denitrification (*nirS*, *nirK*) all peaked at intermediate succession (35 yr) (14). Studies in deglaciated forelands show different results, with one showing that denitrification (*nirS*) and ammonia oxidation (ammonia oxidizing bacteria-*amoA*) decreased with succession age (22) and another study showing the opposite pattern (23). In forest ecosystems, these processes can create a significant sink for greenhouse gases such as CO_2 and CH_4 (24) and a sink for N_2O under certain conditions (25). As a result, land use change and deforestation can contribute significantly to global greenhouse gases emissions (26), and the function as greenhouse gas sink may take >100 years to recover (27).

Advances in DNA sequencing methods in recent years have greatly increased the capability to characterize environmental microbial communities beyond what can be done using qPCR. However, the widely used approach of targeted 16S rRNA amplicon sequencing mainly focuses on characterizing dynamics of community structure including changes in diversity and taxonomic composition, while changes in ecological function must be inferred from these patterns (21). In contrast, shotgun metagenomic approaches sequence DNA randomly from all genetic material present in environmental samples, thus identifying the relative abundance of many functional genes directly (28, 29). They also provide a comprehensive view of the relative abundances of thousands of functional genes present in an environmental sample, including those difficult to target with qPCR because of high phylogenetic diversity (21). This new capability has provided increased understanding of microbial function and change in response to environmental variation or disturbance, especially in complicated environments like soil (30). As a result, this approach can be a more reliable assessment of changes in potential function among microbial communities, and thus provides valuable additional insights into the mechanisms by which microorganisms interact with the surrounding environment.

In this study, we analyzed the metagenomes of soil microbial communities in a chronosequence of reforested reclaimed mine sites spanning 6 to 31 years plus unmined reference sites. Our objectives were to determine: (a) whether changes in the taxonomic composition as characterized by metagenomics followed a similar pattern along chronosequence ages as was previously found using amplicon sequencing; (b) how the taxonomic composition and functional genes changed during ecosystem succession and what those changes indicated about the recovery of ecosystem functions; (c) how taxa and functional genes associated with nitrogen cycling and greenhouse gas emission reflect changes previously reported in these ecosystem processes at these sites.

4.2 Materials and methods

The soil samples for this study were collected in September 2014 in the same triplicate plots described in Chapter 3 following the same procedures. Sampling three plots from five different chronosequence sites (ages 6, 12, 22, 31 and unmined reference) resulted in 15 total metagenomes. The samples were transported to the lab on ice and stored at -80 °C until DNA was extracted from approximately 0.25 g soil using the PowerSoil Soil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). DNA concentrations were measured with Qubit 2.0 fluorometer (Invitrogen, USA).

For shotgun metagenomics, DNA concentrations from all samples were adjusted equally with PCR water (MoBio Laboratories, Carlsbad, CA, USA) and sheared to 300bp using an M220 Focused-ultrasonicator (Covaris, USA) in 50ul tubes for 90 seconds. All 15 DNA libraries were then multiplexed and sequenced on the Illumina Hiseq platform in a 101bp paired-end sequencing run at the Virginia Tech Biocomplexity Institute. The raw sequence reads were uploaded to MG-RAST (31)for quality control and annotation. The reads were annotated with the Genbank database for taxonomy and the Subsystem database for functional genes. Annotations were filtered with a minimum identity of 80%, a minimum aligned length of 20bp and an E-value cutoff of 1×10^{-5}, which is stricter than the default value. In the output files of MG-RAST, the annotations of functional genes were listed as enzymes encoded by the genes and classified to three levels of broad categories (level1, level2 and level3) in Subsystem database. The metagenome sequences can be accessed with project ID mgp16415 in the MG-RAST database.

Variations in the relative abundances of functional genes and phylogenetic composition among the samples were analyzed with both univariate and multivariate methods and visualized with the R program. The sequences for each category (level1, level2 and function) were rarefied to the minimum number of sequences (831,310) among all samples. Bray-Curtis distances between sample pairs were calculated based on species composition and visualized with non-metric multidimensional scaling (NMDS) using the 'metaMDS' function in package 'vegan' (32), and the phyla significantly correlated with the ordinations were selected with function 'envfit' and plotted. The Bray-Curtis distances were also calculated using functional gene relative abundances, with the significantly correlated level1 categories included in Figure 4.2b. The relative abundances of functional genes and taxa were averaged for each site, z-transformed and visualized as heatmaps using the R package 'pheatmap'. To analyze the variations of functional genes and taxa across chronosequence ages, generalized linear models were applied to test the significance of correlations between chronosequence ages and taxa/function with R function 'glm'.

4.3 Results

4.3.1 Shotgun metagenome sequencing results

Shotgun metagenome sequencing resulted in approximately 227 million total sequences, averaging about 15 million per sample. Among all the sequences, 2.81% sequences failed to pass quality control, 2.71% were identified as neither rRNA nor proteins, 6.61% were identified as rRNA, 20.8% were identified as annotated proteins, and 67.1% were identified as unknown proteins (Figure 4.1). The percentages are similar to those reported in other soil metagenome studies (21, 33, 34).

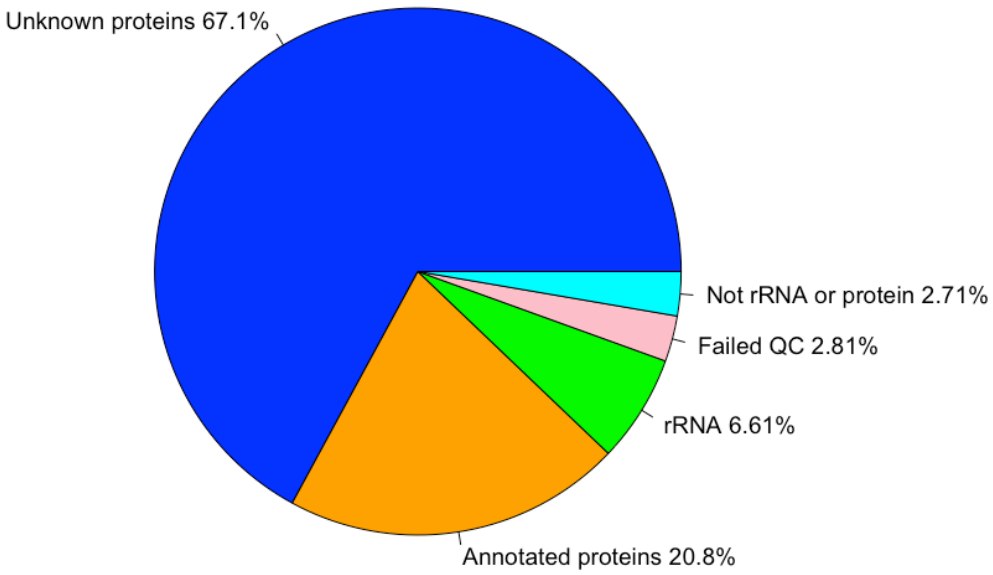


Figure 4.1. Averaged percentage of sequences assigned to predicted features.

4.3.2 Changes in OTU and functional gene relative abundances

To understand microbial changes with regard to both taxa and functions, we calculated the Bray-Curtis distances between sample pairs based upon both OTU and functional gene relative abundances, respectively, and visualized the patterns with NMDS. As shown in Figure 4.2a, the patterns of β-diversity based upon OTU relative abundances generally transition along the chronosequence ages from 6 yr to 31 yr, which is consistent with the patterns of these same communities characterized with 16S amplicon sequencing and presented in Chapter 3. The relative abundances of phyla Actinobacteria, Acidobacteria, Bacteroidetes, Verrucomicrobia, Alphaproteobacteria, Deltaproteobacteria and Gammaproteobacteria were significantly

correlated with the ordinations of samples (P = 0.001-0.012) (Table 4.1), with Actinobacteria and Bacteroidetes being more abundant in earlier ages and Acidobacteria more abundant in later ages.

In contrast, the ordination plot based on relative abundances of functional genes shows a different pattern (Figure 4.2b); the unmined sites were more similar to 22 yr sites and shifted away from the 6 - 31 yr trajectory. The changes in relative abundance of some level1 function categories were significantly correlated with the ordinations (P = 0.001-0.043), including respiration (Res), Virulence, Disease and Defense (VDD), Stress Response (SR), Phages, Prophages, Transposable elements, Plasmids (PPTP), Cofactors, Vitamins, Prosthetic Groups, Pigments (CVPP), Fatty Acids, Lipids, and Isoprenoids (FLI), Nitrogen Metabolism (NM), Phosphorus Metabolism (PhM) and Protein Metabolism (PrM) (Table 4.2). PPTP, SR and VDD increased with chronosequence ages, while CVPP and FLI decreased, and Res was higher at unmined sites. The correlation coefficients and P-values are shown in Table 4.1 and Table 4.2.

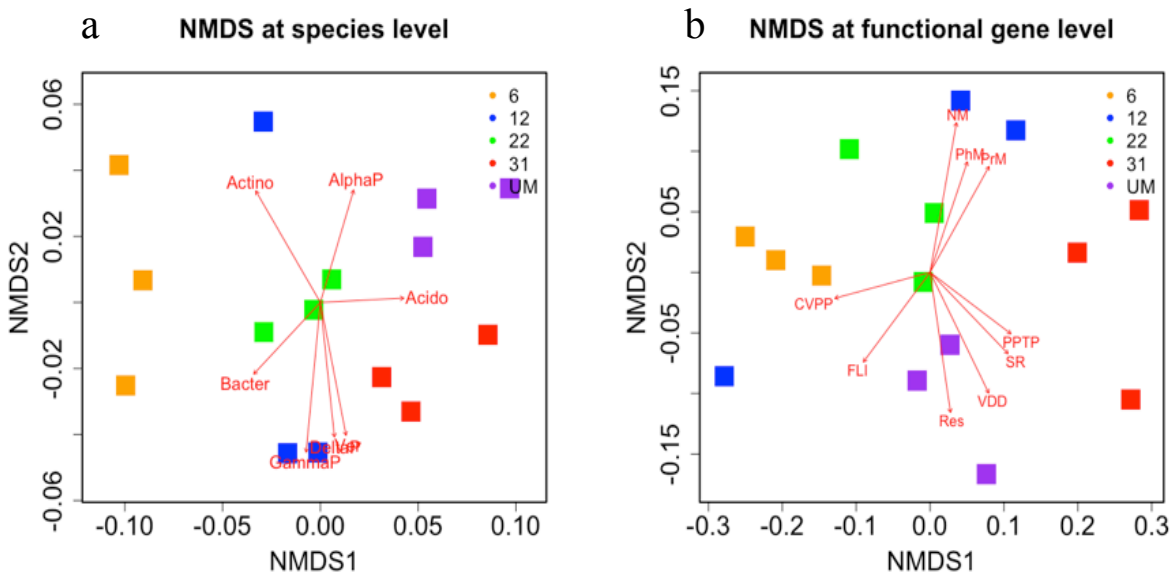


Figure 4.2. NMDS analyses based on species relative abundance (a) and functional gene relative abundance (b). Keys to the abbreviations were shown in Table 4.1 and Table 4.2.

Table 4.1. Results of correlation analyses between taxa and β-diversity patterns (*P*-values based on 999 permutations).

Taxa	Abbreviations	r^2	P-value	
Acidobacteria	Acido	0.6598	0.003	**
Actinobacteria	Actino	0.8066	0.001	***
Bacteroidetes	Bacter	0.5893	0.008	**
Verrucomicrobia	Ver	0.6477	0.002	**
Alphaproteobacteria	AlphaP	0.5231	0.012	*
Betaproteobacteria	BetaP	0.2221	0.235	
Deltaproteobacteria	DeltaP	0.6189	0.002	**
Gammaproteobacteria	GammaP	0.7618	0.001	***

Table 4.2. Results of correlation analyses between Level1 functional category and functional gene shifting patterns (*P*-values based on 999 permutations).

Level1 functional category	Abbreviation	r^2	P-value	
Amino Acids and Derivatives	AAD	0.4235	0.043	*
Carbohydrates	CH	0.2979	0.126	
Cell Division and Cell Cycle	CDCC	0.4987	0.012	*
Cofactors, Vitamins, Prosthetic Groups, Pigments	CVPP	0.8829	0.001	***
DNA Metabolism	DM	0.1694	0.354	
Fatty Acids, Lipids, and Isoprenoids	FLI	0.7039	0.002	**
Membrane Transport	MT	0.1504	0.362	
Metabolism of Aromatic Compounds	MAC	0.1648	0.338	
Nitrogen Metabolism	NM	0.8553	0.001	***
Nucleosides and Nucleotides	NN	0.5651	0.013	*
Phages, Prophages, Transposable elements, Plasmids	PPTP	0.7532	0.001	***
Phosphorus Metabolism	PhM	0.5623	0.008	**
Protein Metabolism	PrM	0.7263	0.001	***
RNA Metabolism	RM	0.2978	0.123	
Respiration	Res	0.7313	0.004	**
Stress Response	SR	0.8104	0.001	***
Sulfur Metabolism	SM	0.02	0.879	
Virulence, Disease and Defense	VDD	0.8397	0.001	***

4.3.3 Variations among high taxonomic levels across ages

Based on similarities to rRNA entries in the MG-RAST database, 97.64% of sequences were classified as Bacteria, 0.59% were classified as Archaea, 1.61% were classified as Eukaryota and 0.02% were classified as viruses. The dominant microbial phyla or proteobacterial classes were Alphaproteobacteria, Actinobacteria and Betaproteobacteria, which respectively represent 34.1%, 18.5% and 12.5% of all rRNA genes (Figure 4.3). Acidobacteria (6.8%) represents relatively less of the shotgun metagenome database than it did in the amplified 16S rRNA sequences in Chapter 3, where it was around 21.5%, while Alphaproteobacteria were abundant when characterized with both methods.

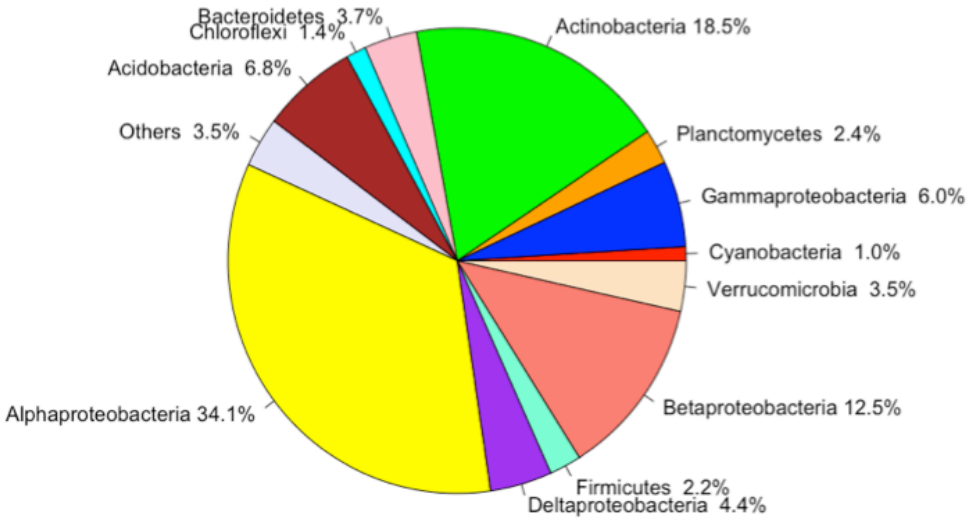


Figure 4.3. Averaged relative abundance of phyla or classes (within Proteobacteria) across all samples.

The heatmap in Figure 4.4 shows the variations of dominant microbial phyla and proteobacterial classes (relative abundance > 1%) across chronosequence ages. Actinobacteria were more abundant at 5 yr sites, while Alphaproteobacteria were more abundant at the unmined reference sites. More generally, the relative abundance of Bacteroidetes was higher at earlier ages while that of Acidobacteria was higher at later ages (Figure 4.4). To further test these trends, linear regression was performed to quantify the significance of the taxonomic variations across ages (6 - 31 yr) (Figure 4.5). The relative abundance of Actinobacteria decreased significantly with age ($P=0.0036$), while the relative abundances of Acidobacteria, Deltaproteobacteria, Chlorobi, Euryarchaeota, Aquificae, Nitrospirae, Therotogae and Firmicutes increased significantly ($P=0.0009$-0.0358). Because of the dominance of Alphaproteobacteria, I also analyzed order-level variation in this phylum, including Caulobacterales, Sphingomonadales, Parvularculales, Rhodobacterales, Rickettsiales, Rhizobiales and Rhodospirillales (Figure 4.6). Among these, the relative abundances of Caulobacterales and Sphingomonadales decreased significantly with ages

(P=0.0005-0.0199), while that of Rhodospirillales increased significantly (P=0.0008) (Figure 4.7).

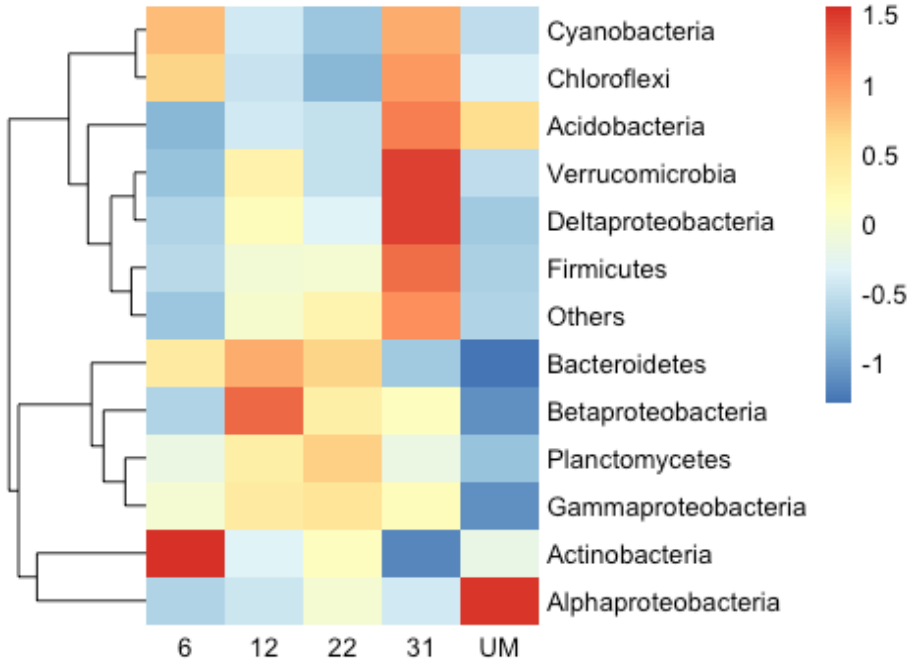


Figure 4.4. The heatmap shows changes in the relative abundance of bacterial phyla or proteobacterial classes over succession ages. The key represents the z-scores of the relative abundance of the taxa.

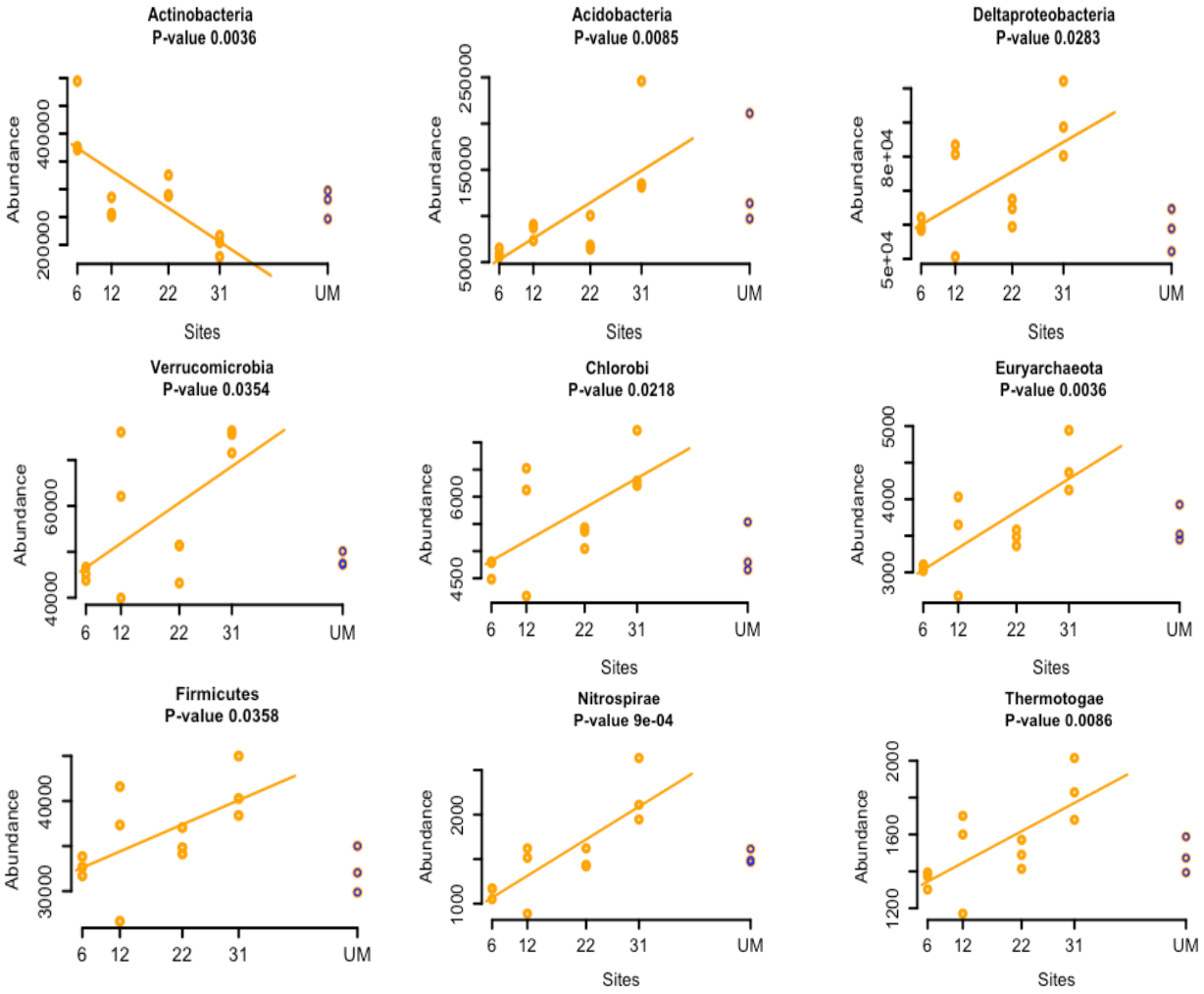


Figure 4.5. The phyla or classes (within Proteobacteria) with shifting trends significantly correlated with succession ages.

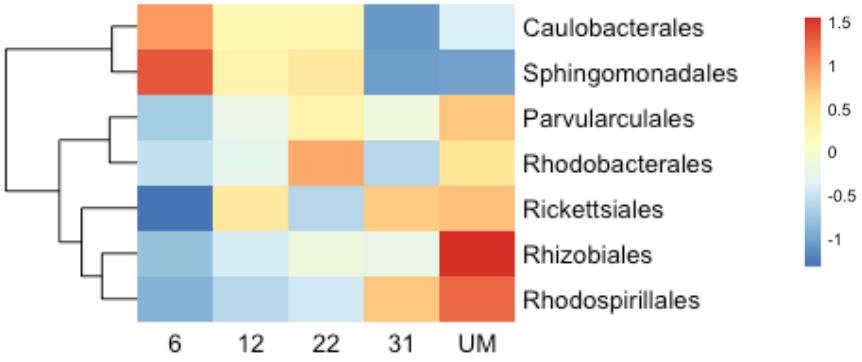


Figure 4.6. The heatmap shows changes in the relative abundance of alphaproteobacterial orders across succession ages. The key represents the z-scores of the relative abundance of the taxa.

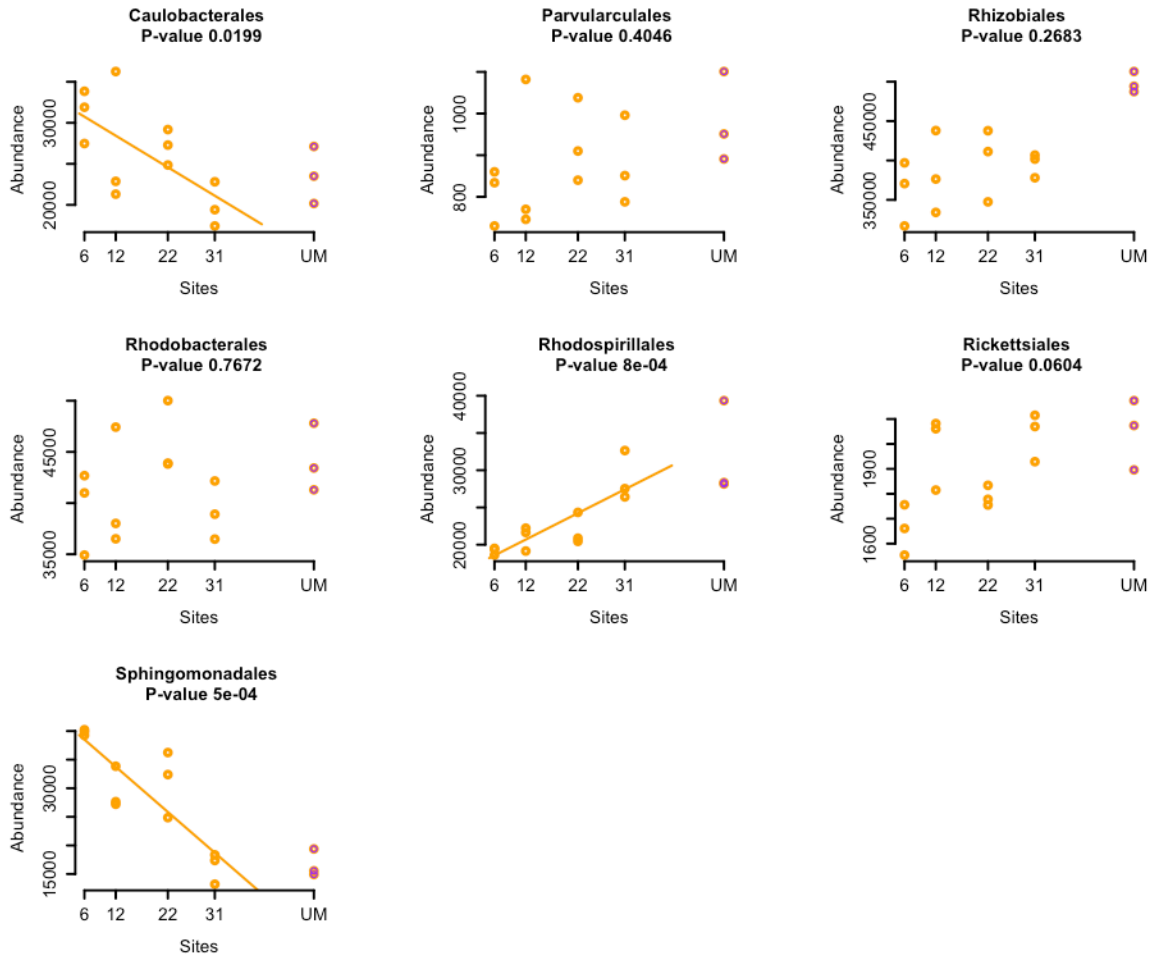


Figure 4.7. Linear regression between the relative abundance of alphaproteobacterial orders and chronosequence ages.

4.3.4 Variations of broad functional categories across ages

The functional genes were clustered into 21 level 1 categories based on the Subsystem database. The dominant categories were Clustering-based subsystems (CBS), Carbohydrates (Ch), Protein Metabolism (PrM) and Amino Acids and Derivatives (AAD), representing 14.2%, 11.8%, 10.4% and 10.3% of the total annotated genes (Figure 4.8). Among all the categories, Respiration (Res), Virulence, Disease and Defense (VDD), Stress Response (SR), Phages, Prophages, Transposable elements, Plasmids (PPTP) were more abundant in later stages (Figure 4.9). Photosynthesis (Phs)

genes had high relative abundances at 6 yr sites, while the relative abundances of genes encoding for Potassium metabolism (PoM) and Metabolism of Aromatic Compounds (MAC) were higher at unmined sites. Among these categories, the relative abundance of PPTP, SR, VDD and Membrane Transport (MT) increased significantly with age (P=0.0024-0.0182), while the relative abundance of FLI, CBS and CVPP decreased significantly (P=0.0004-0.399) (Figure 4.10). The level2 categories were also examined to find out those changing significantly with chronosequence ages. The relative abundances of genes related to Cell division, CO_2 fixation and Monosaccharides metabolism decreased significantly, while genes related to Polysaccharides, Cold shock, Oxidative stress, Pathogenicity islands, Phages, Prophages, Resistance to antibiotics and toxic compounds, and Toxins and superantigens increased significantly (P=0.001-0.0191) (Figure 4.11).

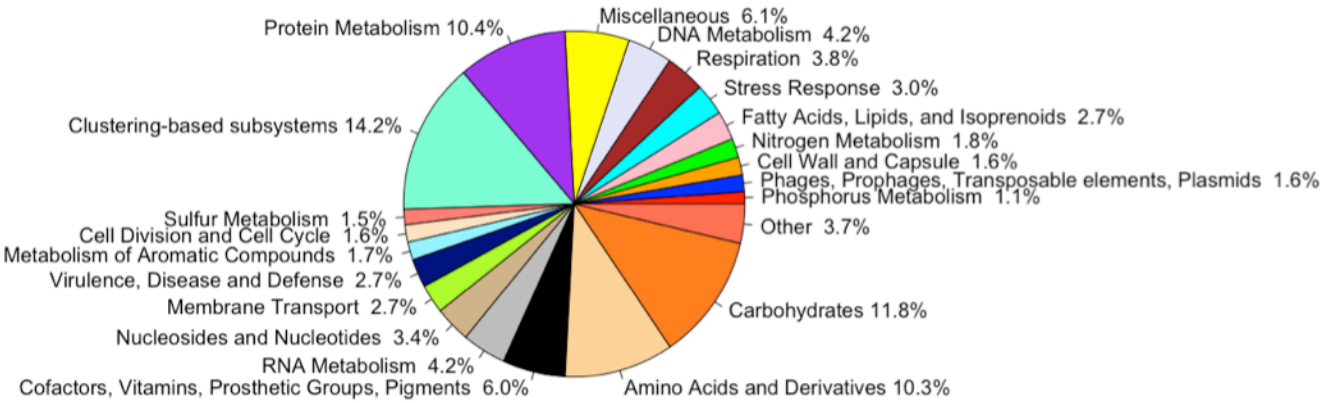


Figure 4.8. Averaged composition of level1 functional category across all samples.

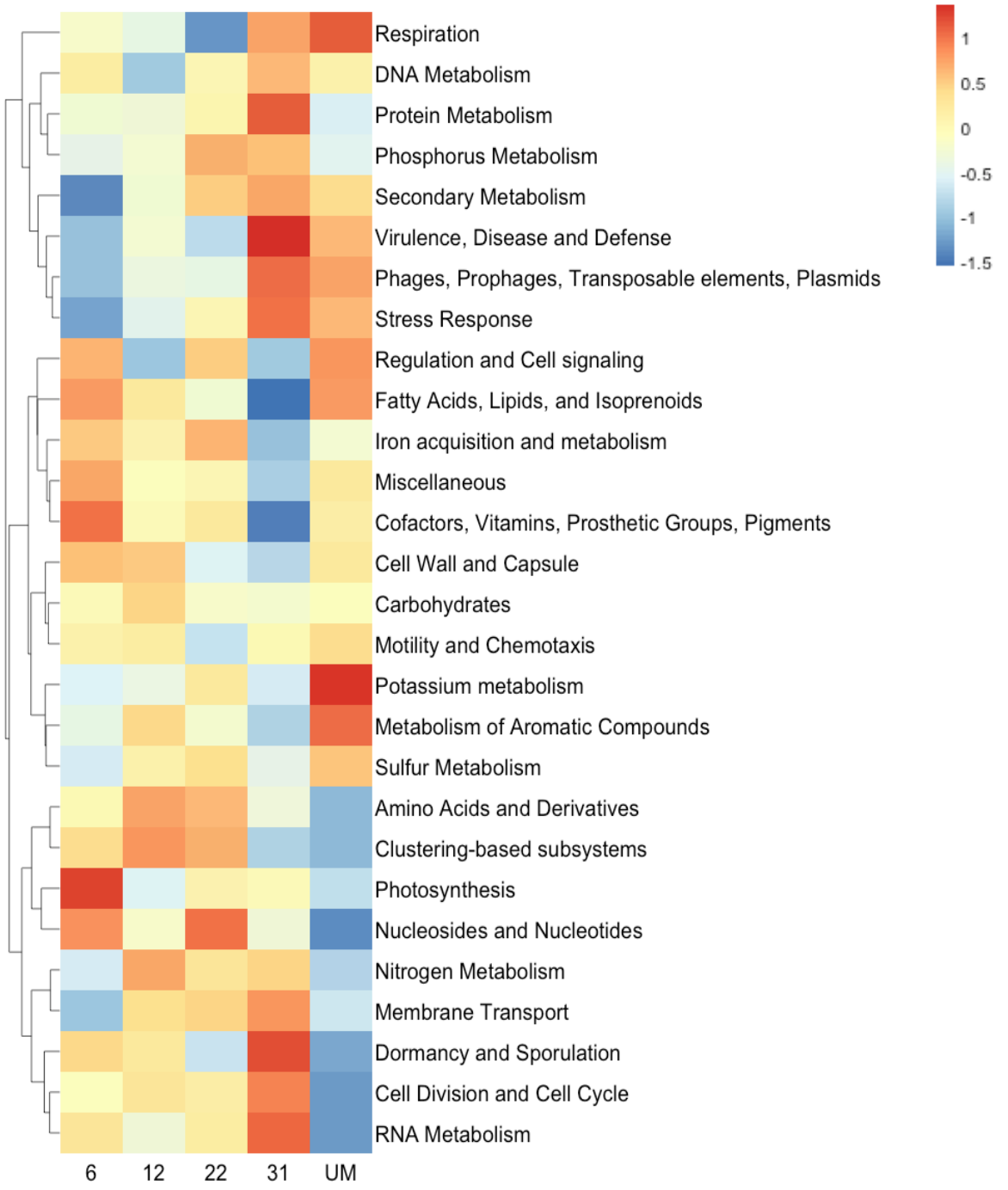


Figure 4.9. The heatmap shows the variations of level1 functional category across succession ages. The key represents the z-scores of the category abundance.

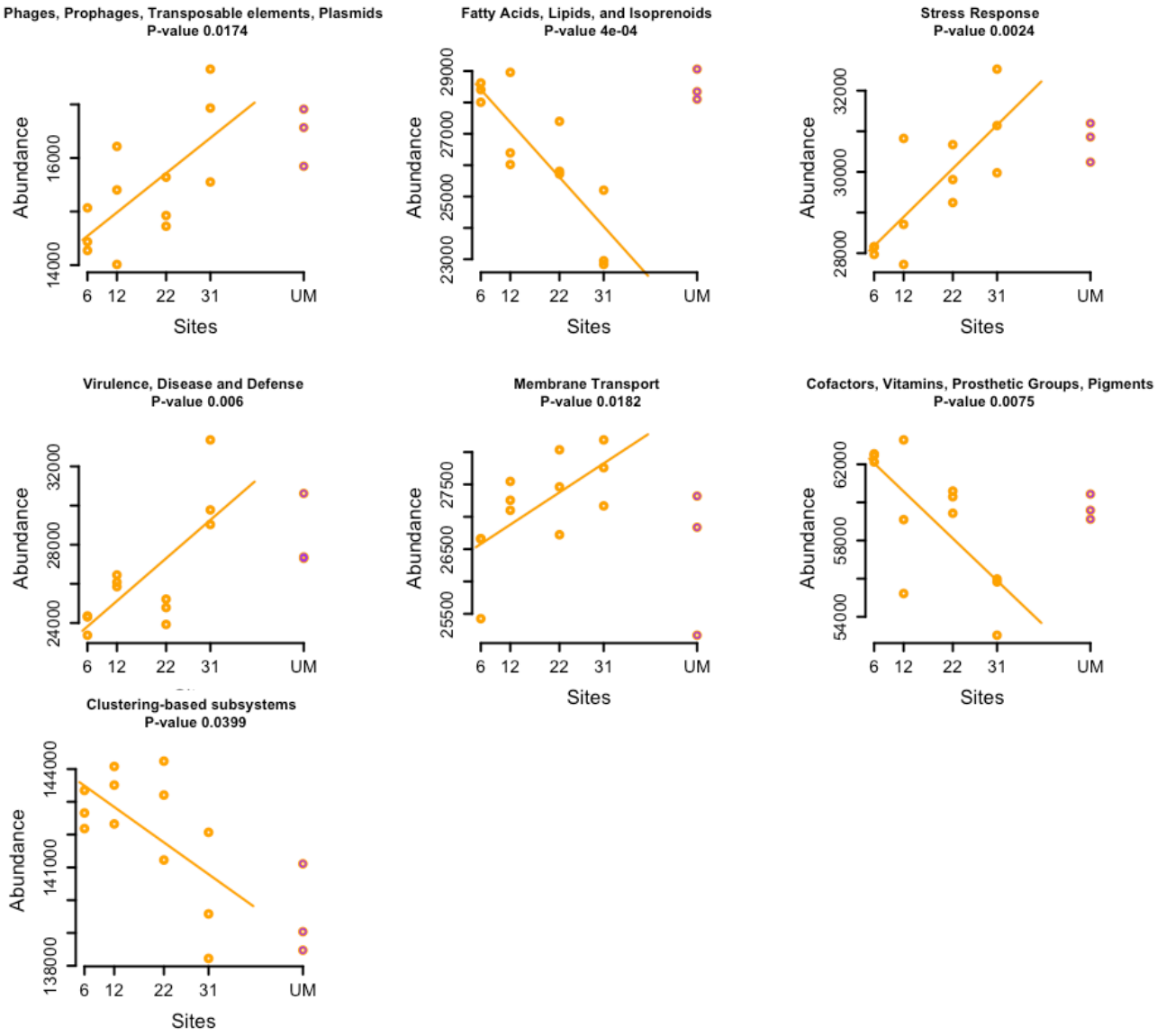


Figure 4.10. The level1 functional category with shifting trends significantly correlated

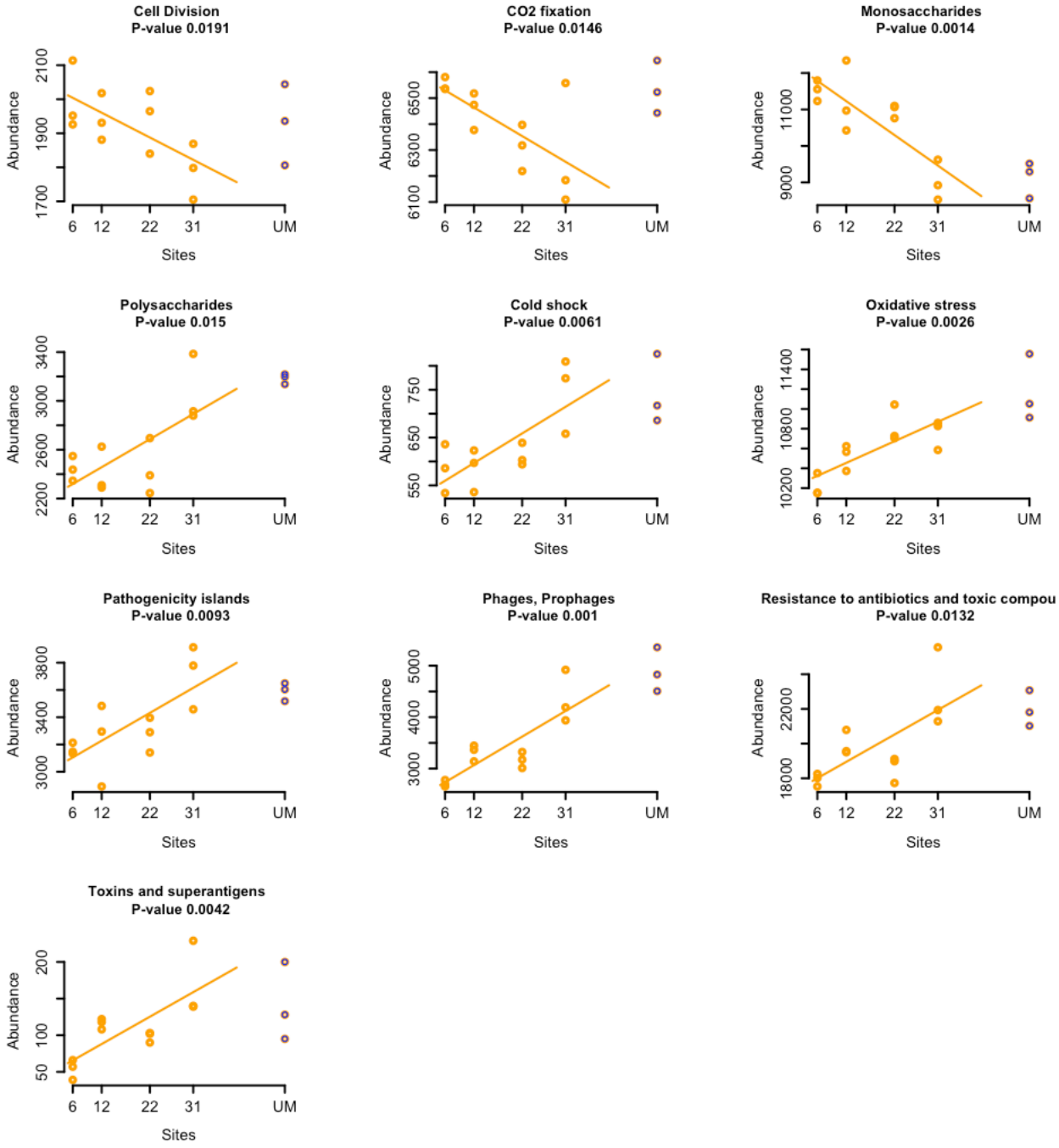


Figure 4.11. The level2 functional category with shifting trends significantly correlated with succession ages.

4.3.5 Variations taxa and functional genes involved in nitrogen cycle

In order to understand the potential change of microbial nitrogen cycling in the chronosequence, we specifically focused on the relative abundances of some taxa and functional genes known to affect nitrogen transformations. For nitrification, ammonia-oxidizing bacteria (AOB) and nitrite-oxidizing bacteria (NOB) both increased significantly with chronosequence age (P=0.026 and 0.0098), but the relative abundance of NOB was the highest at the unmined sites while the relative abundance of AOB at unmined sites were lower than 22- and 31-yr sites (Figures 4.12 & 4.14). This was also further complicated because taxa within the same functional group showed different trends with chronosequence age. For example, ammonia oxidizing microbes (relative abundance > 0.005% of all sequences), *Nitrosomonas* and *Nitrosococcus* (AOB) increased significantly from 6 to 31 yr sites (P=0.030 and 0.0072) and the trends of *Nitrosospira* (AOB) were close to being statistically significant (P=0.053), but they were all less abundant at unmined sites (Figure 4.14). In contrast, *Nitrosopumilus* (the most abundant ammonia-oxidizing archaea (AOA)) reached the highest relative abundance at the unmined sites (Figure 4.12). Furthermore, patterns of functional gene change were also different, with the relative abundance of ammonia monooxygenase, the key functional gene in the oxidation of ammonium to begin nitrification, more abundant at 5 yr sites than all other sites (Figure 4.14).

Within the NOB, the relative abundance of *Nitrobacter* increased significantly from 6 to 31 yr sites (P=0.020) and reached the highest proportion in unmined sites, while *Nitrococcus*, *Leptospirillum* and *Thermodesulfovibrio* also increased significantly from 6 to 31 yr sites (P=0.00097, 0.0021 and 0.00019) but were then less abundant at unmined sites (Figure 4.12 & 4.14). However, *Nitrobacter* was the most dominant NOB, with a relative abundance of one order of magnitude higher than others, so its shift mainly drove the overall NOB trend. Nitrate

reductase (EC 1.7.99.4), which is also known as nitrite oxidoreductase and is involved in both nitrification and denitrification, was more abundant at the earlier stages (Figure 4.13).

For denitrification, the high phylogenetic diversity of this group and the wide distribution of nitrite reductase genes make it difficult to provide a complete list of denitrifiers, with > 50 potential denitrifying genera belonging to 12 phyla (35, 36). Instead, genes involved in denitrification have been mainly used to estimate the potential activity of denitrifiers in the environment. Across the chronosequence ages, nitrate and nitrite transporters and nitrite reductase (EC 1.7.1.4) were generally higher in relative abundance at earlier stages than later stages (Figure 4.13), which may imply more active denitrification in the earlier stages. The large subunit of nitrite reductase [NAD(P)H], which is the major part of nitrite reductase, also decreased significantly with age (P = 0.0021). In contrast, nitrogenase (EC1.18.6.1), which metabolizes nitrogen fixation, was highest at 30 yr sites and lowest at the unmined sites.

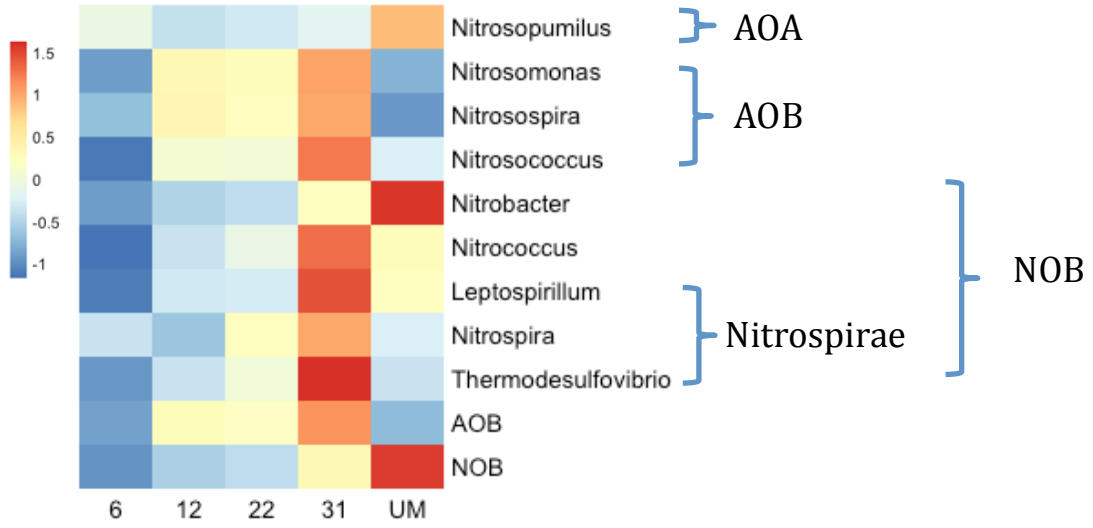


Figure 4.12. The heatmap shows the variations of genera identified as AOA, AOB, NOB (abundance > 0.0005%). The key represents the z-scores of the relative abundance of the taxa.

Table 4.3. The taxonomy of genera identified as AOB and NOB.

Kingdom	Phylum	Class	Order	Family	Geus
Archaea	Thaumarchaeota	incertae sedis	Nitrosopumilales	Nitrosopumilaceae	Nitrosopumilus
Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae	Nitrosomonas
Bacteria	Proteobacteria	Betaproteobacteria	Nitrosomonadales	Nitrosomonadaceae	Nitrosospira
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Nitrobacter
Bacteria	Proteobacteria	Gammaproteobacteria	Chromatiales	Ectothiorhodospiraceae	Nitrococcus
Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Leptospirillum
Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira
Bacteria	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Thermodesulfovibrio

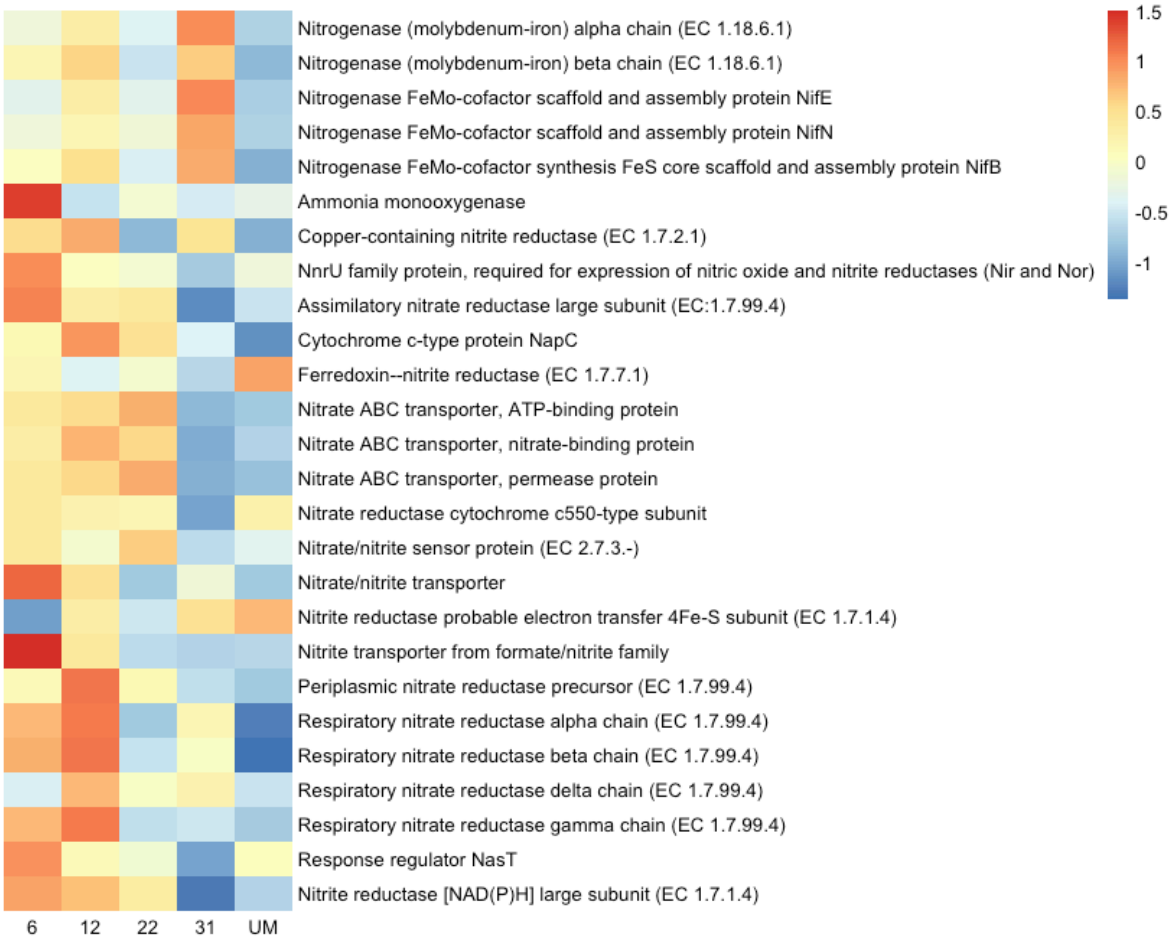


Figure 4.13. The heatmap shows the variations of functional genes involved in nitrogen cycle. The key represents the z-scores of the category abundance.

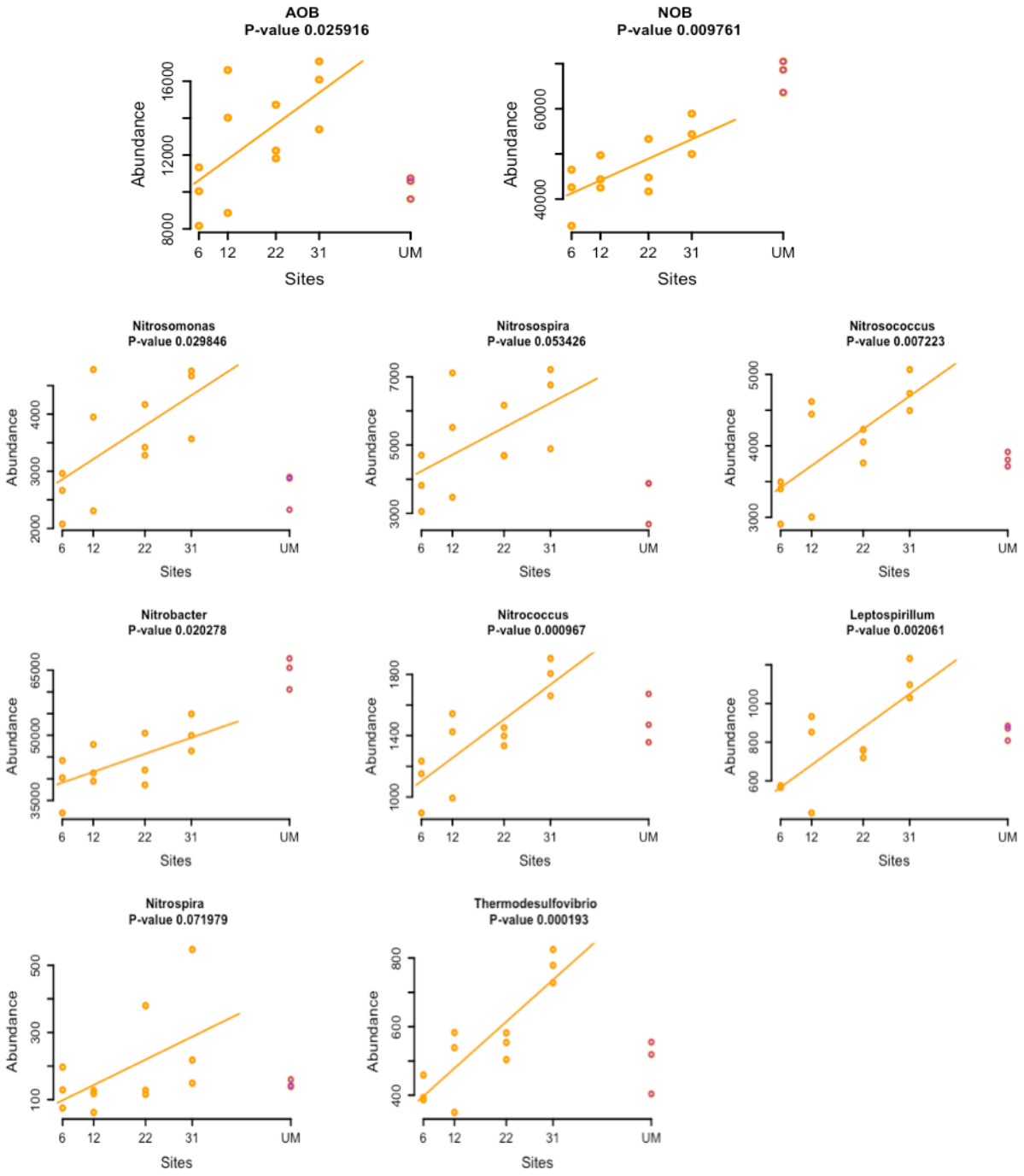


Figure 4.14. AOB, NOB and the genera involved in nitrogen cycle whose shifting trends were significantly correlated with succession ages.

4.3.6 Variations of taxa and functional genes involved in greenhouse gas emissions

Methanotrophs, which metabolize methane for carbon source and energy, comprised 0.30% of all sequences, while the methanogens, which use CO_2 as an electron acceptor during anaerobic respiration and produce methane, comprised 0.013% of all sequences. Within the characterized methanotrophs, *Methylocella*, *Methylobacterium*, *Methylocystis* and *Methylosinus* were most abundant at the unmined sites, and they were also more abundant at 31 yr than earlier stages. In contrast, several other methanotrophic genera were more abundant at 12-31 yr sites (Figure 4.15). However, *Methylocella*, *Methylobacterium*, *Methylocystis* and *Methylosinus* comprised 80.5% of all methanotrophs, which may explain the overall observation that unmined sites operate best as methane sinks (37). For methanogens, they were overall most abundant at 31 yr sites but in much lower relative abundances than the methanotrophs.

With regard to functional genes related to methane cycling, no methyl-coenzyme M reductase genes (the primary gene involved in methanogenesis) were detected, so it appears that changes in methanogenesis may contribute less overall to net fluxes of methane than changes in methanotrophy. In this regard, methane monoxygenase, the functional gene metabolizing methane during methanotrophy, was most abundant in unmined sites (Figure 4.16), which further supports that methane sequestration was only active at the unmined sites. With regard to other greenhouse gases, carbon monoxide dehydrogenase, an enzyme involved in CO_2 production, was less abundant at 30 yr sites, but otherwise there was no difference among other sites (Figure 4.16). Similarly, nitric oxide reductase and nitrous oxide reductase, which are involved in N_2O production, did not show strong significant trends across ages (Figure 4.16).

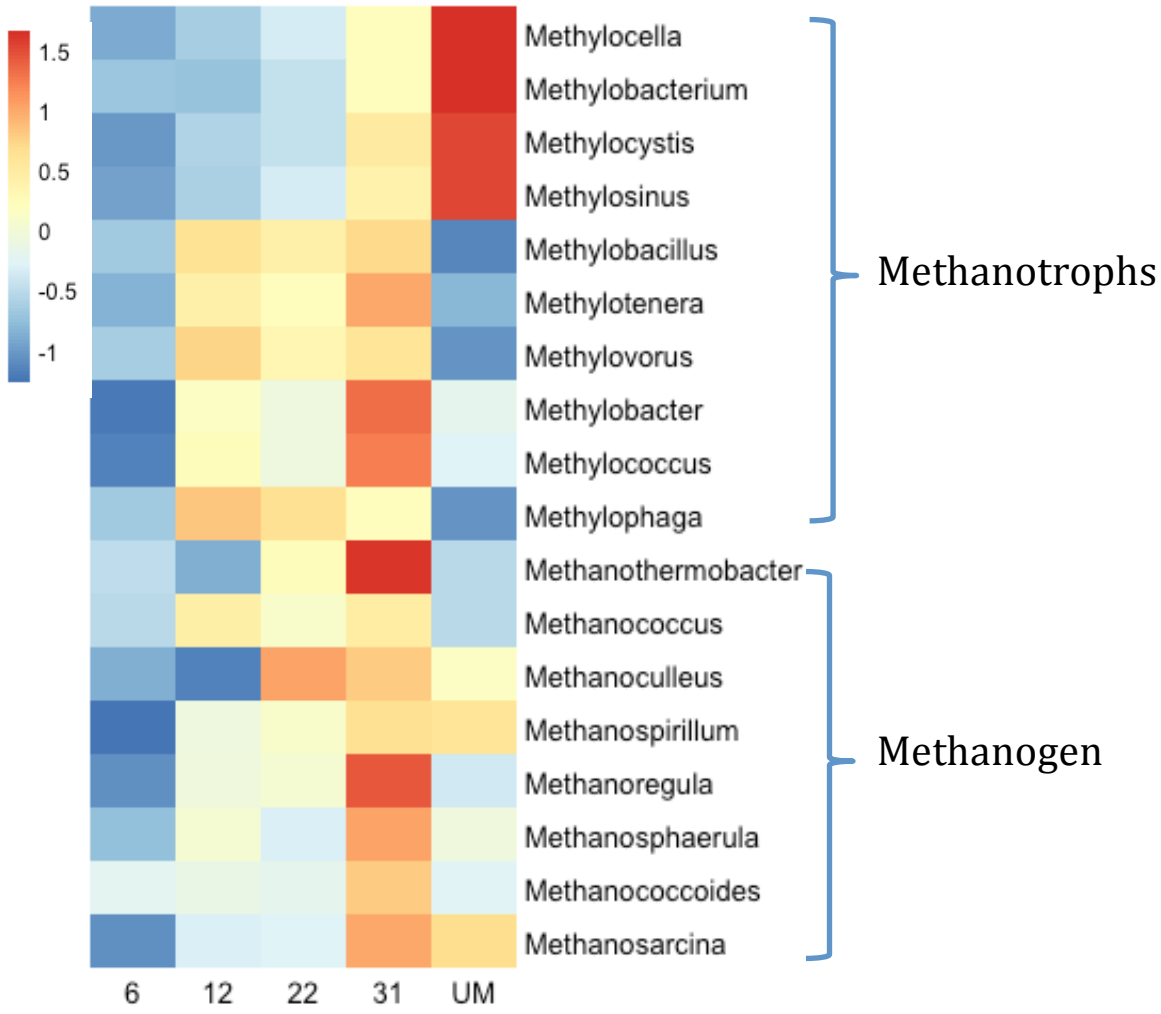


Figure 4.15. The heatmap shows the variations of genera identified as methanogens and methanotrophs (abundance > 0.0005%). The key represents the z-scores of the relative abundance of the taxa.

Table 4.4. The taxonomy of genera identified as methanogens and methanotrophs.

Kingdom	Phylum	Class	Order	Family	Genus
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	Methylocella
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylocystaceae	Methylocystis
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylocystaceae	Methylosinus
Bacteria	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylobacillus
Bacteria	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylotenera
Bacteria	Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylovorus
Bacteria	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylococcaceae	Methylobacter
Bacteria	Proteobacteria	Gammaproteobacteria	Methylococcales	Methylococcaceae	Methylococcus
Bacteria	Proteobacteria	Gammaproteobacteria	Thiotrichales	Piscirickettsiaceae	Methylophaga
Archaea	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae	Methanothermobacter
Archaea	Euryarchaeota	Methanococci	Methanococcales	Methanococcaceae	Methanococcus
Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanomicrobiaceae	Methanoculleus
Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Methanospirillaceae	Methanospirillum
Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Unclassified	Methanoregula
Archaea	Euryarchaeota	Methanomicrobia	Methanomicrobiales	Unclassified	Methanosphaerula
Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae	Methanococcoides
Archaea	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae	Methanosarcina

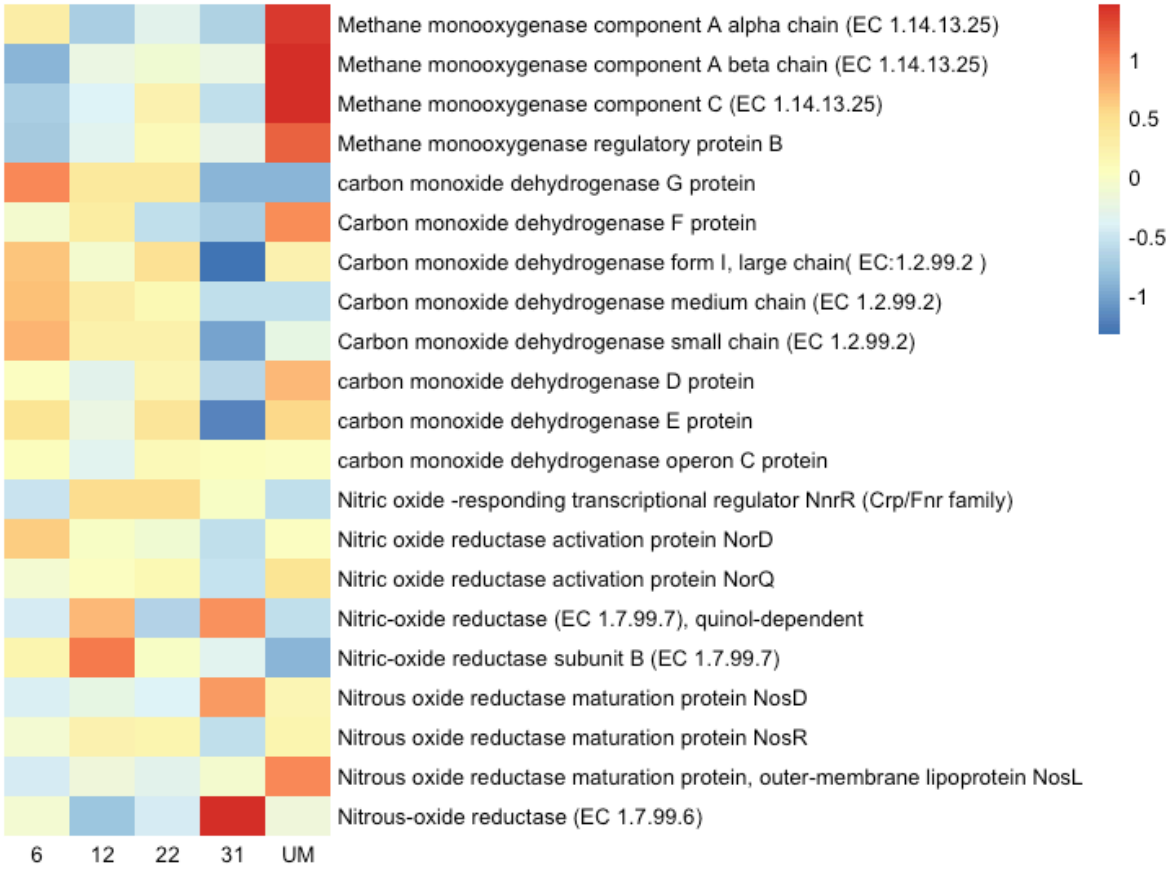


Figure 4.16. The heatmap shows the variations of functional genes involved in the production of greenhouse gas. The key represents the z-scores of the functional gene abundance.

4.4 Discussion

4.4.1 Taxonomic change during succession

The β-diversity patterns and taxonomic changes, as characterized by shotgun metagenome sequencing, were generally similar to those identified using amplicon sequencing in Chapter 3. Based on Bray-Curtis distances among the samples, microbial communities were again found to be separated along chronosequence age, with the community structure becoming more similar to unmined references as age increased (Figure 3.2a). Given that the current metagenome approach directly sequenced the DNA extracted from soil samples without further amplification, the resultant data should be more accurate in terms of relative abundance, without potential bias from amplicon sequencing. As a result, this work has helped verify the broad patterns describing recovery of a complex soil microbial community during the restoration of a forest ecosystem.

Significant increases of Acidobacteria, Verrucomicrobia and Nitrospirae (38-41) coupled with significant decreases of Actinobacteria (33, 42) also supported the shift from copiotrophic groups to oligotrophic groups described in Chapter 3, which also agrees with the increasing soil C:N ratio that was previously reported at these sites (37). Again, this suggests that the recalcitrant C compounds increased and labile C substrates decreased with chronosequence age, which was also supported by the increase of genes involved in polysaccharide metabolism and the decrease of genes involved in monosaccharide metabolism (Figure 4.11). Because of the high relative abundance of Alphaproteobacteria, we further examined the shifting trends of orders belonging to Alphaproteobacteria. Among the 7 alphaproteobacterial orders, Caulobacterales and Sphingomonadales significantly decreased while Rhodospirillales increased significantly. Given that Sphingomonadales are known to interact with plants and utilize root exudates (43),

development of the tree canopy and reduced cover of grasses in later ages could explain the decrease of Sphingomonadales, given that root biomass has been shown to decrease with reduced grass dominance in other systems (44) and grass rooting occurs more closely to surface (where these samples were collected) compared with deeper tree rooting in mine soils (45) . In contrast, Rhodospirillales are abundant in forest soils compared with grassland soils (46) and are generally oligotrophs that are negatively correlated with soil nutrients such as total dissolved nitrogen and dissolved organic carbon (47). Overall, our results support that the change of soil nutrients, especially the ratio of labile:recalcitrant carbon and change from grass- to tree-dominated vegetation, are important factors driving microbial succession in the recovery of forest ecosystems.

4.4.2 Functional change during succession

Changes in relative abundances of functional genes also indicated that the availability of resources and microbial interactions play important roles in microbial succession. For example, significant increases in relative abundances of genes were found related to Phages, Prophage, Transposable elements and Plasmids (PPTP), Stress Response (SR), Virulence, Disease and Defense (VDD) and Membrane Transport (MT), and significant decreases of Fatty Acids, Lipids, and Isoprenoids (FLI), Clustering-based subsystems (CBS) and Cofactors, Vitamins, Prosthetic Groups, Pigments (CVPP). Soil is an important reservoir of phages, prophages and transposable elements (48, 49), and these mobile genetic elements play important roles in shaping microbial communities (50, 51). For example, the transposable elements are essential in the construction and spread of catabolic pathways among bacteria, which lead to a more rapid adaptation of the bacterial community to new xenobiotics (52, 53). This suggests that the increase of PPTP during succession could result from the adaption of bacterial community to the appearance of new

compounds in litter deposition due to the vegetation development. It also suggests that the activity of transposable elements could be a response to stresses on the community such as nutrient deficiency or temperature change (54, 55), which explains the concurrent increase of PPTP and SR.

The relative abundance of genes involved in VDD could serves as an indicator of stress-related environments (56), which could be related with nutrient deficiency from increased competition within microbial community. For example, bacteria have been shown to increase production of antibiotics and bacteriocin when detecting nutrient competition (57). In this study, among genes involved in VDD, the relative abundance of genes involved in toxins, superantigens, and resistance to antibiotics and toxic compounds significantly increased with chronosequence age, which indicates that competition increased among microorganisms for nutrients during succession (58). Transporter proteins are more abundant in low nutrient conditions in order to increase the chance of substrate acquisition (59, 60), which could explain the increase of MT with age. Genes involved in secondary metabolism also increased in later stages, and synthesizing and secreting secondary metabolites are another mechanism involved in virulence and defense (56). When viewed together, these broad trends also suggest that interactions among microorganisms become more abundant and complex as ecosystems develop, which agrees with the increased complexity inherent in the interaction networks presented in Chapter 3. It also further supports that the increased complexity could be related to mechanisms for dealing with the nutrient deficiency stress.

Other metagenome changes also provided insights into changes in microbial processes that occur during ecosystem succession. Photosynthesis-related genes were most abundant at 6 yr sites, where the undeveloped tree canopy allowed the most light at the soil surface. Genes involved in

the metabolism of aromatic compounds (MAC) were most abundant at unmined sites, where the buildup of recalcitrant organic carbon was also most abundant (37). Respiration-related genes also increased with succession age, which are consistent with the observation by Avera et al. that both substrate induced respiration and total microbial biomass increased with chronosequence age (37). In addition, others have observed that soil microbial respiration rates decrease when the microbial community shifts from oligotrophic to copiotrophic groups following N addition (61-63). Although copiotrophs usually have higher activity and higher turnover rates, they are less likely to access the recalcitrant C pool, which is the major part of soil organic carbon (33). Thus, the respiration rate of the overall microbial community is expected to increase across chronosequence ages, with the decrease of copiotrophs and increase of oligotrophs. Coupled with this change, genes involved in cell division decreased along chronosequence age, which also suggests a decrease of copiotrophs, as they tend to have higher growth rates.

4.4.3 Potential changes in nitrogen cycle during succession

AOB and NOB both increased significantly with succession age, which agrees with the corresponding significant increase of total N reported by Avera et al. (37). AOA reached the highest relative abundance at unmined sites, differing from the trends of AOB, which suggests an AOB-AOA shift with age. This agrees with previous reports that AOB usually inhabit ammonium rich environments, while AOA are less dependent on ammonium concentrations (64), suggesting these two groups could be respectively *r*- and *k*-selected. Although the total N was the highest at unmined sites (37), N in soil tends to get embedded in complex organic compounds over time, after which soil nitrogen bioavailability highly depends on the depolymerization of N bound in soil organic matter (SOM) (65).

Similar taxonomic shifts also occurred within the NOB. *Nitrobacter*, *Nitrococcus* and *Nitrospirae* all increased with chronosequence age, but *Nitrobacter* was highly abundant at unmined sites while the others were not. *Nitrobacter* has a lower nitrite affinity than *Nitrospira*, indicating that *Nitrobacter* is more adapted to environments with higher nitrite concentration (66). This shift suggests that the unmined sites have a higher concentration of nitrite, which could also be related to the shift from AOB to AOA in unmined sites. Although the contribution of archaea and bacteria to ammonia oxidation is largely unknown, archaeal *amoA* genes have been found to be more abundant than bacterial *amoA* genes in most soil environments, which agrees with these results (67, 68).

In contrast, functional genes involved in ammonia and nitrite oxidation showed different trends. The relative abundance of ammonia monooxygenase genes was highest in the 6 yr sites, with no obvious differences among other sites. However, the detected abundance of ammonia monooxygenase was extremely low, averaging only six copies of genes for each sample, while ~10^4 copies of AOB 16S rRNA were detected. Thus, the trends need further verification. Nitrite oxidoreductase was generally higher in the earlier chronosequence ages, but it is important to remember that it metabolizes the nitrite-nitrate transformation in both directions. As a result, it is difficult to distinguish its contribution to nitrification specifically. Nitrite oxidoreductase is also less intensely studied because of its high phylogenetic diversity, and the 16S rRNA of nitrifiers are usually used for estimating nitrite oxidation in the environment on a molecular basis (69).

Because of the broad phylogenetic diversity of denitrification, it is difficult to estimate the abundance of denitrifiers from taxonomic data sets. However, genes related with nitrite reduction generally decreased with chronosequence age, implying higher denitrifying activity in younger ages, which is consistent with previous studies (22, 70). Finally, although the change of

functional genes provided considerable information about potential shifts in nitrogen cycle, the abundance of a gene does not equal to enzyme activity, so additional experimental transcriptomic or enzymatic work would be required to verify mechanisms.

4.4.4 Changes in greenhouse gas emission during succession

The low relative abundances of methanotrophs and methane monooxygenase genes at all sites except the unmined sites are in agreement with the finding of Avera et al. that the reforested soils at these same sites have not recovered their function as a methane sink (37). Methane is an important greenhouse gas, and its emissions provide the second largest contribution to global warming following carbon dioxide (71). Temperate forest soils work as an important methane sink, contributing to around 30-50% of methane oxidation (72). Land use change can alter methanotrophic communities (73), and deforested soils can switch to a net source of methane from a net sink (74). This work confirms that reversing these trends could take significant time (i.e., decades) following reforestation.

Methanotrophs have been divided into type-I (e.g. *Methylococcus* and *Methylobacter*) and type-II (e.g. *Methylosinus and Methylocystis*), and they showed different trends in our sites, with a shift from type-I to type-II with age, which is consistent with a previous study (72). The trends of type-II methanotrophs were consistent with methane monoxygenase, both reaching the highest relative abundance at the unmined sites, which agrees with the observation that the relative abundance of type-II methanotrophs were significantly correlated with methane oxidation (72). The relative abundance of type-II was much higher than type-I (80% vs. 20% of all methanotrophs), and the four dominant type-II methanotrophs were all members of class

Alphaproteobacteria, while the 6 dominant type-I methanotrophs belong to Betaproteobacteria and Gammaproteobacteria.

Although all of the type-II methanotrophs increased significantly with age (P = 0.00014-0.0025), their relative abundances were still much lower than the unmined sites, which suggests that the methanotrophic community was recovering but will take much longer than 31 years to become comparable to undisturbed forest soils. Methanotrophs living at low methane concentrations are extremely vulnerable to disturbances (75, 76), and the recovery time of the soil methane sink after the disturbance of land use change have been estimated to be >100 years (27). Unfortunately, factors controlling this recovery are currently unknown (77) but they are likely to include not only abiotic factors but also interactions with methanogens and *r*-selected methanotrophs.

Overall, this work provides a comprehensive view of changes in relative abundance of functional genes and microbial community structure using shotgun metagenomics during microbial recovery following ecosystem disturbance, and it elucidates possible mechanisms driving microbial succession after disturbance and their potential ecological implications. The shift of microbial taxa and functional genes both indicated that a transition from labile to recalcitrant carbon played an important role in driving microbial change. The resource availability change likely led to the transition from copiotrophic to oligotrophic groups and increased competition between microorganisms. Microbial succession not only occurred at the overall microbial community level but also was apparent within functional groups, such as the transition between AOB and AOA within ammonia oxidizers and the transition between Nitrospirae and Nitrobacter within nitrite oxidizers. The analysis of community structure and functional genes involved in methane metabolism also indicated that the extremely long recovery time of methanotrophs

could be the reason that the forest soils did not recover its function as methane sinks after 31 years.

References

1. Bonan GB. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. science 320:1444-1449.
2. Akselsson C, Westling O. 2005. Regionalized nitrogen budgets in forest soils for different deposition and forestry scenarios in Sweden. Global Ecology and Biogeography 14:85-95.
3. Bukata AR, Kyser TK. 2005. Response of the nitrogen isotopic composition of tree-rings following tree-clearing and land-use change. Environmental science & technology 39:7777-7783.
4. Baldrian P, Kolarik M, Stursova M, Kopecky J, Valaskova V, Vetrovsky T, Zifcakova L, Snajdr J, Ridl J, Vlcek C, Voriskova J. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J 6:248-58.
5. Baldrian P, Kolařík M, Štursová M, Kopecký J, Valášková V, Větrovský T, Žifčáková L, Šnajdr J, Rídl J, Vlček Č. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. The ISME journal 6:248-258.
6. Song W, Kim M, Tripathi BM, Kim H, Adams JM. 2015. Predictable communities of soil bacteria in relation to nutrient concentration and successional stage in a laboratory culture experiment. Environmental microbiology.
7. De Vries FT, Van Groenigen JW, Hoffland E, Bloem J. 2011. Nitrogen losses from two grassland soils with different fungal biomass. Soil Biology and Biochemistry 43:997-1005.
8. Six J, Frey S, Thiet R, Batten K. 2006. Bacterial and fungal contributions to carbon sequestration in agroecosystems. Soil Science Society of America Journal 70:555-569.
9. Kolb S. 2009. The quest for atmospheric methane oxidizers in forest soils. Environmental microbiology reports 1:336-346.
10. Knelman JE, Schmidt SK, Lynch RC, Darcy JL, Castle SC, Cleveland CC, Nemergut DR. 2014. Nutrient addition dramatically accelerates microbial community succession. PLoS One 9:e102609.
11. Zhou J, Deng Y, Zhang P, Xue K, Liang Y, Van Nostrand JD, Yang Y, He Z, Wu L, Stahl DA, Hazen TC, Tiedje JM, Arkin AP. 2014. Stochasticity, succession, and

environmental perturbations in a fluidic ecosystem. Proc Natl Acad Sci U S A 111:E836-E845.
12. Paver SF, Hayek KR, Gano KA, Fagen JR, Brown CT, Davis‐Richardson AG, Crabb DB, Rosario‐Passapera R, Giongo A, Triplett EW. 2013. Interactions between specific phytoplankton and bacteria affect lake bacterial community succession. Environ Microbiol 15:2489-2504.
13. Zumsteg A, Luster J, Göransson H, Smittenberg R, Brunner I, Bernasconi S, Zeyer J, Frey B. 2012. Bacterial, Archaeal and Fungal Succession in the Forefield of a Receding Glacier. Microb Ecol 63:552-564.
14. Salles JF, e Silva MCP, Dini-Andreote F, Dias AC, Guillaumaud N, Poly F, van Elsas JD. 2017. Successional patterns of key genes and processes involved in the microbial nitrogen cycle in a salt marsh chronosequence. Biogeochemistry 132:185-201.
15. Zak DR, Grigal DF, Gleeson S, Tilman D. 1990. Carbon and nitrogen cycling during old-field succession: constraints on plant and microbial biomass. Biogeochemistry 11:111-129.
16. Yan E-R, Wang X-H, Guo M, Zhong Q, Zhou W, Li Y-F. 2009. Temporal patterns of net soil N mineralization and nitrification through secondary succession in the subtropical forests of eastern China. Plant and Soil 320:181-194.
17. Cline LC, Zak DR. 2015. Soil microbial communities are shaped by plant‐driven changes in resource availability during secondary succession. Ecology 96:3374-3385.
18. Banning NC, Gleeson DB, Grigg AH, Grant CD, Andersen GL, Brodie EL, Murphy DV. 2011. Soil Microbial Community Successional Patterns during Forest Ecosystem Restoration. Appl Environ Microbiol 77:6158-6164.
19. Dini-Andreote F, de Cassia Pereira e Silva M, Triado-Margarit X, Casamayor EO, van Elsas JD, Salles JF. 2014. Dynamics of bacterial community succession in a salt marsh chronosequence: evidences for temporal niche partitioning. ISME J 8:1989-2001.
20. Sun S, Li S, Avera BN, Strahm BD, Badgley BD. 2017. Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration. Applied and Environmental Microbiology:AEM. 00966-17.
21. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 109:21390-5.
22. Brankatschk R, Töwe S, Kleineidam K, Schloter M, Zeyer J. 2011. Abundances and potential activities of nitrogen cycling microbial communities along a chronosequence of a glacier forefield. The ISME journal 5:1025-1037.
23. Zeng J, Lou K, Zhang C-J, Wang J-T, Hu H-W, Shen J-P, Zhang L-M, Han L-L, Zhang T, Lin Q. 2016. Primary Succession of Nitrogen Cycling Microbial Communities Along the Deglaciated Forelands of Tianshan Mountain, China. Frontiers in Microbiology 7.
24. Böttcher H, Verkerk PJ, Gusti M, HavlÍk P, Grassi G. 2012. Projection of the future EU forest CO2 sink as affected by recent bioenergy policies using two advanced forest management models. Gcb Bioenergy 4:773-783.
25. Chapuis‐Lardy L, Wrage N, Metay A, CHOTTE JL, Bernoux M. 2007. Soils, a sink for N2O? A review. Global Change Biology 13:1-17.

26. Fearnside PM. 2000. Global warming and tropical land-use change: greenhouse gas emissions from biomass burning, decomposition and soils in forest conversion, shifting cultivation and secondary vegetation. Climatic change 46:115-158.
27. Smith K, Dobbie K, Ball B, Bakken L, Sitaula B, Hansen S, Brumme R, Borken W, Christensen S, Priemé A. 2000. Oxidation of atmospheric methane in Northern European soils, comparison with other ecosystems, and uncertainties in the global terrestrial sink. Global Change Biology 6:791-803.
28. Thomas T, Gilbert J, Meyer F. 2012. Metagenomics-a guide from sampling to data analysis. Microbial informatics and experimentation 2:3.
29. von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N, Bork P. 2007. Quantitative phylogenetic assessment of microbial communities in diverse environments. Science 315:1126-30.
30. White III RA, Callister SJ, Moore RJ, Baker ES, Jansson JK. 2016. The past, present and future of microbiome analyses. Nature Protocols 11:2049-2053.
31. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A. 2008. The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics 9:386.
32. Jari Oksanen FGB, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens and Helene Wagner. 2015. vegan: Community Ecology Package. R package version 22-1 http://CRANR-projectorg/package=vegan.
33. Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford MA, Knight R. 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. ISME J 6:1007-17.
34. Leff JW, Jones SE, Prober SM, Barberán A, Borer ET, Firn JL, Harpole WS, Hobbie SE, Hofmockel KS, Knops JM. 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proceedings of the National Academy of Sciences 112:10967-10972.
35. Shapleigh JP. 2013. Denitrifying prokaryotes, p 405-425, The prokaryotes. Springer.
36. Wang Z, Zhang X-X, Lu X, Liu B, Li Y, Long C, Li A. 2014. Abundance and diversity of bacterial nitrifiers and denitrifiers and their functional genes in tannery wastewater treatment plants revealed by high-throughput sequencing. PLoS One 9:e113603.
37. Avera BN, Strahm BD, Burger JA, Zipper CE. 2015. Development of ecosystem structure and function on reforested surface-mined lands in the Central Appalachian Coal Basin of the United States. New For 46:683-702.
38. Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA, Knight R, Fierer N. 2011. The under-recognized dominance of Verrucomicrobia in soil bacterial communities. Soil Biology and Biochemistry 43:1450-1455.
39. Noll M, Matthies D, Frenzel P, Derakshani M, Liesack W. 2005. Succession of bacterial community structure and diversity in a paddy soil oxygen gradient. Environmental Microbiology 7:382-395.
40. Hui N, Jumpponen A, Francini G, Kotze DJ, Liu X, Romantschuk M, Strömmer R, Setälä H. 2017. Soil microbial communities are shaped by vegetation type and park age in cities under cold climate. Environmental Microbiology.

41. Davis KE, Sangwan P, Janssen PH. 2011. Acidobacteria, Rubrobacteridae and Chloroflexi are abundant among very slow‐growing and mini‐colony‐forming soil bacteria. Environmental microbiology 13:798-805.
42. Ramirez KS, Craine JM, Fierer N. 2012. Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. Global Change Biology 18:1918-1927.
43. el Zahar Haichar F, Marol C, Berge O, Rangel-Castro JI, Prosser JI, Balesdent J, Heulin T, Achouak W. 2008. Plant host habitat and root exudates shape soil bacterial community structure. The ISME journal 2:1221-1230.
44. Reich PB, Peterson DW, Wedin DA, Wrage K. 2001. Fire and vegetation effects on productivity and nitrogen cycling across a forest–grassland continuum. Ecology 82:1703-1719.
45. Clark EV, Zipper CE. 2016. Vegetation influences near-surface hydrological characteristics on a surface coal mine in eastern USA. Catena 139:241-249.
46. Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Daniel R. 2011. Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils. PLoS One 6:e17000.
47. King AJ, Freeman KR, McCormick KF, Lynch RC, Lozupone C, Knight R, Schmidt SK. 2010. Biogeography and habitat modelling of high-alpine bacteria. Nature communications 1:53.
48. D'Costa VM, Griffiths E, Wright GD. 2007. Expanding the soil antibiotic resistome: exploring environmental diversity. Current opinion in microbiology 10:481-489.
49. Luo W, Xu Z, Riber L, Hansen LH, Sørensen SJ. 2016. Diverse gene functions in a soil mobilome. Soil Biology and Biochemistry 101:175-183.
50. Koskella B, Meaden S. 2013. Understanding bacteriophage specificity in natural microbial communities. Viruses 5:806-823.
51. Sun CL, Relman DA. 2013. Microbiota's' little helpers': bacteriophages and antibiotic-associated responses in the gut microbiome. Genome biology 14:127.
52. Top EM, Springael D. 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Current Opinion in Biotechnology 14:262-269.
53. Springael D, Top EM. 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. Trends in microbiology 12:53-58.
54. Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: co‐evolution or useful parasites? Heredity 85:101-106.
55. McClintock B. 1984. The significance of responses of the genome to challenge. Science 226:792-801.
56. Aires T, Serrão EA, Engelen AH. 2016. Host and environmental specificity in bacterial communities associated to two highly invasive marine species (Genus Asparagopsis). Frontiers in microbiology 7.
57. Cornforth DM, Foster KR. 2015. Antibiotics and the art of bacterial war. Proceedings of the National Academy of Sciences 112:10827-10828.

58. Cytryn E. 2013. The soil resistome: the anthropogenic, the native, and the unknown. Soil Biology and Biochemistry 63:18-23.
59. Russo DA, Couto N, Beckerman AP, Pandhal J. 2016. A metaproteomic analysis of the response of a freshwater microbial community under nutrient enrichment. Frontiers in Microbiology 7.
60. Georges AA, El-Swais H, Craig SE, Li WK, Walsh DA. 2014. Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. The ISME journal 8:1301-1313.
61. Janssens I, Dieleman W, Luyssaert S, Subke J-A, Reichstein M, Ceulemans R, Ciais P, Dolman AJ, Grace J, Matteucci G. 2010. Reduction of forest soil respiration in response to nitrogen deposition. Nature Geoscience 3:315-322.
62. Liu L, Greaver TL. 2010. A global perspective on belowground carbon dynamics under nitrogen enrichment. Ecol Lett 13:819-28.
63. Treseder KK. 2008. Nitrogen additions and microbial biomass: A meta‐analysis of ecosystem studies. Ecology letters 11:1111-1120.
64. Ollivier J, Töwe S, Bannert A, Hai B, Kastl EM, Meyer A, Su MX, Kleineidam K, Schloter M. 2011. Nitrogen turnover in soil and global change. FEMS Microbiology Ecology 78:3-16.
65. Finzi AC, Berthrong ST. 2005. The uptake of amino acids by microbes and trees in three cold‐temperate forests. Ecology 86:3345-3353.
66. Blackburne R, Vadivelu VM, Yuan Z, Keller J. 2007. Kinetic characterisation of an enriched Nitrospira culture with comparison to Nitrobacter. Water Research 41:3033-3042.
67. Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol G, Prosser J, Schuster S, Schleper C. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442:806-809.
68. Chen XP, Zhu YG, Xia Y, Shen JP, He JZ. 2008. Ammonia‐oxidizing archaea: important players in paddy rhizosphere soil? Environmental Microbiology 10:1978-1987.
69. Ke X, Angel R, Lu Y, Conrad R. 2013. Niche differentiation of ammonia oxidizers and nitrite oxidizers in rice paddy soil. Environmental microbiology 15:2275-2292.
70. Kandeler E, Deiglmayr K, Tscherko D, Bru D, Philippot L. 2006. Abundance of narG, nirS, nirK, and nosZ genes of denitrifying bacteria during primary successions of a glacier foreland. Applied and environmental microbiology 72:5957-5962.
71. Shindell DT, Faluvegi G, Koch DM, Schmidt GA, Unger N, Bauer SE. 2009. Improved attribution of climate forcing to emissions. Science 326:716-718.
72. Nazaries L, Tate KR, Ross DJ, Singh J, Dando J, Saggar S, Baggs EM, Millard P, Murrell JC, Singh BK. 2011. Response of methanotrophic communities to afforestation and reforestation in New Zealand. The ISME journal 5:1832-1836.
73. MacDonald J, Skiba U, Sheppard L, Ball B, Roberts J, Smith K, Fowler D. 1997. The effect of nitrogen deposition and seasonal variability on methane oxidation and nitrous oxide emission rates in an upland spruce plantation and moorland. Atmospheric Environment 31:3693-3706.

74. Zerva A, Mencuccini M. 2005. Short-term effects of clearfelling on soil CO 2, CH 4, and N 2 O fluxes in a Sitka spruce plantation. Soil Biology and Biochemistry 37:2025-2036.
75. Bodelier PL, Roslev P, Henckel T, Frenzel P. 2000. Stimulation by ammonium-based fertilizers of methane oxidation in soil around rice roots. Nature 403:421-424.
76. Priemé A, Christensen S, Dobbie KE, Smith KA. 1997. Slow increase in rate of methane oxidation in soils with time following land use change from arable agriculture to woodland. Soil Biology and Biochemistry 29:1269-1273.
77. Ho A, Lüke C, Frenzel P. 2011. Recovery of methanotrophs from disturbance: population dynamics, evenness and functioning. The ISME journal 5:750-758.

Chapter 5. Models of bacterial co-occurrence groups are predictive of bacterial succession dynamics during forest restoration.

Abstract

Soil microbiota play vital roles in biogeochemical processes and constitute an essential part of terrestrial ecosystems, but their ecology is difficult to examine directly because of high diversity and complexity of interactions. Modeling is a useful tool for predicting microbial community dynamics over space and time, and has been applied in different types of systems to identify the linkages between microbial communities and their environments. However, it is particularly difficult to model soil microbiota because of the extremely high taxonomic richness. Furthermore, due to uncertainty about coherence of ecological function at high microbial taxonomic levels, simplifying community structure data by grouping OTUs at the phylum or class levels may mask significant trends of some important OTUs. We present a method of instead grouping OTUs based on their co-occurrence patterns and tested the ability of the resulting groups to predict community dynamics and environmental factors within artificial neural networks (ANN). Bacterial co-occurrence groups (BCGs) were built based on their consistent co-occurrence patterns in soils from reclaimed mine sites representing a chronosequence of ages since reforestation to follow the development of the forest ecosystem. The abundances of BCGs and previously published measurements of environmental factors were incorporated into an ANN in order to test the utility of BCGs as predictors of changes in community structure and environmental factors. Compared with grouping OTUs into phylum or class, BCGs simplified microbial communities further, with only eleven groups to model, but more accurately reproduced patterns of community structure change, with a higher similarity to

the patterns characterized at OTU level. BCGs also predicted community structures and environmental factors more accurately than phylum and class when used as input information in ANN models.

5.1 Introduction

Soil microorganisms play essential roles in ecosystem function; they mediate the biogeochemical processes and influence the establishment of the plant community. The recent advances of next-generation sequencing have made it tractable to examine the taxonomic structure of numerous microbial communities from different ecosystems or different stages of ecosystem development (1, 2). However, besides shifts of taxonomic structure, information about relationships among different microbial groups and ecological processes is also important for understanding the role of microbiota in maintaining ecosystem processes (1, 3). Unfortunately, because of the complexity of most microbial communities and the inability to culture the majority of microorganisms, it is difficult to track these relationships and interactions *in situ*, particularly in soil systems (4-6).

Modeling can be a valuable tool to predict shifts in microbial community structure and to infer important interactions among microbes and external factors (7). For example, the generalized Lotka-Volterra model was used to predict intestinal bacterial community dynamics under antibiotic perturbations (8) and an ecosystem reliability model was used to identify connections between bacterial community structure and nitrate removal in a denitrifying bioreactor (4). However, in contrast to relatively simple systems like the human gut and bioreactors, soil microbial communities exhibit complex structures, with thousands of or even tens of thousands of species that potentially involve even more interactions (9, 10). As a result, it is

computationally expensive to model the soil microbiome at the species level, overwhelming to interpret the results with thousands of predicting variables, and tends to cause overfitting in some models when there are more variables than observations.

One method to overcome the unmanageable complexity and simplify soil microbial community structure is to group operational taxonomic units (OTUs) at higher taxonomic levels. For example, order abundance has been used in modeling microbial communities in sea and river water, which, after excluding less abundant orders, resulted in 24 and 7 groups respectively as input variables for modeling (3, 11). However, soil microbiota are still much more diverse (12), with thousands of microbial OTUs comprising hundreds of orders occurring in a gram of soil (13, 14). Furthermore, while models based on high taxonomic levels such as phylum or class do simplify the structure, the coherence of ecological function at these taxonomic levels remains controversial (15). For example, all known organisms in phylum Cyanobacteria are photoautotrophic, and all known members of phylum Chlamydiae are obligate intra-cellular symbionts or parasites (15). In contrast, phylum Proteobacteria includes an extremely high diversity of functions, including the class Alphaproteobacteria, which is considered one of the most ecologically diverse groups (15). Even lower taxonomic ranks, such as family and genus have shown different levels of ecological coherence in different cases (15). In addition to uncertainty about correlation between taxonomy and ecological function, bacteria belonging to the same taxa may actually have counteractive ecological roles that, as a result of grouping OTUs into higher taxonomy ranks, could potentially average their trends out and remain undetected.

As opposed to taxonomic grouping, another method to simplify intricate community structure is to group OTUs based on co-occurrence across multiple samples. For example, microorganisms in lake water communities were grouped based on their different peak time during a year for modeling the community dynamics and the influence of environmental factors (16). Another approach to this challenge has been to use weighted correlation network analysis (also known as weighted gene co-expression network analysis; WGNCA) (17). While WGCNA was originally designed to identify highly correlated gene expression patterns among complex transcriptome data sets, it has also been successfully used for identifying groups of highly correlated microbial OTUs (18-21). Previously, the application of WGCNA has been limited to identifying microbial groups related to clinical traits or environmental factors with correlation analysis, but it is also possible to identify groups of co-occurrent OTUs that can be used to simplify and model microbial community structure in complex systems.

In this study, we present a modeling approach incorporating simplified bacterial community structures that reproduce the recovery and intra-annual patterns of bacterial communities previously described during ecosystem recovery following reforestation of reclaimed mine lands (10). The bacterial co-occurrence groups (BCGs) were generated using WGCNA based on their co-occurrence patterns across a chronosequence of reclaimed mine soils representing different ages since reforestation and intra-annual sampling time points. The patterns of soil community structure change across the chronosequence ages were characterized using the BCG abundances, and ANNs incorporating BCG abundance were built to test the ability of BCGs to predict community structure and some previously published environmental factors. Our specific objectives were to test: (1) whether the BCGs could be identified that show different intra- and interannual patterns and different taxonomic compositions, which indicates that they represent

different ecological responses to environment change; (2) whether BCGs could re-create patterns of community structures more reliably than phylum and class, with a higher similarity to the patterns characterized using thousands of OTUs; (3) whether the ANN built with BCGs would have lower errors when predicting community structure and environmental factors, compared with an ANN built on phylum or class abundances.

5.2 Materials and methods

5.2.1 Data collection and processing

The data used in this study are amplicon sequencing libraries of bacterial communities in soils collected from a chronosequence since reforestation of reclaimed mine lands (5, 11, 21 and 30 years (yr) post reforestation) and a nearby unmined reference forest site. Details on the methods for characterizing bacterial community structure were described in detail by Sun et al. (10) and in Chapter 3. Briefly, soil was sampled at four separate times in 2013 (May, July, September and October) from triplicate plots randomly located within each site. Details of sampling sites, plot design and sampling methods were further described in the "Soil dynamics" section of Avera et al. (22). The V4 region of the 16S rRNA of soil sample DNA was sequenced on the Illumina Miseq platform and raw reads are available in the NCBI BioProject database (accession number PRJNA324696).

The taxonomic composition of each sample at the phylum, class, and OTU levels were used as input data in this study. Soil properties analyzed from these samples and described by Avera et al. (22) were also used as input data, which included temperature (Temp) (°C), moisture (%) , available NH_4^+ and NO_3^- (mg N cm^{-2} day^{-1}); gas fluxes of CO_2 (mg CO_2-C m^{-2} h^{-1}), CH_4 (μg CH_4-C m^{-2} h^{-1}) and N_2O (μg N_2O-N m^{-2} h^{-1}); and total microbial biomass C (TMB)

(mg C g^{-1} dry soil) and substrate-induced respiration (SIR) (μg CO_2-C mg^{-1} soil h^{-1}). All environmental variables above were normalized with the formula:

$$\boldsymbol{e\_norm_{ij}} = (\boldsymbol{e_{ij}} - \mathbf{min}(\boldsymbol{e_j}))/(\mathbf{max}(\boldsymbol{e_j}) - \mathbf{min}(\boldsymbol{e_j})) \quad (1)$$

where e_norm_{ij} is the normalized parameter j of sample i and e_{ij} is the observed parameter j of sample i. Euclidean distances between pair-wise samples were calculated based on the normalized values and visualized by principle component analysis (PCoA) (function "princomp" in R package 'vegan' (23)). Analysis of similarity (ANOSIM) was used to test the significance of differences among groups of samples (function 'anosim' in R package 'vegan').

5.2.2 Generation of BCGs

The R package 'WGCNA' (17) was used to construct BCGs from bacterial OTU abundance data. The bacterial OTU abundance matrix represented the number of sequences belonging to each OTU in each sample, and the total sequence count of each sample was rarefied to 97,000 for normalization. OTUs with lower abundances (CPM (counts per million) <6) or non-significant variance (FDR (False discovery rate) > 0.05) were excluded from downstream analysis. The filtered OTU matrix was $\log_2$ transformed, scaled by subtracting means, and divided by the standard deviations. The scaled OTU matrix was analyzed using the 'blockwiseModules' command to identify groups of highly correlated OTUs using the value of 7 for the soft-thresholding power (used in power transformation to suppress low correlations that are likely noise), selected using the 'pickSoftThreshold' command, 50 for the minimum group size, and 0.25 for mergeCutHeight. As a result, 11 BCGs were identified and their eigengenes (first principal components), taxonomic compositions and abundances were calculated for each sample. The Bray-Curtis distances between all sample pairs were re-calculated based on their

BCG compositions, phylum compositions, class compositions and OTU compositions using the 'vegdist' function in the R package 'vegan' (23). The correlations among distance matrices were calculated using Mantel test ('mantel' function in R package 'vegan').

5.2.3 Artificial neural network model

The development of artificial neural networks (ANN) was inspired by the interconnected network of neurons in the brain; mimicking these natural processes, model neurons receive input from other neurons, process that input, and pass the output to downstream neurons (24). Because of their capability to reveal sophisticated interactions and recognize complicated patterns (24, 25), we applied an ANN model here to predict the trends of BCG abundances and changes in environmental factors. The BCG abundances and environmental factors were combined as one matrix, and all values were normalized to a range from 0 to 1 with the formula:

$$\boldsymbol{m\_norm_{ij}} = (\boldsymbol{m_{ij}} - \mathbf{min}(\boldsymbol{m_j}))/(\mathbf{max}(\boldsymbol{m_j}) - \mathbf{min}(\boldsymbol{m_j})) \quad (2)$$

where m_norm_{ij} is the normalized parameter j of sample i and m_{ij} is the observed parameter j of sample i. Because of missing data for some environmental variables, only 57 of 60 samples could be used for modeling. The total dataset was divided into a training set and a validation set, with data from 43 randomly selected samples used for training and the remaining 14 samples used for validation (3-fold cross-validation) (26), and then the ANN model was cross-validated 100 times.

The ANN models were generated from the training dataset with R package ‘neuralnet’ (27). With the interactions among BCGs and their relationships with environmental factors calculated, the abundance of one BCG was predicted from the abundances of other BCGs and environmental factors, and hence one model was generated for each BCG. Similarly, each of the environmental factors was predicted from the abundances of 11 BCGs and the other environmental factors. The ANN was trained with a resilient backpropagation algorithm with weight backtracking (28) and consisted of four layers: one input layer, two hidden layers (8 and 3 neurons respectively), and one output layer, which are similar to other studies (26, 29). The model was also trained using data from only 3 of the 4 sampling months and tested for the remaining month to evaluate its ability to predict the age-related community structure changes in the remaining month. To test its prediction of intra-annual community structure change, the model was trained with 4 chronosequence ages and tested with the remaining age. When modeling with phylum or class abundance, the procedures were the same except that the phylum or class abundances were used instead of BCG abundances.

5.2.4 Bayesian Network

Bayesian networks are used to identify relationships among networks of variables (30). Specifically, they calculate the conditional dependencies among a set of variables, and the directed acyclic graph (DAG) built from the dependencies represents the probabilistic relationships among microbial groups and environmental factors. Unlike some other network models, Bayesian networks can capture multiple types of relationships and are relatively robust with small datasets (31). In order to infer relationships between BCGs and environmental factors in this study, a Bayesian network was constructed from the same normalized matrix as the ANN with R package ‘bnlearn’ (32). With the interactions among BCGs, environmental factors,

chronosequence ages, and sampling months considered, any of those mentioned above could be the parent node (the node directing to the given node) for another, except that chronosequence age and sampling month were not assigned parent nodes and temperature and moisture were not assigned parent nodes besides age and month, as the chronosequence age and sampling month are not dependent on other factors. The hill-climbing algorithm was used to construct the model, and the Bayesian Information Criterion was used as the network score for evaluating its performance. In the generated network, each node is one of the factors mentioned earlier, and each directed edge (from parent node to child node) indicates that change of the child node is conditionally dependent on change of the parent node.

5.3 Results

5.3.1 Abundance and taxonomic composition of BCGs

Analysis of OTU abundance data collected from all 60 samples (3 replicate plots in 5 chronosequence ages collected in 4 different months) using WGCNA successfully identified clusters of OTUs that were significant correlated and resulted in the formation of 11 BCGs. The cluster dendrogram in Figure 5.1 represents the clustering of bacterial OTUs based on the similarity of their co-occurrence patterns, and each BCG (represented by different colors) comprised of OTUs that are clustered together. The eigengenes (first principle component) of BCGs showed high variations in intra- and interannual patterns, indicating that the BCGs responded differently to environmental change (Figure 5.2). For example, BCGs 1, 3, 4 and 9 are characterized mainly by intra-annual shifts among months, while the remaining BCGs were related to changes among chronosequence ages (Figure 5.2). The OTUs belonging to the same

BCG tend to have similar shifting trends, which could indicate that they react similarly to environmental change and could have similar or related functional roles in the environment.

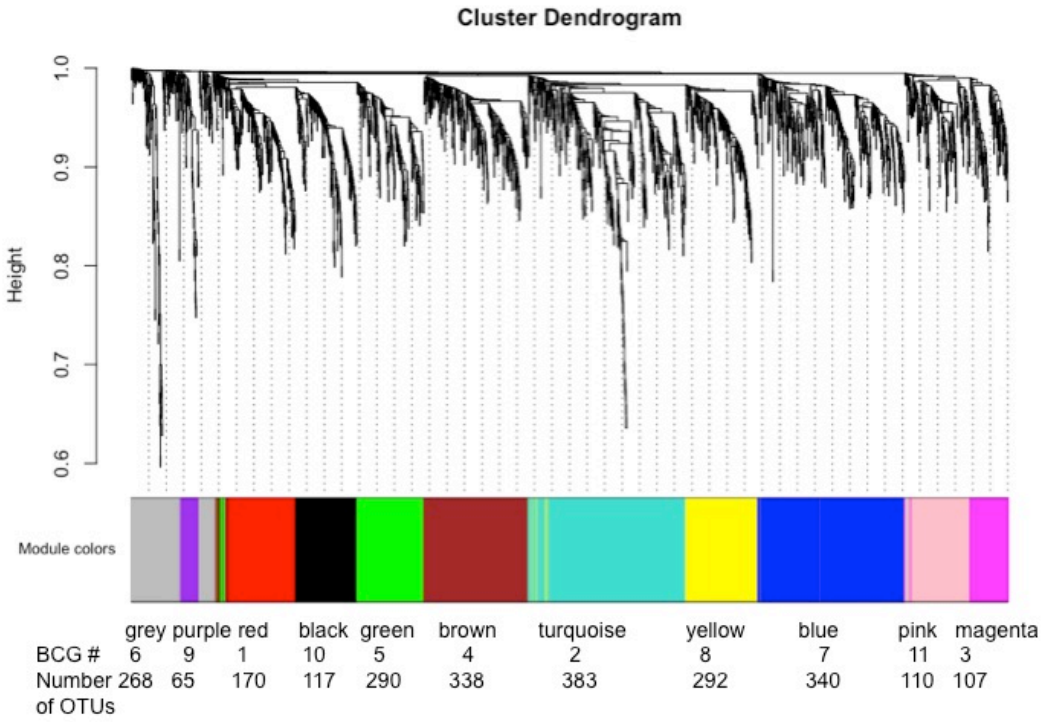


Figure 5.1. Cluster dendrogram of BCGs identified with WGCNA and the number of OTUs within each BCG. The cluster dendrogram represents the similarity of co-occurrence patterns of bacterial OTUs.

Within each BCG, OTUs belonged to between 12 and 22 different phyla, with different relative abundances of each phylum in each BCG (Figure 5.3). Some phyla, such as Acidobacteria, Verrucomicrobia and Bacteroidetes, appeared in all BCGs, but most of BCGs are dominated by only one or two specific phyla. For example, many of the OTUs in BCGs 1, 4 and 6 belong to Planctomycetes (41.6% and 38.8% and 34.9%, respectively). The eigengenes of BCG 1 and 4 are higher in September and October than May and July, while the eigengenes of BCG 6 do not shown clear patterns regarding to sampling month. Likewise, Acidobacteria and Planctomycetes are dominant in BCG 9 (36% and 25.5%), and its eigengenes are also higher in September and

October. In contrast, Bacteroidetes are dominant in BCG3 (31.1% of OTUs), and the eigengenes of BCG 3 are higher in May and July. With respect to chronosequence age, Actinobacteria were dominant in BCG 11 (33%), which decreases with increasing age, while Alphaproteobacteria are dominant in BCG 8 (25.5%), which increases.

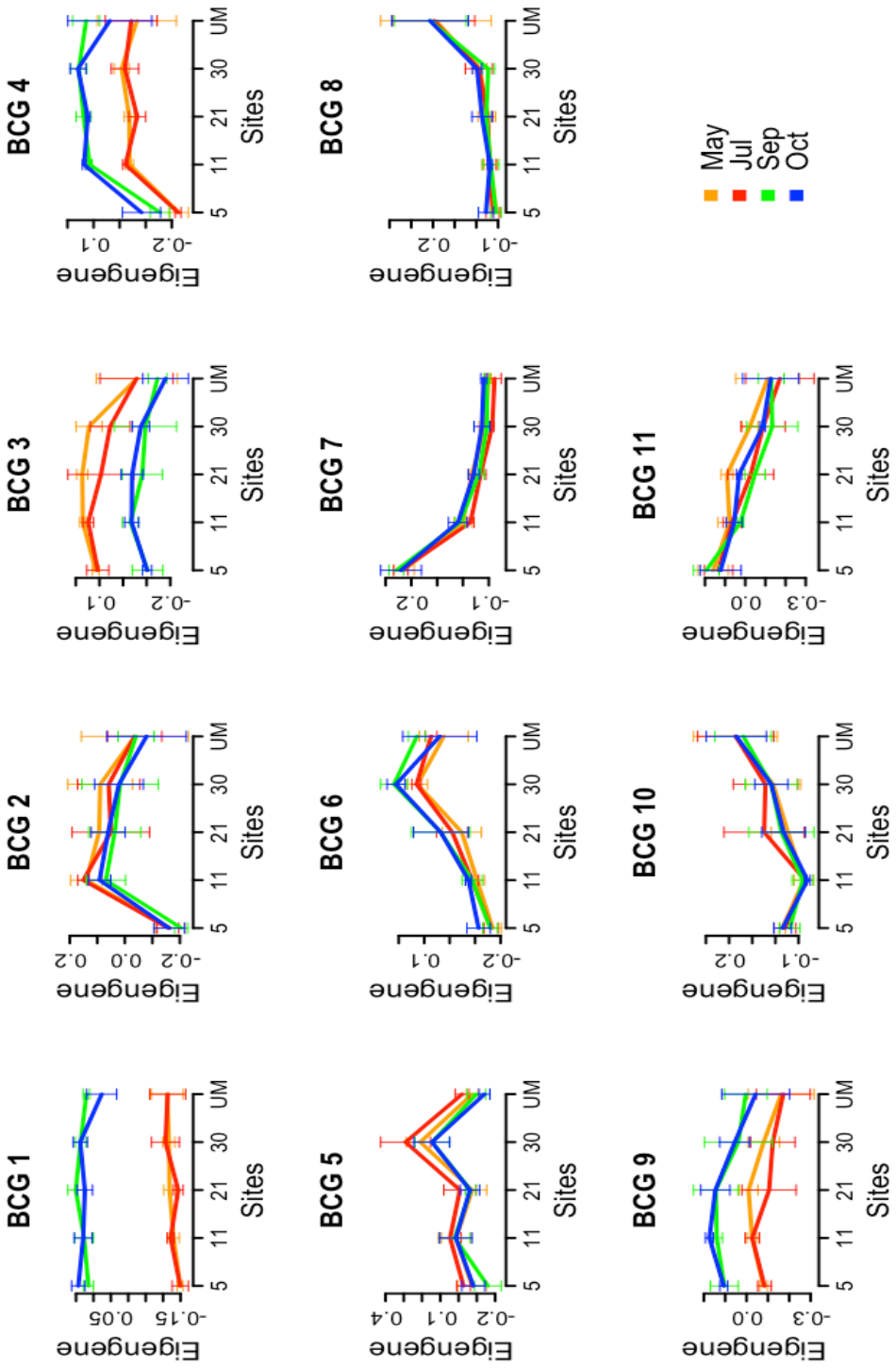


Figure 5.2. Trends of the eigengenes of the 11 BCGs across sampling months and chronosequence ages.

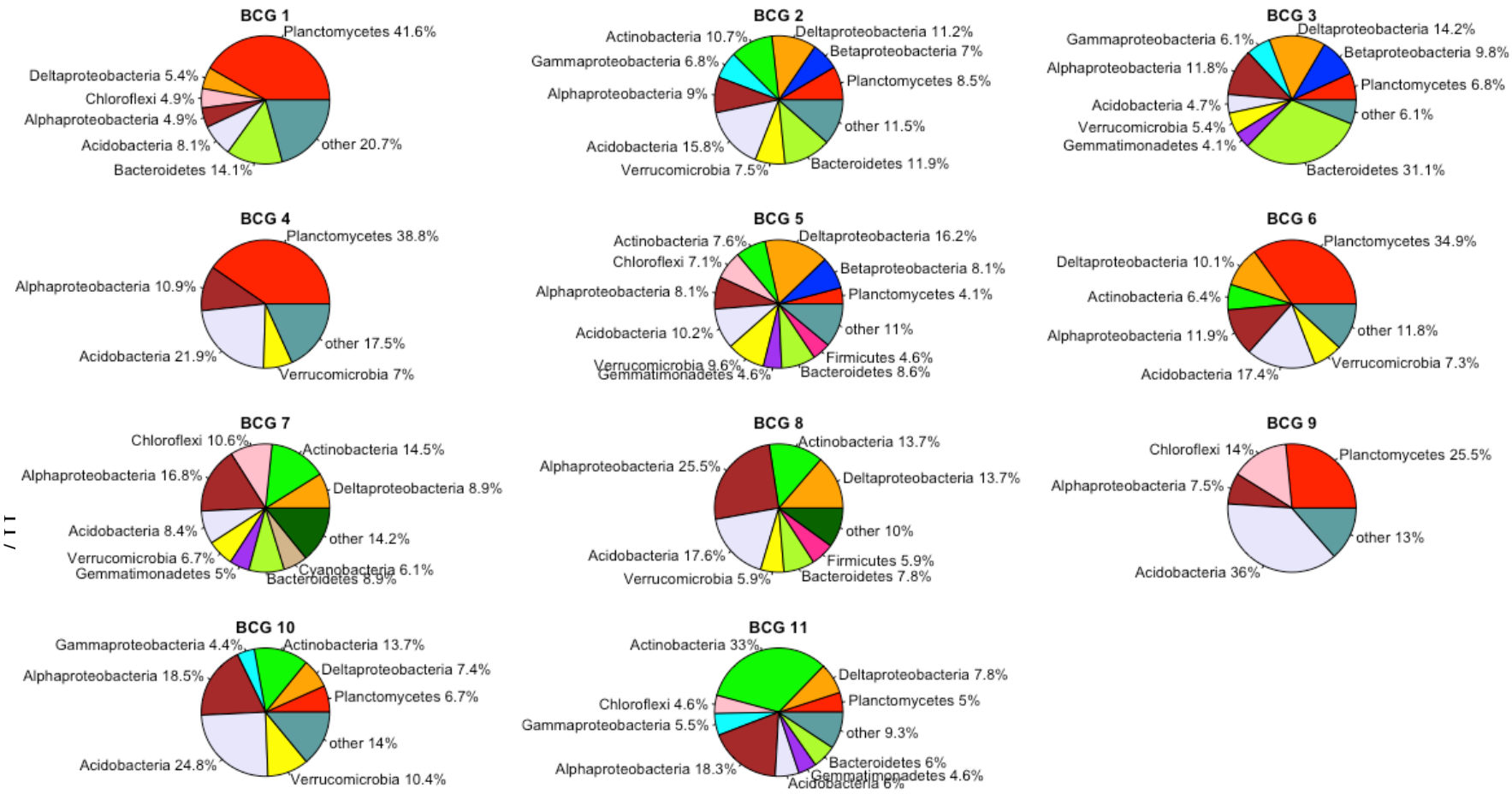


Figure 5.3. Taxonomic compositions of the 11 BCGs at phylum level or class level for Proteobacteria.

5.3.2 BCGs reproduced intra-annual and succession dynamics of the total bacterial community

The relative abundances of 11 BCGs successfully reproduced the patterns of chronosequence age- and month-related community structure change observed when analyzing the total OTU data presented in Chapter 3 (Figure 5.4). For example, the NMDS plot based solely on BCG abundance is very similar to the plot based on the abundances of all OTUs, with May and July clearly separated from September and October, and succession trajectory following the chronosequence age (Figure 5.4a). In contrast, plots that were based on phylum or class abundances did not reproduce the age-related patterns correctly (Figure 5.4c&d). For example, in the NMDS plot of phylum abundance, one sample from unmined sites was closer to the 5 yr samples than other unmined samples, while in the NMDS plot of class abundance, the shifting trends do not follow the trajectory with chronosequence ages as closely as the NMDS plot of OTU abundance. ANOSIM was used to test the significance of the differences between pairwise community structures characterized respectively by BCG, phylum and class (Table 5.1, 5.2 and 5.3), and the results were compared to the differences of community structures characterized by OTU abundance. The significance of differences between months detected by BCG, phylum and class were all the same to those detected by OTU abundance, with May and July significantly different from September and October. However, they showed different power in detecting the differences between ages. While ANOSIM with OTU abundance found that all ages are significantly different, BCG, phylum and class all failed to detect differences between certain age groups. BCG abundance failed to detect the differences of 2 pairs out of 10 pairs (11 yr and 21 yr, 21yr and 30 yr) (Table 5.1). Phylum abundance failed to detect the differences of 8 pairs out of 10 pairs, with it only revealing the 5-30 and 5-UM differences correctly (Table 5.2). Class

abundance failed to detect the differences of 3 pairs out of 10 pairs (11 yr and 21 yr, 11 yr and 30 yr, 21 and 30) (Table 5.3).

The pair-wise Bray-Curtis distances of all samples calculated from BCG abundances were significantly correlated to those calculated from the abundances of all 16,891 OTUs (mantel test, r=0.925, p=0.001). While the distances calculated from phylum and class abundances are also significantly correlated with distances calculated from OTU abundances (mantel test; phylum: r=0.555, p=0.001; class: r=0.847, p=0.001), the correlations are not as strong, especially for phylum, as the one calculated with BCG abundance. Likewise, the distances calculated from OTU, BCG, phylum and class abundances were all significantly correlated with those calculated from environmental variables, but the strength of correlations decreased in the order of OTU > BCG > phylum > class (mantel test; OTU: r=0.306, p=0.0004; BCG: r=0.256, p=0.0004; phylum: r=0.152, p=0.0114; class: r=0.131, p=0.0192). Moreover, the number of BCGs is less than the number of phyla and far less than the number of classes (11 vs 49 and 164, respectively), which means that, in addition to better explanatory power, BCGs also have the benefit of further model simplification.

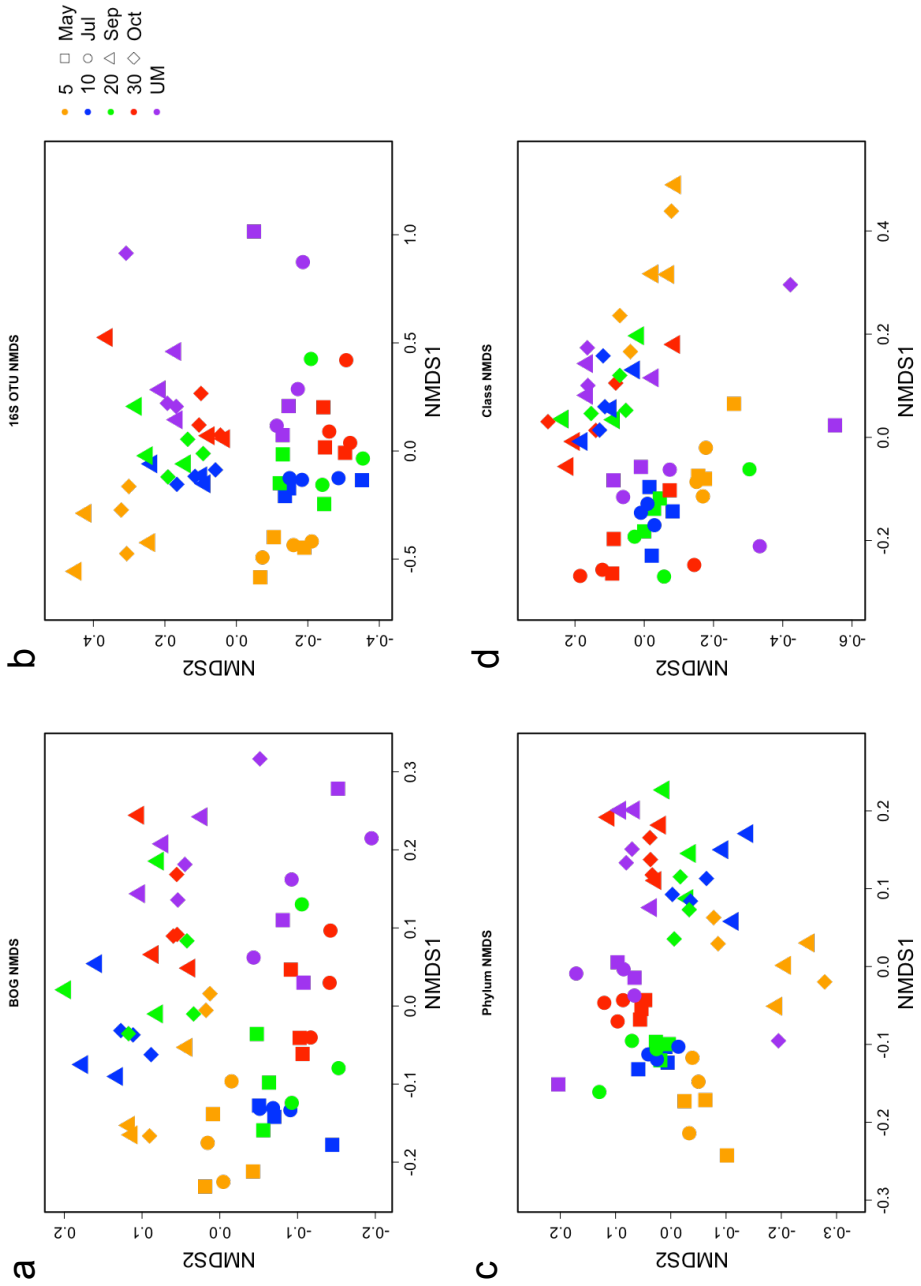


Figure 5.4. Bacterial β-diversity visualized with NMDS based on OTU (a), BCG (b), phylum (c) and class (d) abundance data.

Table 5.1. Results of significance tests of the differences between pairwise community structures characterized with BCG abundance regarding to chronosequence ages and sampling months with ANOSIM (based on 999 permutations; data in each cell is R-statistic/P-value).

a.

Month	*May*	*Jul*	*Sep*	*Oct*
May		-0.028/0.736	0.539/0.001	0.512/0.001
Jul	-0.028/0.736		0.448/0.001	0.425/0.001
Sep	0.539/0.001	0.448/0.001		-0.037/0.845
Oct	0.512/0.001	0.425/0.001	-0.037/0.845	

b.

Chronosequence Stage	*5*	*11*	*21*	*30*	*UM*
5		0.580/0.001	0.522/0.001	0.710/0.001	0.711/0.001
11	0.580/0.001		0.017/0.238	0.267/0.007	0.508/0.001
21	0.522/0.001	0.017/0.238		0.095/0.071	0.270/0.003
30	0.710/0.001	0.267/0.007	0.095/0.071		0.206/0.007
UM	0.711/0.001	0.508/0.001	0.270/0.003	0.206/0.007	

Table 5.2. Results of significance tests of the differences between pairwise community structures characterized with phylum abundance regarding to chronosequence ages and sampling months with ANOSIM (based on 999 permutations; data in each cell is R-statistic/P-value).

a.

Month	*May*	*Jul*	*Sep*	*Oct*
May		-0.029/0.806	0.757/0.001	0.779/0.001
Jul	-0.029/0.806		0.713/0.001	0.723/0.001
Sep	0.757/0.001	0.713/0.001		0.010/0.335
Oct	0.779/0.001	0.723/0.001	0.010/0.335	

b.

Chronosequence Stage	*5*	*11*	*21*	*30*	*UM*
5		0.056/0.148	0.123/0.055	0.279/0.003	0.291/0.001
11	0.056/0.148		-0.052/0.922	0.050/0.169	0.086/0.084
21	0.123/0.055	-0.052/0.922		0.014/0.279	0.038/0.203
30	0.279/0.003	0.050/0.169	0.014/0.279		0.009/0.308
UM	0.291/0.001	0.086/0.084	0.038/0.203	0.009/0.308	

Table 5.3. Results of significance tests of the differences between pairwise community structures characterized with class abundance regarding to chronosequence ages and sampling months with ANOSIM (based on 999 permutations; data in each cell is R-statistic/P-value).

a.

Month	*May*	*Jul*	*Sep*	*Oct*
May		0.0003/0.423	0.789/0.001	0.731/0.001
Jul	0.0003/0.423		0.854/0.001	0.772/0.001
Sep	0.789/0.001	0.854/0.001		-0.021/0.672
Oct	0.731/0.001	0.772/0.001	-0.021/0.672	

b.

Chronosequence Stage	*5*	*11*	*21*	*30*	*UM*
5		0.231/0.017	0.158/0.037	0.327/0.005	0.322/0.002
11	0.231/0.017		-0.016/0.455	0.094/0.08	0.309/0.001
21	0.158/0.037	-0.016/0.455		0.023/0.228	0.150/0.017
30	0.327/0.005	0.094/0.08	0.023/0.228		0.193/0.012
UM	0.322/0.002	0.309/0.001	0.150/0.017	0.193/0.012	

5.3.3 Artificial neural network and Bayesian network with BCG

The artificial neural network model takes BCG abundances and environmental factors as input information and successfully predicted the intra- and interannual changes in bacterial community structure, with lower prediction errors than ANN models based upon phylum or class. The predicted BCG abundance matrices in 100 cross-validations were all significantly correlated with the actual abundance matrices (mantel test, r=0.80-0.97, p<=0.001). The rRMSE (relative root-mean-square error) of BCG abundances was 0.20 on average, and the rRMSE of environmental factors was 0.31 on average (mean values for TMB = 0.229; SIR = 0.212; CO_2 = 0.248; CH_4 = 0.366; N_2O = 0.289; NH_4 = 0.498; and NO_3 = 0.327). The ANN models were also built with phylum and class abundances in order to compare the prediction accuracy. When phylum or class abundances were used for building the models, the rRMSEs were respectively 0.30 and 0.27 for the abundances of taxa and 0.33 and 0.34 for environmental factors, which are all significantly higher than the model built with BCGs (Welch Two Sample t-test for community

structure: phylum: t = -7.77, P-value < 0.01; class: t = -9.55, P-value < 0.01; Welch Two Sample t-test for environmental factors: phylum: t = -2.12, P-value = 0.035; class: t = -2.09, P-value = 0.037) (Figure 5.5).

In order to visualize the predicted intra-annual and overall succession dynamics of the bacterial communities, we trained the ANN model using only three months or four chronosequence ages, respectively, and tested its ability to predict each remaining one month or chronosequence age. The Bray-Curtis distances calculated from the predicted BCG abundances were visualized with NMDS. The predicted BCG abundances mostly reproduced the intra-annual and chronosequence dynamics of bacterial community change, with a clear separation of May and July from September and October, and a successional trajectory that followed chronosequence ages (Figure 5.6).

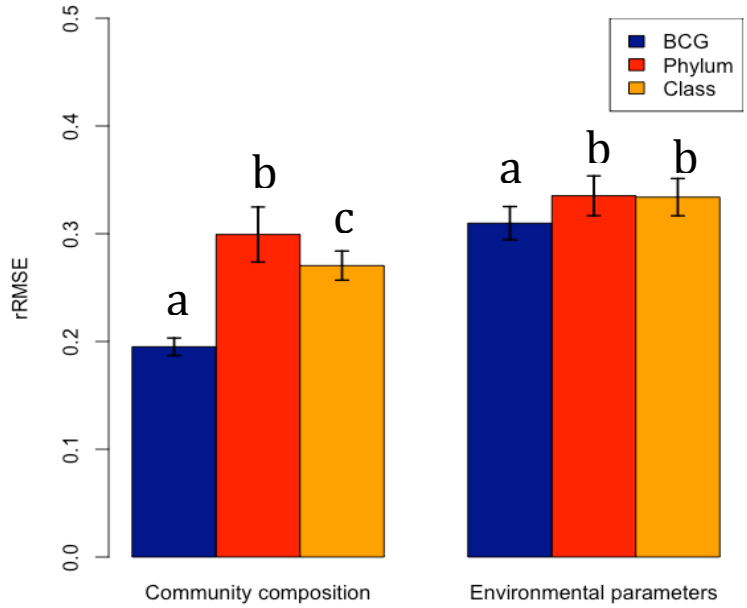


Figure 5.5. The performance of ANN models built with BCG, phylum and class abundance are evaluated with the rRMSEs when predicting community structure and environmental factors.

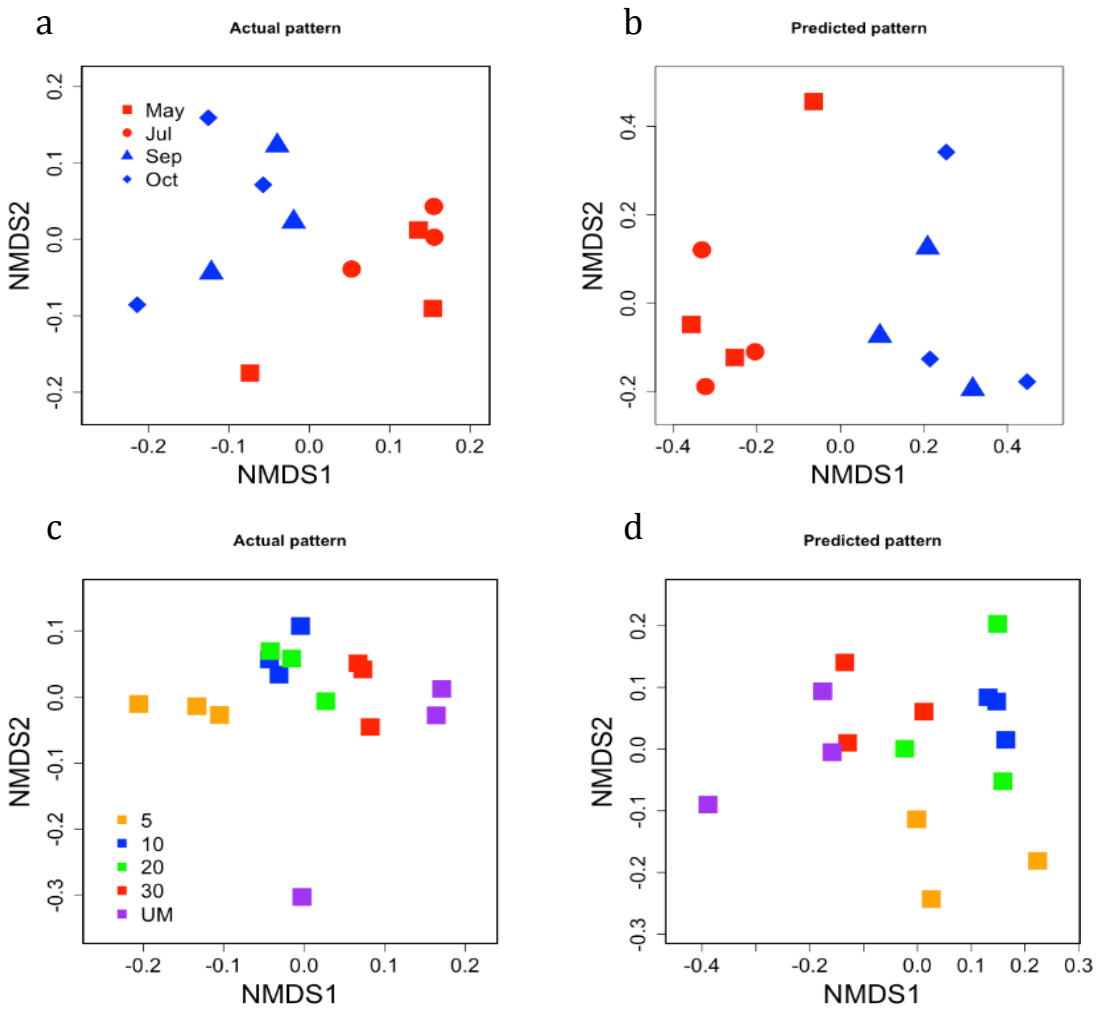


Figure 5.6. NMDS plots of the Bray-Curtis distances calculated from actual BCG abundances (a and c) and predicted BCG abundances (b and d). (a and b) The ANN model was trained with data of four ages and tested with one age in order to test whether the predicted intra-annual patterns could separate May and July from September and October. (c and d) The ANN model was trained with data of three months and tested with one month in order to test whether the predicted recovery patterns could follow the trajectory of ages.

We constructed the Bayesian network to better understand potential relationships between BCGs and environmental factors (Figure 5.7). Based on the results of the Bayesian network, BCGs 1, 3, 4, 6, 9, 10, and 11 were primarily influenced by chronosequence age. In contrast, sampling

month is the primary driver of BCGs 2, 5, 7 and 8, while temperature drives BCGs 2 and 9. Age, temperature and BCG 2 all contribute to changes in TMB, while month, TMB and most BCGs (9 out of 11) contribute to changes in SIR. Month, temperature and BCG 6 influence NH_4^+, while age, moisture and BCG 3 influence NO_3^-. For changes in gas fluxes in soil, month influences CO_2, while month, TMB, CO_2 and BCG 3 influence CH_4, and month and BCG 6 influence N_2O. While these relationships inferred from the change of BCG abundance and environmental factors by the Bayesian network indicate potentially important interactions between BCG abundances and other aspects of the ecosystem that warrant further study, it is important to remember these results do not prove a mechanistic linkage.

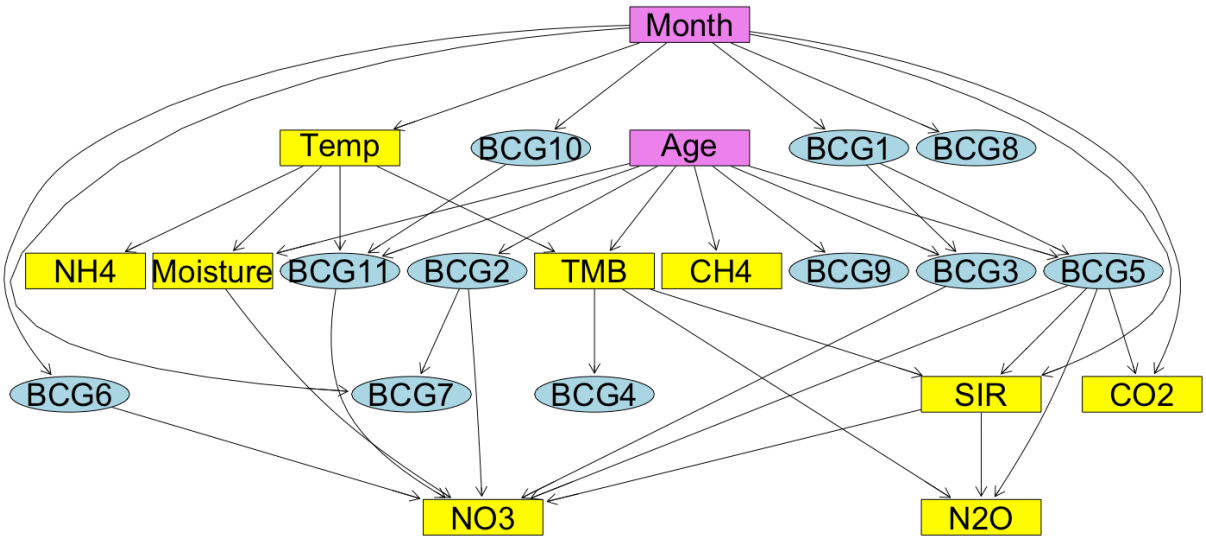


Figure 5.7. The possible interactions between BCGs and environmental factors identified by Bayesian network.

5.4 Discussion

Because of the large numbers of taxa identified in microbial communities, particularly in highly diverse soil systems, it can be difficult to identify meaningful linkages between changes in microbial community structure and ecosystem processes when using microbial data binned at low taxonomy levels (e.g., genus or species) (16). On the other hand, ongoing uncertainty about the ability to predict ecological function from classification at high taxonomic levels of bacterial phylogeny (e.g., phylum and class) make it difficult to simplify microbial community data in a meaningful way (15). To address this obstacle, we used an alternative method for simplifying microbial community data, independent of taxonomy, which can potentially be used for improved modeling of changes in soil microbial community structure and related environmental factors.

Microorganisms are highly influenced by the environments they inhabit, but the characteristics of those microhabitats are hard to measure directly. As a result, microorganisms can share similar niches and coexist in the same microhabitats, and groups of ecologically similar bacteria are commonly described as ecotypes (1, 33). In fact, this aspect of microbial life is so critical that it has been considered as one of the secondary defining properties for the definition of a microbial species (34). Different ecotypes are likely also physically and biochemically different, which could possibly explain their ecological separation (35). By including the co-occurrence patterns in the analysis of changes in bacterial communities, it is potentially possible to gain new insights into how bacterial ecotypes operate *in situ* and reveal the influence of the environment in which they live (36).

The presence, absence, or changes in abundance of microorganisms can provide information about the environmental conditions that play long-term roles in shaping microbial communities. Successful identification of BCGs with different occurrence patterns in this study indicates that the groups of microorganisms may inhabit different niches and play different ecological roles. Following this theory, the BCGs were hypothesized to be a better approach for simplifying data describing microbial community structure while still being able to characterizing the bacterial community changes compared to binning into phylum or class. In addition, the BCGs not only provide phylogenetic information but also have advantages as modeling factors to predict other environmental conditions. To test this hypothesis ANN were used because they are particularly valuable for modeling data of high complexity given their multi-layer structure and observation-inferred non-linear functions (24). The results of the ANN model supported this hypothesis in that they provided predictions with significantly lower errors from models with BCGs compared to those with phylum or class, for both community structure and environmental factors. They were also better at reproducing previously observed patterns of bacterial community structure at both intra-annual time scales and along the chronosequence age.

The 11 BCGs identified in our study were made up of different taxonomic compositions at the phylum level and showed unique changes in abundance across chronosequence ages and sampling months as well. Some phyla, such as Acidobacteria, Verrucomicrobia and Bacteroidetes, appeared in all BCGs, which indicate that organisms belonging to the same phylum do not necessarily behave similarly as ecosystems change. However, many BCGs were dominated by particular phyla, such as Planctomycetes, which suggests that these phyla might be more ecologically coherent, meaning that they contain more OTUs that respond in a similar way to ecological change. In addition, the shifts of some BCGs across chronosequence ages or

sampling months showed similar patterns as the shifts observed by their dominant taxa as described in Chapter 3. For example, BCGs 1 and 4 are more abundant in September and October, and their dominant phylum, Planctomycetes, displayed the same trends when summarized as a group (10). Planctomycetes abundances are known to be moisture-dependent (37, 38), which suggests that other OTUs in the same BCGs may also be similarly affected by soil moisture. Likewise, the decrease of BCG 11 with chronosequence age was also consistent with the changes in the dominant phylum Actinobacteria, which is typically considered as copiotrophic. BCG 3 and its dominant taxa Bacteroidetes both decreased from May and July to September and October (10), and Bacteroidetes were generally more abundant in dry soils (13). Cyanobacteria are more abundant in BCG 7 than other BCGs, and BCG 7 is most abundant at 5 yr sites, which agrees with that Cyanobacteria are photoautotrophs and its abundance decreased with the reduced sunlight from close tree canopy (10). Overall, the various trends of changes in relative abundance in BCGs in our study suggest a potential pattern of succession among BCGs, which likely resulted from their different niches and environmental change. Similar patterns of seasonal succession among groups of co-varying organisms has also been observed in surface seawater samples over time (21), suggesting this approach has potential beneficial applications across different types of ecosystems and temporal scales.

The Bayesian network illustrated that BCGs are under the influence of different factors, with some primarily influenced by chronosequence age and some others mainly affected by sampling month. Temperature, moisture and other factors also have varied effects on different BCGs, which taken all together, further supports that BCGs inhabit separate niches and play dissimilar ecological roles. However, some BCGs are only under the influence of chronosequence age or month, but not linked with any of the environmental variables, suggesting they are related to

other environmental factors not included in our dataset. The mechanisms of BCG interactions inferred by the Bayesian network could be difficult to verify, but some of the inferred interactions between environmental factors were consistent with the findings by Avera et al., such as the influence of chronosequence age on methane flux and microbial biomass (22).

It is also important to note that some of the relationships inferred by Bayesian network are different from what the trends of eigengenes indicate, such as that the eigengenes of BCG 4 and 9 showed clear month-related difference while Bayesian network considered them independent of month. Eigengenes and Bayesian network independently inferred the most likely relationships based upon different approaches. Furthermore, the eigengenes (Figure 5.2) represent only the first principle component, which only explains part of the variance, while the Bayesian network calculated the relationships based on the best performance of the overall network. Although the Bayesian network has more explanatory power in deducing uncertain relationships from complex datasets than widely used methods like correlations (39), it is important to note that the relationships inferred by a Bayesian network are not equivalent to verified causative relationships between BCGs and environmental factors. However, they provide a valuable framework for reducing complex data sets to testable hypotheses that can provide focus to follow-up experimental work to investigate possible mechanisms inferred from observational data in the field.

Overall, the application of BCGs to simplify complex soil microbiota data shows promising predictive power even using a relatively simple data structure. Given that the rapidly expanding sequencing technologies, modeling, and computational advances are being widely adopted to shed light on the myriad of interactions between microorganisms and their surrounding

ecosystem, continued development of tools will be required to manage the complex data sets resulting from this work. Although the BCGs in this work were developed for studying soil bacterial communities in the development of a forest ecosystem, this approach could applied to other systems with highly diverse microbial communities and complex structures. Still, further development of this approach is required to maximize the value of information about BCGs and their interactions with environment. For example, the BCG concept here should be tested in other systems including artificially created microbial community, and in systems where data quantifying more environmental factors are available for analysis. Other system variables could be integrated into the model to provide more insight into the ecological roles of BCGs. The improvements in understanding of the soil microbial community dynamics and microbiome-environment interactions gained through this approach could greatly improve our knowledge about microbial controls of ecosystem functions.

References

1. Konopka A, Lindemann S, Fredrickson J. 2015. Dynamics in microbial communities: unraveling mechanisms to identify principles. ISME J 9:1488-95.
2. Faust K, Lahti L, Gonze D, de Vos WM, Raes J. 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current opinion in microbiology 25:56-66.
3. Larsen PE, Field D, Gilbert JA. 2012. Predicting bacterial community assemblages using an artificial neural network approach. Nature methods 9:621-625.
4. Zilles JL, Rodríguez LF, Bartolerio NA, Kent AD. 2016. Microbial community modeling using reliability theory. The ISME journal.
5. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, Xia LC, Xu ZZ, Ursell L, Alm EJ. 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. The ISME journal.
6. Abreu NA, Taga ME. 2016. Decoding molecular interactions in microbial communities. FEMS Microbiology Reviews:fuw019.
7. Larsen PE, Gibbons SM, Gilbert JA. 2012. Modeling microbial community structure and functional diversity across time and space. FEMS microbiology letters 332:91-98.

8. Stein RR, Bucci V, Toussaint NC, Buffie CG, Rätsch G, Pamer EG, Sander C, Xavier JB. 2013. Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota. PLoS Comput Biol 9:e1003388.
9. Hartmann M, Frey B, Mayer J, Mäder P, Widmer F. 2015. Distinct soil microbial diversity under long-term organic and conventional farming. The ISME journal 9:1177-1194.
10. Sun S, Li S, Avera BN, Strahm BD, Badgley BD. 2017. Soil bacterial and fungal communities show distinct recovery patterns during forest ecosystem restoration. Applied and Environmental Microbiology:AEM. 00966-17.
11. Staley C, Gould TJ, Wang P, Phillips J, Cotner JB, Sadowsky MJ. 2014. Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. Frontiers in microbiology 5:524.
12. Kirchman DL. 2012. Processes in microbial ecology. Oxford University Press.
13. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 109:21390-5.
14. Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I, Schrumpf M, Daniel R. 2016. Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. Scientific Reports 6.
15. Philippot L, Andersson SG, Battin TJ, Prosser JI, Schimel JP, Whitman WB, Hallin S. 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Microbiol 8:523-529.
16. Dam P, Fonseca LL, Konstantinidis KT, Voit EO. 2016. Dynamic models of the complex microbial metapopulation of lake mendota. NPJ Systems Biology and Applications 2:16007.
17. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.
18. Tong M, McHardy I, Ruegger P, Goudarzi M, Kashyap PC, Haritunians T, Li X, Graeber TG, Schwager E, Huttenhower C. 2014. Reprograming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. The ISME journal 8:2193-2206.
19. Ma B, Wang H, Dsouza M, Lou J, He Y, Dai Z, Brookes PC, Xu J, Gilbert JA. 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. ISME J.
20. Tong M, Li X, Parfrey LW, Roth B, Ippoliti A, Wei B, Borneman J, McGovern DP, Frank DN, Li E. 2013. A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. PloS one 8:e80702.
21. Bryant JA, Aylward FO, Eppley JM, Karl DM, Church MJ, DeLong EF. 2016. Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. The ISME journal 10:1308-1322.
22. Avera BN, Strahm BD, Burger JA, Zipper CE. 2015. Development of ecosystem structure and function on reforested surface-mined lands in the Central Appalachian Coal Basin of the United States. New For 46:683-702.
23. Jari Oksanen FGB, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens and Helene Wagner. 2015.

vegan: Community Ecology Package. R package version 22-1 http://CRANR-projectorg/package=vegan.
24. Krogh A. 2008. What are artificial neural networks? Nature biotechnology 26:195-197.
25. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C. 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature medicine 7:673-679.
26. Winter C, Hein T, Kavka G, Mach RL, Farnleitner AH. 2007. Longitudinal changes in the bacterial community composition of the Danube River: a whole-river approach. Applied and environmental microbiology 73:421-431.
27. Günther F, Fritsch S. 2010. neuralnet: Training of neural networks. The R journal 2:30-38.
28. Riedmiller M, Braun H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm, p 586-591. *In* (ed), IEEE,
29. Singh KP, Basant A, Malik A, Jain G. 2009. Artificial neural network modeling of the river water quality—a case study. Ecological Modelling 220:888-895.
30. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics 20:3594-3603.
31. Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118-127.
32. Scutari M. 2010. bnlearn: Bayesian network structure learning. R package.
33. Ribeck N, Lenski RE. 2015. Modeling and quantifying frequency-dependent fitness in microbial populations with cross-feeding interactions. Evolution 69:1313-1320.
34. De Queiroz K. 2007. Species concepts and species delimitation. Systematic biology 56:879-886.
35. Booth A, Mariscal C, Doolittle WF. 2016. The modern synthesis in the light of microbial genomics. Annual Review of Microbiology 70:279-297.
36. Fuhrman JA. 2009. Microbial community structure and its functional implications. Nature 459:193-199.
37. Bajerski F, Wagner D. 2013. Bacterial succession in Antarctic soils of two glacier forefields on Larsemann Hills, East Antarctica. FEMS Microbiol Ecol 85:128-42.
38. Waring BG, Hawkes CV. 2015. Short-term precipitation exclusion alters microbial responses to soil moisture in a wet tropical forest. Microbial ecology 69:843-854.
39. Uusitalo L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecological modelling 203:312-318.

Chapter 6. Conclusions.

This study provided a comprehensive analysis of the microbial recovery during reforestation and ecosystem restoration of reclaimed mine sites by characterizing the abundance, diversity, taxonomic composition, potential interaction patterns, and functional gene composition of the soil microbial community. Generally, bacterial and fungal communities were both recovering from the disturbance of surface mining and became more similar to unmined sites with age. However, these two branches of microbiota showed differing characteristics during their recovery, with bacteria having higher intra-annual variability and increased interactions with age. Bacteria have a higher turnover rates and are more sensitive to environmental change, which supports the finding that they exhibited higher intra-annual variability during the study. In contrast, tight symbiotic coupling between specific fungi and plant species is a likely factor driving the high degree of age specificity among fungal communities.

The changes in taxonomic and functional gene compositions observed in metagenome sequencing from these soils suggest that microbial recovery is likely under the influence of nutrient availability and concurrent succession of the plant community. The taxonomic changes among microbial communities suggest a transition from copiotrophic to oligotrophic organisms with increasing chronosequence age and this transition was supported by decreases in genes related with cell division, as oligotrophs tend to have lower turnover rates. This copiotroph-oligotroph transition was also supported by increases of genes encoding for polysaccharide metabolism and decreases of monosaccharide-related genes, which also indicated that the biochemical forms of available nutrients were becoming more complex. This agreed with an increase of co-occurrence between microbes in constructed interaction networks and increased abundances of genes related to transposable elements. Species-occurrence is mainly driven by

metabolic dependencies (1), and transposable elements are important in transferring catabolic pathways, which could help microorganism rapidly adapt to xenobiotics (foreign substances to microbial community) (2, 3). The increased metabolic coupling was likely an adaptation by the community to metabolize new and more complex organic compounds being made available in the soil from increased litter deposition during ecosystem development. The reduced nutrient availability also appears to have increased competition among microorganisms based upon an increase of genes involved in toxins, superantigens, and resistance to antibiotics and toxic compounds as well as an increase in the number of negative relationships in co-occurrence networks. Bacteria known up-regulate the expression of toxin and antibiotics related genes under nutrient competition (4), and transposable elements can also be induced by nutrient stress to the organism (5).

The taxa and functional genes involved in nitrogen cycle and greenhouse gases were examined separately to establish linkages with previously measured ecological functions at these same sites. The taxonomic changes among these specialist organisms further suggested transitions between *r*- and *k*-selected organisms within the same functional groups, such as the shifts between AOB and AOA, *Nitrospira* and *Nitrobacter*, and between type-I and type-II methanotrophs, which implies that successional shifts not only occurred in the overall microbial community but also within certain functional groups. The genes involved in nitrogen cycle indicated a higher nitrogen cycling activity at younger chronosequence ages, but standing NH_4^+ and NO_3^- concentrations in the same sites did not show significant differences (6), suggesting that N cycling rates of the same limiting stock likely decreased. However, the important ecological function of forest soils as methane sinks was not recovered after 31 years (6). The relative abundance of methanotrophs increased significantly from 6 to 31 years, but both

methanotrophs and methane monooxygenase genes at all previously-mined sites were much less abundant than the unmined sites. This agrees with previous conclusions that methanotrophs are extremely vulnerable to disturbance and could take as long as > 100 years to recover (7-9).

The taxonomic data collected in this study was used to test the performance of BCGs, a method for simplifying microbial community structure by grouping OTUs based on their occurrence patterns instead of phylogenetic relationships. The high microbial richness in complex environments like soil has become an obstacle for modeling the microbiome, and grouping OTUs based only on their phylogenetic linkages may mask the trends of some OTUs because of the variance of ecological function among members of high taxonomic levels. The BCGs identified in this ecosystem performed better as microbial groups than either phylum and class with regards to characterizing the recovery patterns of the bacterial community and predicting both community structure and environmental factors, while at the same time simplifying microbial community structure and models to an even greater degree. Furthermore, BCGs had different taxonomic compositions, intra-annual and age-related dynamics, and were also under the influence of different environmental factors based upon Bayesian Networks, which supports the theory that BCGs could be a good mechanism for identifying new microbial ecotypes. Given their simplified structure but still superior predictive power, BCGs are a promising approach for modeling other complex microbiome systems.

References

1. Zelezniak A, Andrejev S, Ponomarova O, Mende DR, Bork P, Patil KR. 2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. Proc Natl Acad Sci U S A 112:6449-6454.
2. Top EM, Springael D. 2003. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. Current Opinion in Biotechnology 14:262-269.

3. Springael D, Top EM. 2004. Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. Trends in microbiology 12:53-58.
4. Cornforth DM, Foster KR. 2015. Antibiotics and the art of bacterial war. Proceedings of the National Academy of Sciences 112:10827-10828.
5. Capy P, Gasperi G, Biémont C, Bazin C. 2000. Stress and transposable elements: co-evolution or useful parasites? Heredity 85:101-106.
6. Avera BN, Strahm BD, Burger JA, Zipper CE. 2015. Development of ecosystem structure and function on reforested surface-mined lands in the Central Appalachian Coal Basin of the United States. New For 46:683-702.
7. Bodelier PL, Roslev P, Henckel T, Frenzel P. 2000. Stimulation by ammonium-based fertilizers of methane oxidation in soil around rice roots. Nature 403:421-424.
8. Priemé A, Christensen S, Dobbie KE, Smith KA. 1997. Slow increase in rate of methane oxidation in soils with time following land use change from arable agriculture to woodland. Soil Biology and Biochemistry 29:1269-1273.
9. Smith K, Dobbie K, Ball B, Bakken L, Sitaula B, Hansen S, Brumme R, Borken W, Christensen S, Priemé A. 2000. Oxidation of atmospheric methane in Northern European soils, comparison with other ecosystems, and uncertainties in the global terrestrial sink. Global Change Biology 6:791-803.

Appendix A

Supplementary materials for Chapter 3

Table S1. Results of significance tests of the differences between pairwise bacterial communities regarding to chronosequence ages and sampling months with ANOSIM (based on 999 permutations; data in each cell is R-statistic/P-value).

a.

Month	**May**	**Jul**	**Sep**	**Oct**
May		-0.005/0.437	0.473/0.001	0.469/0.001
Jul	-0.005/0.437		0.471/0.001	0.434/0.001
Sep	0.473/0.001	0.471/0.001		-0.036/0.846
Oct	0.469/0.001	0.434/0.001	-0.036/0.846	

b.

Chronosequence Stage	**5**	**11**	**21**	**30**	**UM**
5		0.672/0.001	0.583/0.001	0.803/0.001	0.720/0.001
11	0.672/0.001		0.134/0.037	0.375/0.001	0.588/0.001
21	0.583/0.001	0.134/0.037		0.179/0.014	0.3040.001
30	0.803/0.001	0.375/0.001	0.179/0.014		0.311/0.001
UM	0.720/0.001	0.588/0.001	0.304/0.001	0.311/0.001	

TableS2. Results of significance tests of the differences between pairwise fungal communities regarding to chronosequence ages and sampling months with ANOSIM (based on 999 permutations; data in each cell is R-statistic/P-value).

a.

Month	**May**	**Jul**	**Sep**	**Oct**
May		-0.014/0.585	0.038/0.146	0.010/0.349
Jul	-0.014/0.585		0.016/0.287	0.014/0.322
Sep	0.038/0.146	0.016/0.287		-0.077/0.996
Oct	0.010/0.349	0.014/0.322	-0.077/0.996	

b.

Chronosequence Stage	**5**	**11**	**21**	**30**	**UM**
5		0.360/0.001	0.595/0.001	0.549/0.001	0.523/0.001
11	0.360/0.001		0.232/0.001	0.381/0.001	0.478/0.001
21	0.595/0.001	0.232/0.001		0.306/0.001	0.391/0.001
30	0.549/0.001	0.381/0.001	0.306/0.001		0.235/0.001
UM	0.523/0.001	0.478/0.001	0.391/0.001	0.235/0.001	

TableS3. Results of correlation analyses between environmental factors and β-diversity patterns (*P*-values based on 999 permutations).

	bacteria		**fungi**	
	r^2	**P-value**	**r^2**	**P-value**
Temp	0.26646	0.001*	0.3064	0.001*
Moisture	0.15699	0.012*	0.3213	0.001*
TMB	0.3084	0.001*	0.4289	0.001*
SIR	0.18498	0.006*	0.2208	0.003*
CO_2	0.02561	0.486	0.0401	0.316
CH_4	0.07376	0.117	0.0032	0.953
N_2O	0.02426	0.498	0.0463	0.294
NH_4	0.12262	0.041*	0.1	0.061
$N0_3$	0.06886	0.128	0.1999	0.004*

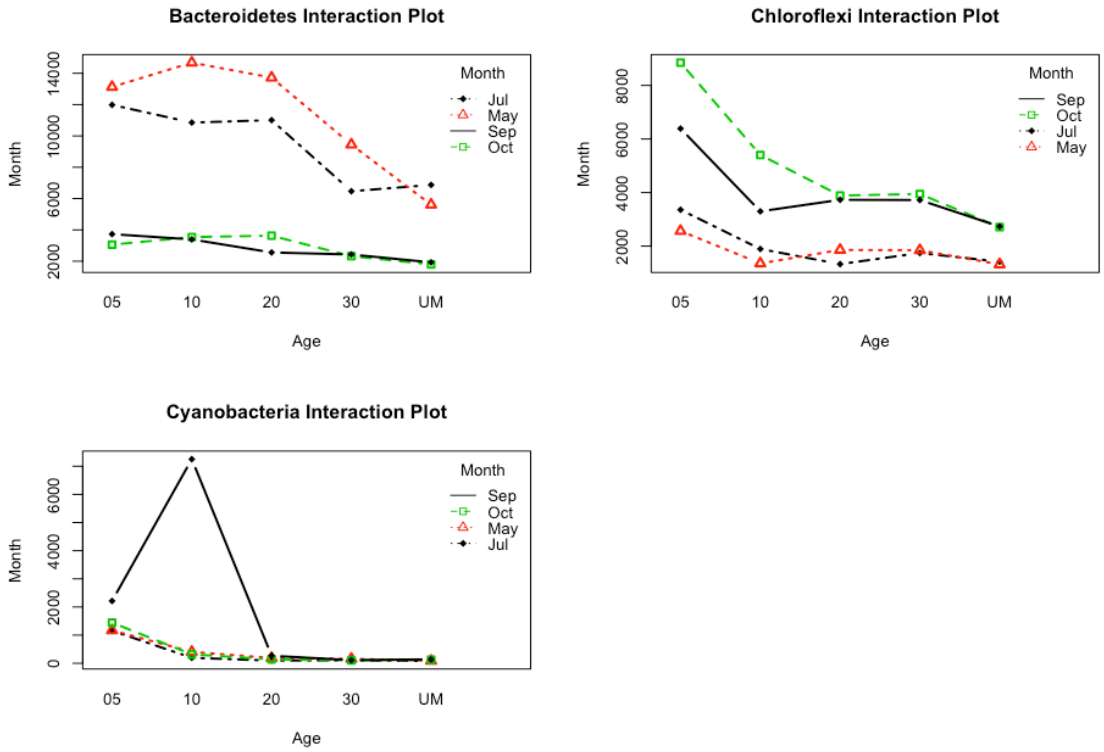
Table S4. Statistical tests of bacterial phyla (classes for phylum Proteobacteria) and fungal classes with ANOVA (data in each cell is F-statistic/P-value).

a. Bacterial taxa

***F*/*P*-value**	**Chronosequence age**	**Month**
Acidobacteria	3.28/0.02	5.73/0.002
Actinobacteria	9.70/1.40E-5	2.22/0.1
*Bacteroidetes**	12.32/1.28E-6	85.9/<2.2E-16
*Chloroflexi**	20.41/3.16E-9	40.1/3.88E-12
*Cyanobacteria**	3.65/0.013	3.53/0.023
Gemmatimonadetes	21.4/1.67E-9	33.8/4.82E-11
Nitrospirae	12.1/1.56E-6	4.50/0.008
Plancotomycetes	3.21/0.02	76.8/<2E-16
Alphaproteobacteria	8.81/3.36E-5	0.51/0.6804
Betaproteobacteria	7.74/0.0001	15.8/6.26E-7
Deltaproteobacteria	8.20/6.31E-5	55.3/2.7E-14
Gammaproteobacteria	0.49/0.74	12.5/6.70E-5
Verrucomicrobia	6.60/0.0004	1.22/0.31
WS3	5.43/0.001	0.19/0.90

*The interactions between chronosequence ages and months are significant for these 3 taxa. But as shown in the figure below, the general trends of Bacteroidetes and Chloroflexi are similar, which would have little influence on effects, while the interaction for Cyanobacteria is strong

and has influence on the main effects.



b. Fungal taxa

F/P-value	**Chronosequence age**	**Month**
Dothideomycetes	6.77/0.0003	4.78/0.006
Eurotiomycetes	3.47/0.016	0.63/0.60
Leotiomycetes	4.23/0.006	2.40/0.082
Pezizomycetes	0.85/0.50	1.27/0.30
Sordariomycetes	9.94/1.21E-5	3.57/0.0226
Agaricomycetes	2.28/0.08	0.69/0.57
Chytridiomycetes	2.28/0.08	1.97/0.13
Rozellomycota	4.99/0.002	0.21/0.89
*Incertae sedis in Zygomycota_*	9.32/2.2E-5	6.32/0.21

c. Results of Tukey's HSD post-hoc comparisons for pairwise differences in bacterial and fungal taxa when the overall effect is significant (*P*-value < 0.05) with letters denoting differences and the mean relative abundance of each taxon in each group.

	Chronosequence age (yr)					**Month**			
	5	**11**	**21**	**30**	**UM**	**May**	**Jul**	**Sep**	**Oct**
Bacteria									
Acidobacteria	16.54%	20.23%	24.22%	23.36%	23.07%	17.96%	18.63%	25.15%	24.20%
	a	a	a	a	a	a	ab	b	ab
Actinobacteria	14.21%	8.90%	8.13%	6.40%	8.82%				
	a	b	b	b	b				
Bacteroidetes*	8.22%	8.37%	7.97%	5.33%	4.18%	11.67%	9.73%	2.89%	2.96%
	a	a	a	a	a	a	a	b	b
Chloroflexi*	5.45%	3.08%	2.78%	2.90%	2.11%	1.84%	2.01%	4.10%	5.11%
	a	b	b	b	b	a	a	b	b
Gemmatimonadetes	2.34%	1.40%	1.30%	1.26%	1.05%	1.85%	2.04%	0.90%	1.07%
	a	b	b	b	b	a	a	b	b
Nitrospirae	0.13%	0.39%	0.48%	1.33%	1.79%	0.45%	0.54%	1.20%	1.10%
	a	a	ab	bc	c	a	a	a	a
Plancotomycetes	6.19%	8.05%	8.15%	8.62%	6.94%	3.97%	3.78%	12.23%	10.37%
	a	a	a	a	a	a	a	b	b
Alphaproteobacteria	17.18%	15.27%	17.02%	16.79%	23.85%				
	a	a	a	a	b				
Betaproteobacteria	7.36%	9.89%	7.74%	7.92%	5.19%	9.62%	9.30%	5.50%	6.07%
	ab	a	ab	ab	b	a	a	b	b
Deltaproteobacteria	5.98%	7.26%	6.51%	8.22%	5.96%	8.11%	9.24%	4.74%	5.06%
	a	a	a	a	a	a	a	b	b
Gammaproteobacteria						6.42%	6.30%	3.83%	3.82%
						a	a	b	b
Verrucomicrobia	6.04%	6.54%	6.69%	8.41%	8.88%				
	a	a	ab	bc	c				
WS3	0.34%	0.70%	0.61%	0.69%	0.34%				
	a	bc	ac	cd	ae				
Fungi									
Dothideomycetes	3.92%	2.58%	1.32%	1.22%	0.44%	1.36%	0.93%	1.93%	3.23%
	a	ab	bc	bc	bc	ab	a	ab	b
Eurotiomycetes	1.34%	0.66%	1.86%	0.42%	0.83%				
	ab	ab	a	b	ab				

Leotiomycetes	0.60%	1.19%	0.66%	4.11%	1.71%				
	a	ab	a	b	ab				
Sordariomycetes	10.56%	4.22%	5.94%	5.29%	3.29%	6.94%	7.28%	4.21%	4.59%
	a	b	b	b	b	a	a	a	a
Rozellomycota	0.25%	0.43%	0.94%	0.24%	0.16%				
	a	ab	bc	a	a				
Incertae sedis in Zygomycota_	0.76%	0.97%	2.15%	3.37%	3.30%	3.04%	2.69%	1.24%	1.51%
	a	a	ab	b	b	a	ab	b	ab

Table S5. Statistical tests of overlapping nodes and edges (fisher's exact test for nodes and permutation test (999 permutations) for edges, data in each cell is P-value).

Adjacent Stages		**5 -11**	**11 -21**	**21 -30**	**30 -UM**
Bacteria	*Nodes*	0.0041	0.001*	0.001*	0.001*
	Edges	0.001*	0.001*	0.001*	0.001*
Fungi	*Nodes*	0.976	0.999	0.999	0.872
	Edges	0.99	0.864	0.637	0.614

H_a: the actual number of overlaps is greater than the number of overlaps in randomized networks
* p-value is smaller or equal to 0.001.

Table S6. Values for degree and closeness centrality of bacterial nodes and the abundance of corresponding OTUs in overlapping edges and in the whole network.

Network 5	**Nodes in 5-11 overlapping edges**	**Nodes in 5 yr network**
Average degree	6.54	3.10
Degree p-value	0.001*	
Average closeness centrality	0.016	0.013
Closeness centrality p-value	0.001*	
Abundance	152	139
Abundance p-value	0.683	

Network 11	**Nodes in 5-11 overlapping edges**	**Nodes in 11-21 overlapping edges**	**Nodes in 11 yr network**
Average degree	9.69	11.13	5.33
Degree p-value	0.006	0.001*	
Average closeness centrality	0.011	0.011	0.009
Closeness centrality p-value	0.001*	0.001*	
Abundance	152	191	138
Abundance p-value	0.660	0.131	

Network 21	**Nodes in 11-21 overlapping edges**	**Nodes in 21-30 overlapping edges**	**Nodes in 21 yr network**
Average degree	17.04	16.63	7.53
Degree p-value	0.001*	0.001*	
Average closeness centrality	0.023	0.023	0.020
Closeness centrality p-value	0.001*	0.001*	
Abundance	191	183	120
Abundance p-value	0.037	0.001*	

Network 30	**Nodes in 21-30 overlapping edges**	**Nodes in 30-UM overlapping edges**	**Nodes in 30 yr network**
Average degree	14.92	13.78	7.16
Degree p-value	0.001*	0.001*	
Average closeness centrality	0.042	0.042	0.039
Closeness centrality p-value	0.001*	0.001*	
Abundance	183	162	121
Abundance p-value	0.001*	0.024	

Network UM	**Nodes in 30-UM overlapping edges**	**Nodes in Unmined network**
Average degree	22.82	11.66
Degree p-value	0.001*	
Average closeness centrality	0.027	0.025
Closeness centrality p-value	0.001*	
Abundance	162	114
Abundance p-value	0.006	

The significance of difference between degree, closeness centrality and abundance were tested with Welch Two Sample t-test.
H_0: the values of two datasets are not different
* p-value is smaller or equal to 0.001.

Table S7. Smallworldness and modularity of bacterial and fungal networks.

Network properties		**Age**				
		5	**11**	**21**	**30**	**Un-mined**
Bacteria	*Smallworldness*	2.31	0.48	0.53	0.73	0.29
	Modularity	0.25	0.25	0.04	0.04	0.02
Fungi	*Smallworldness*	2.13	0.38	0.64	1.67	0.62
	Modularity	0.86	0.88	0.88	0.983	0.986

Smallworldness is calculated with the 'smallworldness' function in the 'qgraph' R package. Modularity is calculated with the 'cluster walktrap' and 'modularity' functions in the 'igraph' R package. The network can be considered as a small world when the smallworldness is higher than 1, and the stricter standard requires smallworldness higher than 3 (1) The networks of 5 yr can be considered as small-world networks with the first standard, while others are not small-world networks.

Table S8. The numbers of bacterial nodes, fungal nodes, bacterial-bacterial, fungal-fungal and bacterial-fungal edges in the bacteria-fungi networks (constructed with 500 most abundant bacterial OTUs and 500 most abundant fungal OTUs).

Age	**Bacterial nodes**	**Fungal nodes**	**Bacterial edges**	**Fungal edges**	**Bacteria-fungi-edges**
5	312	103	613	125	64
11	295	167	733	256	79
21	357	167	1303	266	73
30	359	148	1263	240	129
UM	387	122	2210	188	163

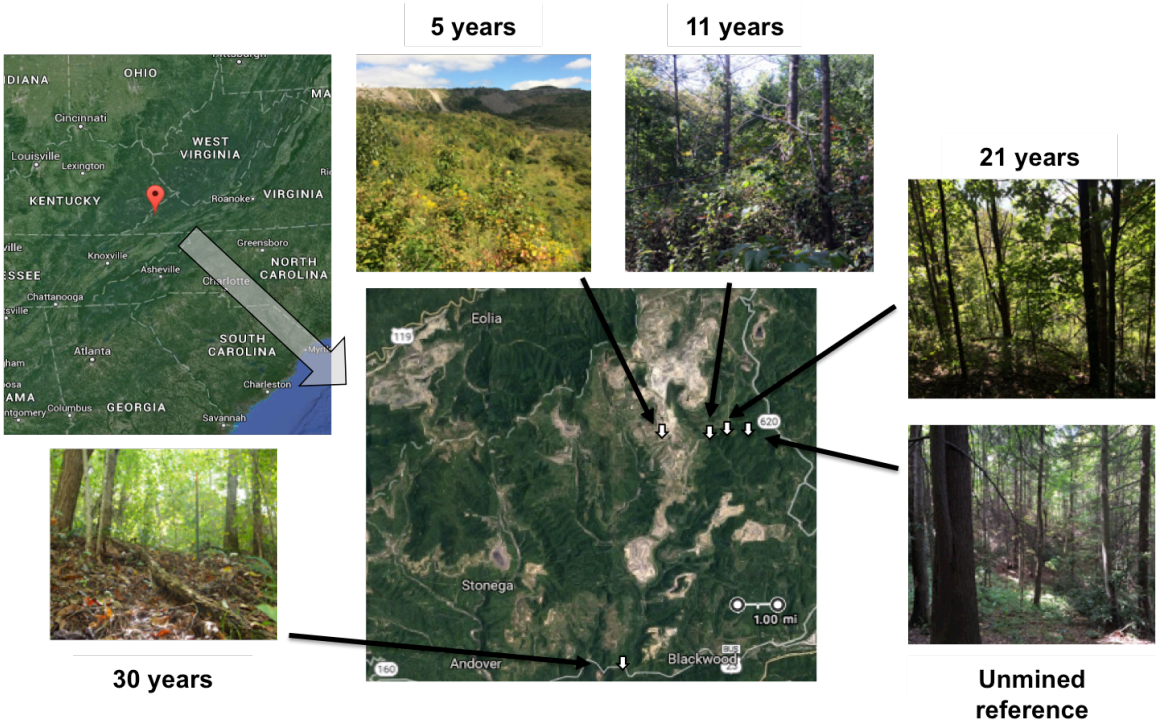


Figure S1. The locations and ages of chronosequence plots built in reclaimed mined sites located in the Powell River Project research area in Wise County, Virginia.

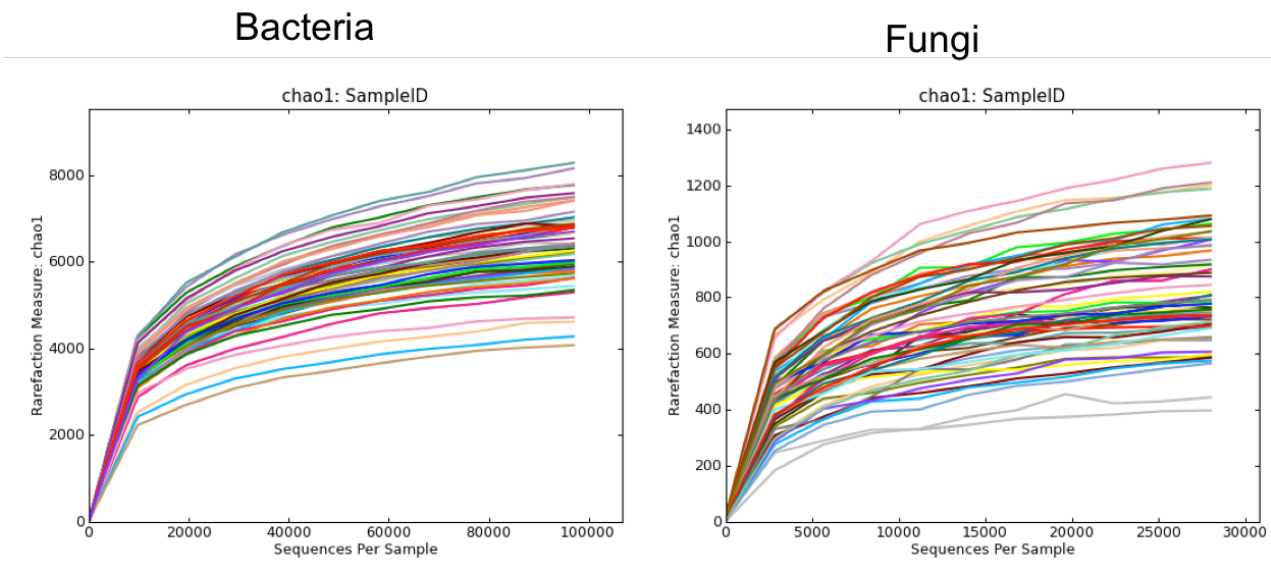


Figure S2. Rarefaction curves of bacterial and fungal richness estimated from sequencing depth.

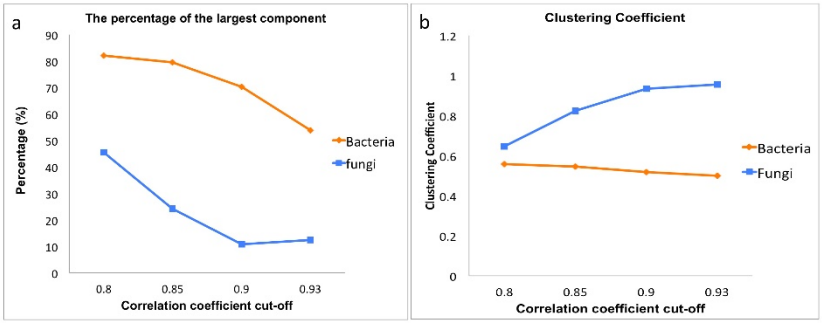


Figure S3. The comparison of bacterial and fungal networks in largest component percentage and clustering coefficient at different correlation coefficient cut-off.

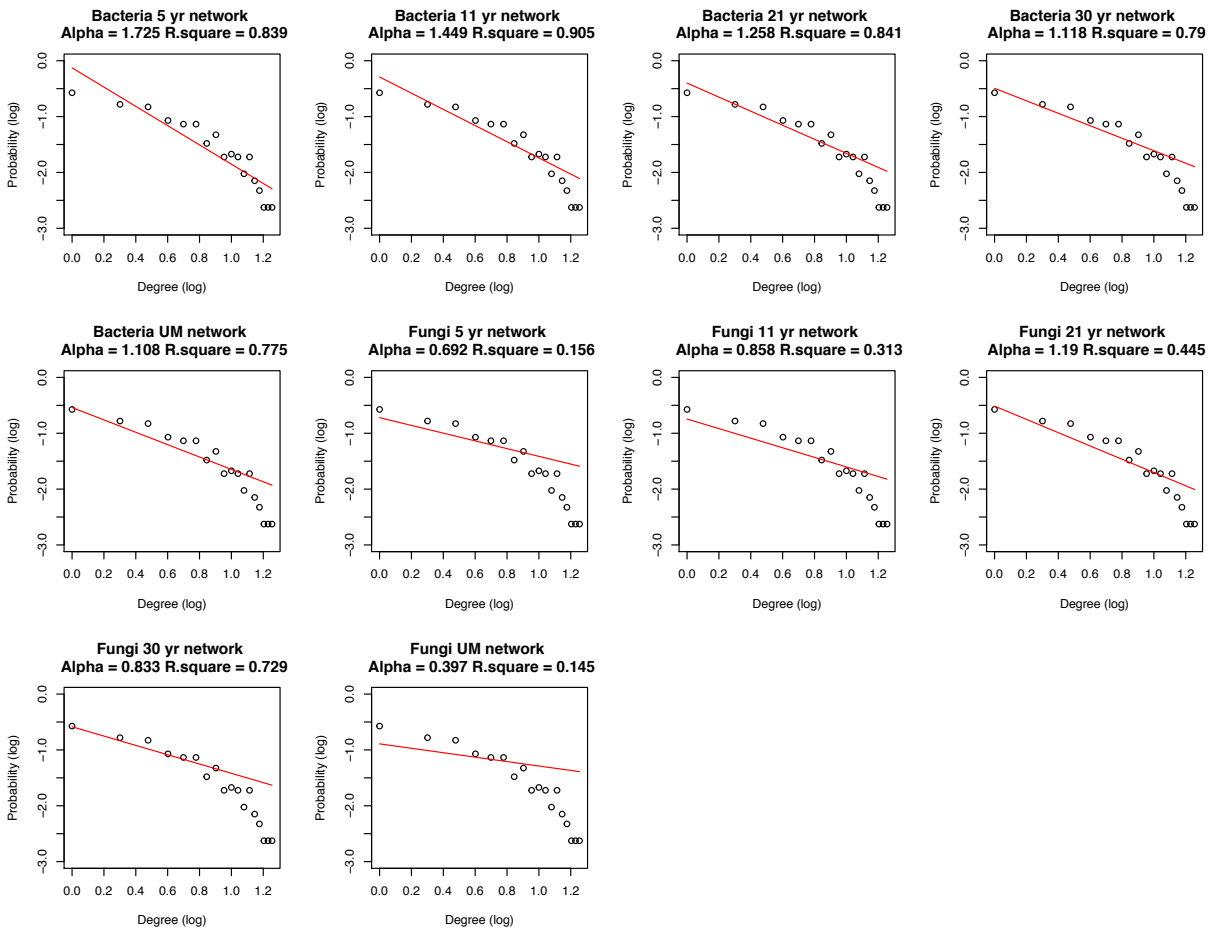


Figure S4. Scatter plots of a power-law degree distribution for bacterial and fungal networks. The networks are usually considered as scale free when alpha is between 2 and 3 (2).

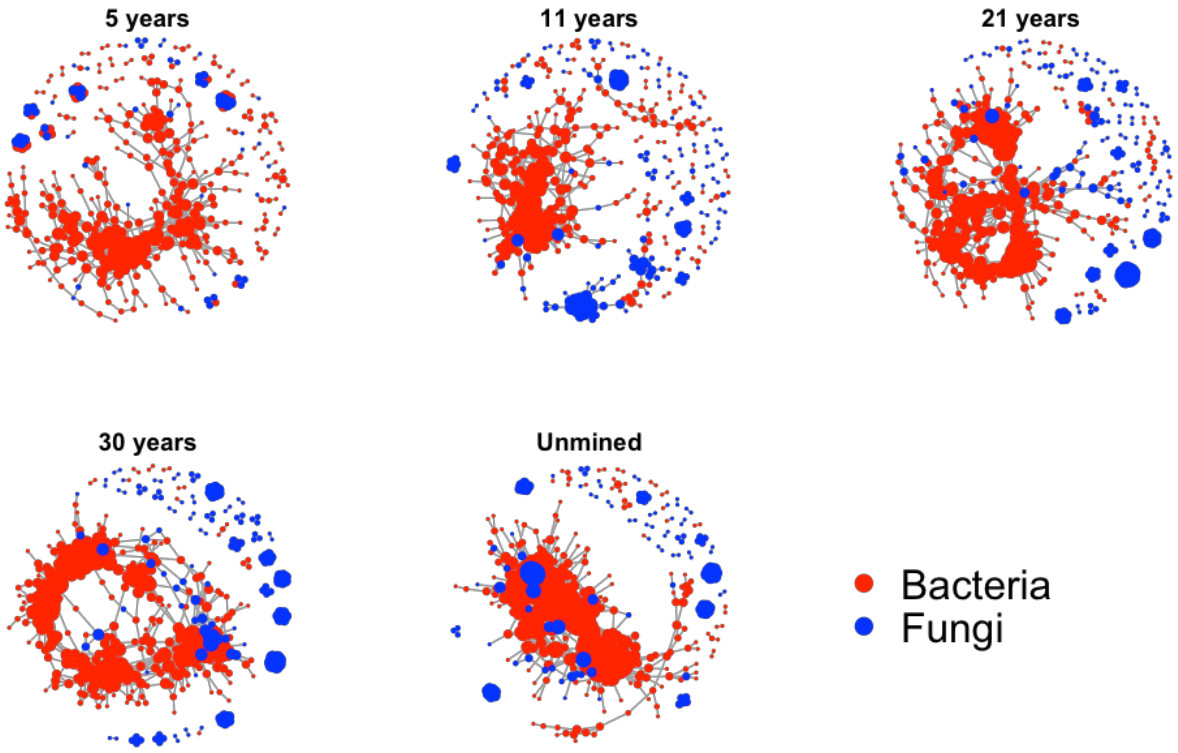
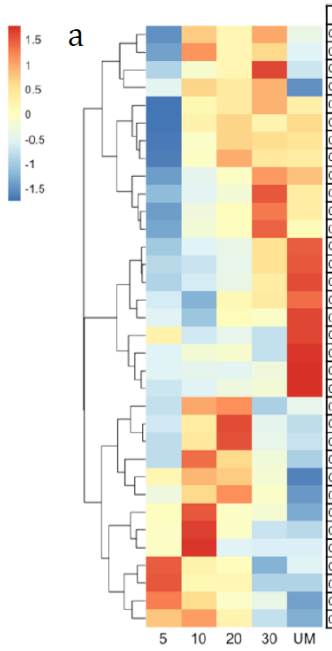
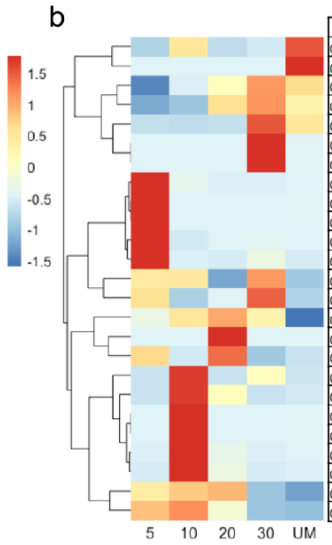


Figure S5. Co-occurrence network analysis of combined bacterial and fungal communities at different chronosequence ages. Each node represents a bacterial (red) or fungal OTU (blue), and an edge represents a spearman correlation with a statistically significant (FDR < 0.05) correlation coefficient > 0.9 or < -0.9. The size of each node is proportional to the square root of its degree.



	Phylum	Class	Order	Family	Genus	Species
OTU_230	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_1210	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_75	Verrucomicrobia	Pedosphaerae	Pedosphaerales	Ellin515	Unclassified	Unclassified
OTU_114	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes	Unclassified
OTU_4	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_25	Acidobacteria	Chloracidobacteria	RB41	Unclassified	Unclassified	Unclassified
OTU_2004	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_2	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes	Unclassified
OTU_864	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_10	Planctomycetes	Planctomycetia	Planctomycetales	Planctomycetaceae	Planctomyces	Unclassified
OTU_17	Proteobacteria	Betaproteobacteria	Unclassified	Unclassified	Unclassified	Unclassified
OTU_1954	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_27	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes	Unclassified
OTU_5	Verrucomicrobia	Spartobacteria	Chthoniobacterales	Chthoniobacteraceae	DA101	Unclassified
OTU_68	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes	Unclassified
OTU_8	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes	Unclassified
OTU_36	Acidobacteria	DA052	Ellin6513	Unclassified	Unclassified	Unclassified
OTU_42	Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	Unclassified
OTU_1	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium	Unclassified
OTU_29	Verrucomicrobia	Spartobacteria	Chthoniobacterales	Chthoniobacteraceae	Candidatus Xiphinematobacter	Unclassified
OTU_511	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes	Unclassified
OTU_33	Proteobacteria	Alphaproteobacteria	Rhizobiales	Unclassified	Unclassified	Unclassified
OTU_47	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_38	Acidobacteria	Acidobacteria-6	iii1-15	Unclassified	Unclassified	Unclassified
OTU_16	Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	Unclassified	Unclassified
OTU_43	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Unclassified	Unclassified
OTU_3	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Sinobacteraceae	Unclassified	Unclassified
OTU_23	Proteobacteria	Betaproteobacteria	Ellin6067	Unclassified	Unclassified	Unclassified
OTU_18	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Methylibium	Unclassified
OTU_6	Cyanobacteria	Chloroplast	Streptophyta	Unclassified	Unclassified	Unclassified
OTU_22	Actinobacteria	Thermoleophilia	Solirubrobacterales	Solirubrobacteraceae	Unclassified	Unclassified
OTU_7	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Hylemonella	Unclassified
OTU_21	Proteobacteria	Betaproteobacteria	IS-44	Unclassified	Unclassified	Unclassified
OTU_26	Chloroflexi	Ellin6529	Unclassified	Unclassified	Unclassified	Unclassified



	Phylum	Class	Order	Family	Genus	Species
OTU_12	Ascomycota	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
OTU_13	Basidiomycota	Agaricomycetes	Agaricales	Hygrophoraceae	Hygrophorus	Unclassified
OTU_6	Zygomycota	Incertae sedis	Mortierellales	Mortierellaceae	Mortierella	Mortierella epicladia
OTU_9	Ascomycota	Sordariomycetes	Hypocreales	Incertae sedis	Ilyonectria	Ilyonectria mors-panacis
OTU_997	Zygomycota	Incertae sedis	Mortierellales	Mortierellaceae	Mortierella	Mortierella rishikesha
OTU_30	Basidiomycota	Agaricomycetes	Agaricales	Psathyrellaceae	Lacrymaria	Lacrymaria lacrymabunda
OTU_2	Basidiomycota	Agaricomycetes	Agaricales	Entolomataceae	Entoloma	Unclassified
OTU_1	Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	Humicola	Unclassified
OTU_36	Ascomycota	Sordariomycetes	Sordariales	Lasiosphaeriaceae	Podospora	Podospora glutinans
OTU_33	Basidiomycota	Agaricomycetes	Agaricales	Hygrophoraceae	Hygrocybe	Hygrocybe conica
OTU_15	Ascomycota	Dothideomycetes	Capnodiales	Davidiellaceae	Unclassified	Unclassified
OTU_22	Ascomycota	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
OTU_3	Ascomycota	Leotiomycetes	Helotiales	Unclassified	Unclassified	Unclassified
OTU_20	Ascomycota	Sordariomycetes	Hypocreales	Bionectriaceae	Clonostachys	Clonostachys rosea f. catenulata
OTU_14	Rozellomycota	unidentified	unidentified	unidentified	unidentified	Rozellomycota sp
OTU_4	Basidiomycota	Agaricomycetes	Sebacinales	Sebacinaceae	unidentified	Helvellosebacina helvelloides
OTU_10	Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	Exophiala	Exophiala equina
OTU_7	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
OTU_16	Basidiomycota	Agaricomycetes	Thelephorales	Thelephoraceae	Unclassified	Unclassified
OTU_42	Basidiomycota	Agaricomycetes	Agaricales	Cortinariaceae	Cortinarius	Unclassified
OTU_32	Basidiomycota	Agaricomycetes	Thelephorales	Thelephoraceae	Unclassified	Unclassified
OTU_18	Basidiomycota	Agaricomycetes	Thelephorales	Thelephoraceae	Tomentella	Tomentella atramentaria
OTU_19	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
OTU_5	Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	Fusarium	Fusarium acutatum
OTU_8	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified

Figure S6. The abundance change and taxonomy of the most abundant OTUs that varied significantly across chronosequence ages for bacteria (a) and fungi (b). The key represents the z-scores of the relative abundance of the taxa. OTUs selected have relative abundance > 0.3% for bacteria or > 0.2% for fungi per sample on average, and are also significantly different between at least one pair of ages with FDR<0.05.

References

1. Humphries MD, Gurney K. 2008. Network 'small-world-ness': a quantitative method for determining canonical network equivalence. PloS one 3:e0002051.
2. Dezső Z, Barabási A-L. 2002. Halting viruses in scale-free networks. Physical Review E 65:055103.

Appendix B

Generic composition of BCGs in Chapter 5

BCG 1

Genus	Relative abundance	Genus	Relative abundance
Thermacetogenium	8.24%	Dokdonella	0.59%
Leptospira	1.18%	Longilinea	1.18%
Rhodoplanes	0.59%	Pasteuria	1.76%
Pseudolabrys	1.76%	Chitinophaga	0.59%
Prosthecobacter	1.76%	Caloramator	1.76%
Planctomyces	2.94%	Phycisphaera	1.18%
Pelotomaculum	1.18%	Cellvibrio	1.18%
Sphingomonas	1.18%	Aciditerrimonas	1.18%
Conexibacter	0.59%	Halochromatium	0.59%
Desulfotomaculum	3.53%	Zavarzinella	17.06%
Thermovibrio	0.59%	Gemmata	13.53%
Dongia	0.59%	Stella	0.59%
Singulisphaera	0.59%	Kushneria	0.59%
Phenylobacterium	0.59%	Verrucomicrobium	0.59%
Haliangium	0.59%	Geobacter	0.59%
Beijerinckia	0.59%	Caulobacter	0.59%
Azospirillum	0.59%	Parasegetibacter	0.59%
Cytophaga	0.59%	Leptospirillum	1.18%
Halorhodospira	1.18%	Thermodesulfovibrio	0.59%
Actinoplanes	0.59%	Luteibacter	0.59%
Phaselicystis	0.59%	Sporolituus	0.59%
Sediminibacterium	1.18%	Tistrella	0.59%
Moorella	0.59%	Thermodesulforhabdus	0.59%
Rubritalea	0.59%	Sporomusa	0.59%
Sporotomaculum	0.59%	Desulfomicrobium	1.18%
Desulfomonile	1.76%	Paenibacillus	0.59%
Caldalkalibacillus	1.18%	Thiodictyon	0.59%
Desulfococcus	0.59%	Methylocapsa	0.59%
Gemmatimonas	1.76%	Asticcacaulis	0.59%
Hyphomicrobium	0.59%	Gelria	0.59%
Chloroflexus	1.18%	Desulfurobacterium	0.59%
Anderseniella	0.59%	Thiomicrospira	0.59%
Opitutus	1.18%	Armatimonas/Armatimonadetes_gp1	0.59%

Mucilaginibacter	1.18%	Howardella	0.59%
Acetomicrobium	0.59%	Sulfobacillus	0.59%
Isosphaera	0.59%		

BCG2

Genus	Relative abundance	Genus	Relative abundance
Thermacetogenium	0.52%	Rhodococcus	0.26%
Leptospira	0.26%	Lysobacter	0.26%
Rhodoplanes	0.52%	Aciditerrimonas	1.04%
Nitrosococcus	0.78%	Methylovirgula	0.26%
Pseudolabrys	1.04%	Halochromatium	0.52%
Curvibacter	0.26%	Zavarzinella	1.04%
Desulfovibrio	1.04%	Gemmata	1.04%
Prosthecobacter	3.13%	Stella	0.26%
Planctomyces	0.78%	Desulfitobacterium	0.26%
Vasilyevaea	0.26%	Kushneria	0.26%
Pelotomaculum	2.09%	Adhaeribacter	0.52%
Solirubrobacter	0.52%	Desulfosoma	0.26%
Terrimonas	0.52%	Haliea	0.26%
Zoogloea	1.04%	Verrucomicrobium	1.04%
Thermanaeromonas	0.26%	Desertibacter	0.26%
Sphingomonas	0.26%	Niastella	0.26%
Ectothiorhodosinus	0.26%	Thermobispora	0.26%
Steroidobacter	1.31%	Thiocystis	0.26%
Conexibacter	2.87%	Ilumatobacter	0.78%
Desulfotomaculum	2.35%	Formivibrio	0.26%
Desulfovirgula	0.26%	Methyloversatilis	0.26%
Thiobacter	0.52%	Propionivibrio	0.26%
Dongia	0.26%	Bauldia	0.26%
Pyramidobacter	1.31%	Paludibacter	0.78%
Pelobacter	0.78%	Agromyces	0.26%
Dasania	0.26%	Anaeromyxobacter	0.26%
Alterococcus	1.83%	Streptosporangium	0.26%
Singulisphaera	0.52%	Azoarcus	0.26%
Phenylobacterium	1.04%	Parasegetibacter	0.52%
Rhodocista	1.31%	Gemella	0.26%
Lewinella	0.52%	Haloferula	0.52%
Chondromyces	1.57%	Thermodesulfovibrio	1.04%
Haliangium	1.04%	Schlesneria	0.26%
Nitrosospira	0.52%	Georgfuchsia	0.26%

Desulfonema	1.57%	Tepidamorphus	0.26%
Azospirillum	1.04%	Dendrosporobacter	0.78%
Cytophaga	1.57%	Hydrogenoanaerobacterium	0.26%
Rhizobium	0.52%	Fulvimonas	0.26%
Burkholderia	1.57%	Patulibacter	0.52%
Limnobacter	0.26%	Flavihumibacter	0.26%
Byssovorax	0.52%	Geothrix	0.26%
Archangium	0.26%	Kofleria	0.26%
Halorhodospira	0.26%	Dethiobacter	1.31%
Oceanospirillum	0.52%	Marinobacter	0.26%
Blastopirellula	0.26%	Desulfomicrobium	0.78%
Actinoplanes	0.78%	Catelliglobosispora	0.26%
Salicola	0.26%	Oceanibaculum	0.26%
Solitalea	1.57%	Amycolatopsis	0.52%
Ectothiorhodospira	0.52%	Stenotrophomonas	0.26%
Phaselicystis	0.78%	Caldilinea	0.26%
Sediminibacterium	0.26%	Cesiribacter	0.26%
Salinispora	0.26%	Escherichia/Shigella	0.52%
Rubritalea	0.26%	Kocuria	0.26%
Dyella	0.26%	Legionella	0.26%
Streptomyces	0.52%	Rhodothermus	0.78%
Desulfomonile	4.44%	Thiodictyon	0.26%
Desulfobotulus	1.31%	Granulosicoccus	0.52%
Caldalkalibacillus	0.26%	Rhodopirellula	0.26%
Desulfococcus	1.04%	Enhygromyxa	0.26%
Gemmatimonas	2.87%	Chelatococcus	0.26%
Hyphomicrobium	0.52%	Thalassomonas	0.26%
Thermovenabulum	0.26%	Acanthopleuribacter	0.26%
Labrys	0.26%	Thioalkalivibrio	0.52%
Actinomadura	0.26%	Sporichthya	0.26%
Desulfosalsimonas	0.78%	Thiohalocapsa	0.52%
Anderseniella	0.52%	Desulfurispora	0.26%
Micromonospora	0.26%	Aeromonas	0.26%
Ktedonobacter	0.26%	Geodermatophilus	0.26%
Flavisolibacter	1.04%	Kitasatospora	0.26%
Oxalophagus	0.26%	Nitratireductor	0.26%
Acetomicrobium	1.31%	Chthonomonas/Armatimonadetes_gp3	0.26%
Shewanella	0.26%	Wohlfahrtiimonas	0.26%
Bellilinea	0.78%	Heliobacterium	0.26%
Dokdonella	0.52%	Anaerobranca	0.26%
Longilinea	0.26%	Aquaspirillum	0.26%
Arenimonas	0.26%	Thermobrachium	0.26%

Pasteuria	0.78%	Alkalispirillum	0.26%
Chitinophaga	0.26%	Nocardiopsis	0.26%
Caloramator	0.52%	Thermoleophilum	0.26%
Caldicellulosiruptor	0.26%	Desulfacinum	0.26%
Denitratisoma	0.52%	Thermomonas	0.52%
Phycisphaera	0.78%	Tessaracoccus	0.26%
Cellvibrio	0.52%	Heliothrix	0.26%
Acidisphaera	0.26%	Aquisalibacillus	0.52%

BCG3

Genus	Relative abundance	Genus	Relative abundance
Rhodoplanes	0.93%	Pasteuria	0.93%
Pseudolabrys	1.87%	Phycisphaera	0.93%
Nitrobacter	0.93%	Thermosporothrix	0.93%
Thermodesulfobium	0.93%	Rhodococcus	0.93%
Prosthecobacter	1.87%	Lysobacter	0.93%
Sulfuritalea	0.93%	Rubrobacter	0.93%
Zoogloea	1.87%	Halochromatium	0.93%
Sphingomonas	0.93%	Zavarzinella	0.93%
Steroidobacter	1.87%	Verrucomicrobium	2.80%
Conexibacter	1.87%	Desertibacter	0.93%
Actinoallomurus	0.93%	Blastochloris	0.93%
Desulfotomaculum	0.93%	Vibrio	0.93%
Desulfovirgula	0.93%	Novosphingobium	0.93%
Thiobacter	0.93%	Ralstonia	0.93%
Pelobacter	0.93%	Anaeromyxobacter	0.93%
Alterococcus	2.80%	Streptosporangium	0.93%
Rhodocista	2.80%	Coprothermobacter	0.93%
Haliangium	2.80%	Kangiella	0.93%
Nitrosospira	0.93%	Thermodesulfovibrio	1.87%
Pseudomonas	0.93%	Collimonas	0.93%
Hydrogenophaga	0.93%	Caldicoprobacter	0.93%
Azospirillum	0.93%	Saccharothrix	0.93%
Rhizobium	1.87%	Stenotrophomonas	0.93%
Burkholderia	0.93%	Caldilinea	0.93%
Limnobacter	0.93%	Lentzea	0.93%
Rudaea	0.93%	Calditerricola	0.93%
Phaselicystis	0.93%	Actinophytocola	0.93%
Sphaerobacter	1.87%	Acanthopleuribacter	0.93%
Streptomyces	1.87%	Schlegelella	1.87%

Desulfomonile	4.67%	Kitasatospora	0.93%
Desulfobotulus	3.74%	Cardiobacterium	0.93%
Sphingobium	0.93%	Chthonomonas/Armatimonadetes_gp3	0.93%
Gemmatimonas	1.87%	Methylothermus	0.93%
Skermanella	0.93%	Pyxidicoccus	0.93%
Ktedonobacter	0.93%	Desulfovirga	0.93%
Massilia	0.93%	Pseudofulvimonas	0.93%
Acetomicrobium	3.74%	Schwartzia	0.93%
Bellilinea	2.80%	Methylophaga	0.93%

BCG4

Genus	Relative abundance	Genus	Relative abundance
Leptospira	0.30%	Arenimonas	0.30%
Nitrosococcus	1.78%	Pasteuria	0.89%
Pseudolabrys	0.89%	Caldicellulosiruptor	3.55%
Eubacterium	0.59%	Phycisphaera	0.89%
Desulfovibrio	0.30%	Cellvibrio	0.30%
Prosthecobacter	3.85%	Acidisphaera	0.30%
Planctomyces	1.18%	Actinocorallia	1.18%
Pelotomaculum	0.89%	Rhodococcus	1.48%
Mycobacterium	0.30%	Aciditerrimonas	0.89%
Sulfuritalea	0.30%	Butyricicoccus	0.30%
Terrimonas	0.30%	Methylovirgula	0.89%
Zoogloea	0.30%	Halochromatium	0.30%
Sphingomonas	1.18%	Zavarzinella	1.18%
Ectothiorhodosinus	0.59%	Stella	0.59%
Steroidobacter	1.78%	Amphibacillus	0.59%
Conexibacter	2.37%	Verrucomicrobium	1.78%
Actinoallomurus	2.66%	Selenomonas	0.30%
Ignatzschineria	0.59%	Niastella	0.30%
Desulfotomaculum	1.78%	Sinosporangium	0.30%
Desulfovirgula	1.18%	Nonomuraea	0.30%
Thiobacter	0.30%	Thiocystis	0.30%
Dongia	0.59%	Bauldia	0.30%
Tepidimicrobium	0.30%	Agromyces	0.30%
Pyramidobacter	0.59%	Anaeromyxobacter	0.30%
Alterococcus	2.96%	Fervidicoccus	0.30%
Singulisphaera	1.18%	Cellulosilyticum	0.30%
Parvibaculum	0.30%	Methylocystis	0.59%

Rhodocista	2.37%	Azoarcus	0.30%
Chondromyces	0.89%	Bosea	0.30%
Haliangium	0.59%	Haloferula	0.30%
Nitrosospira	0.30%	Thermodesulfovibrio	0.30%
Pseudomonas	0.30%	Georgfuchsia	0.30%
Beijerinckia	0.30%	Sporolituus	0.59%
Thermoanaerobacter	0.89%	Hydrogenoanaerobacterium	0.30%
Syntrophothermus	0.59%	Frateuria	0.30%
Desulfonema	0.30%	Tistrella	0.30%
Azospirillum	0.89%	Thiorhodococcus	0.30%
Ruminococcus	0.30%	Angustibacter	0.30%
Cytophaga	0.30%	Kofleria	0.30%
Burkholderia	0.59%	Caldicoprobacter	0.30%
Limnobacter	0.59%	Thermofilum	0.30%
Sulfurihydrogenibium	2.66%	Sporomusa	0.30%
Byssovorax	0.89%	Craurococcus	0.30%
Halorhodospira	1.18%	Desulfomicrobium	0.30%
Blastopirellula	0.59%	Catelliglobosispora	0.30%
Actinoplanes	0.59%	Oceanibaculum	1.18%
Salicola	1.48%	Acidisoma	0.59%
Solitalea	0.30%	Brevundimonas	0.30%
Ectothiorhodospira	0.30%	Crossiella	0.30%
Phaselicystis	0.30%	Legionella	0.30%
Sphaerobacter	0.30%	Subdoligranulum	0.30%
Sediminibacterium	0.30%	Mucispirillum	0.30%
Mesorhizobium	0.30%	Acidiphilium	0.30%
Salinispora	0.30%	Methylosinus	0.30%
Moorella	0.89%	Catenulispora	0.30%
Rubritalea	0.30%	Bdellovibrio	0.30%
Dyella	0.30%	Thioalkalivibrio	0.30%
Streptomyces	0.59%	Paracoccus	0.30%
Desulfomonile	2.07%	Kineococcus	0.30%
Desulfobotulus	0.89%	Desulfurispora	0.30%
Caldalkalibacillus	0.30%	Ferrimicrobium	0.30%
Ruminobacter	0.30%	Levilinea	0.30%
Desulfococcus	0.30%	Wohlfahrtiimonas	0.30%
Gemmatimonas	2.37%	Williamsia	0.30%
Skermanella	0.89%	Ammonifex	0.30%
Methylosoma	0.30%	Methylothermus	0.30%
Hyphomicrobium	0.30%	Pyxidicoccus	0.30%
Chloroflexus	0.89%	Thermoleophilum	0.30%
Actinomadura	1.48%	Albimonas	0.30%
Ktedonobacter	0.89%	Erythrobacter	0.30%

Rhodopila	0.89%	Armatimonas/Armatimonadetes_gp1	0.59%
Flavisolibacter	0.30%	Acholeplasma	0.30%
Ferruginibacter	0.30%	Litoricola	0.30%
Acetomicrobium	1.48%	Thalassobaculum	0.30%
Longilinea	0.30%	Carboxydothermus	0.30%
Perlucidibaca	0.59%	Allocatelliglobosispora	0.30%

BCG5

Genus	Relative abundance	Genus	Relative abundance
Variovorax	0.34%	Nisaea	0.69%
Mycoplasma	0.34%	Niastella	1.72%
Prosthecobacter	1.38%	Thioalkalibacter	0.34%
Planctomyces	2.41%	Devosia	0.69%
Vasilyevaea	0.69%	Thiocystis	0.34%
Pelotomaculum	0.69%	Segetibacter	0.69%
Mycobacterium	0.34%	Siphonobacter	0.34%
Solirubrobacter	4.83%	Bauldia	0.34%
Terrimonas	2.07%	Anaeromyxobacter	0.34%
Zoogloea	0.69%	Catenuloplanes	0.34%
Sphingomonas	2.76%	Bosea	0.34%
Ectothiorhodosinus	0.34%	Actinomycetospora	0.34%
Steroidobacter	1.03%	Haloferula	0.34%
Conexibacter	6.21%	Saxeibacter	0.34%
Pyramidobacter	1.03%	Acetobacterium	0.34%
Alterococcus	0.69%	Branchiibius	0.34%
Phenylobacterium	0.34%	Buttiauxella	0.34%
Rhodocista	0.34%	Kaistia	0.34%
Chondromyces	0.34%	Fulvimonas	0.34%
Haliangium	2.41%	Acidovorax	0.34%
Nitrosospira	0.34%	Subtercola	0.34%
Pseudonocardia	1.03%	Methylobacterium	0.69%
Azospirillum	0.34%	Kribbella	0.34%
Humicoccus	0.34%	Nakamurella	0.34%
Cytophaga	1.38%	Angustibacter	0.34%
Rhizobium	1.03%	Rhodomicrobium	0.34%
Burkholderia	0.34%	Sporomusa	0.69%
Chryseoglobus	0.34%	Dethiobacter	0.34%
Byssovorax	1.03%	Euzebya	0.34%
Blastopirellula	1.03%	Amycolatopsis	0.34%

Actinoplanes	1.38%	Cryptosporangium	0.69%
Cellulomonas	0.34%	Stenotrophomonas	0.34%
Phaselicystis	1.03%	Balneimonas	0.34%
Sphaerobacter	0.34%	Brevundimonas	0.69%
Dactylosporangium	0.34%	Hydrotalea	0.34%
Mesorhizobium	1.03%	Cesiribacter	0.34%
Salinispora	0.34%	Chitinibacter	0.34%
Porphyrobacter	0.34%	Sphaerisporangium	0.34%
Rubritalea	1.03%	Luteimonas	0.34%
Dyella	0.34%	Kocuria	0.34%
Desulfomonile	0.69%	Myxococcus	0.34%
Sphingobium	0.34%	Leptolinea	0.34%
Nocardioides	3.45%	Altererythrobacter	0.34%
Gemmatimonas	4.48%	Tistlia	0.34%
Skermanella	0.34%	Modestobacter	0.34%
Kineosporia	1.03%	Demequina	0.34%
Labrys	0.34%	Sorangium	0.34%
Desulfosalsimonas	0.34%	Fibrobacter	0.34%
Anderseniella	0.34%	Cupriavidus	0.34%
Micromonospora	0.34%	Chelatococcus	0.34%
Ktedonobacter	0.34%	Cycloclasticus	0.34%
Opitutus	0.34%	Desulfitibacter	0.34%
Mucilaginibacter	0.34%	Schlegelella	0.34%
Flavisolibacter	1.03%	Terrabacter	0.34%
Afipia	0.34%	Asticcacaulis	0.34%
Acetomicrobium	0.69%	Marmoricola	0.69%
Bellilinea	1.03%	Kiloniella	0.34%
Longilinea	0.34%	Pleomorphomonas	0.34%
Virgisporangium	0.34%	Ferrimicrobium	0.34%
Chitinophaga	0.34%	Caminibacter	0.34%
Rhodococcus	0.34%	Halothiobacillus	0.69%
Lysobacter	0.69%	Dyadobacter	0.34%
Arthrobacter	0.34%	Finegoldia	0.34%
Aciditerrimonas	1.72%	Phycicoccus	0.34%
Microvirga	0.69%	Umezawaea	0.34%
Bacillus	0.34%	Propionispira	0.34%
Zavarzinella	0.69%	Chryseobacterium	0.34%
Gemmata	0.34%	Pseudoxanthomonas	0.34%
Janibacter	0.34%	Sphingopyxis	0.34%
Crabtreella	0.34%	Amaricoccus	0.34%
Amphibacillus	0.34%	Caldanaerobius	0.34%
Kushneria	1.03%	Aeromicrobium	0.34%
Couchioplanes	0.34%	Olsenella	0.34%

Haliea	0.34%	Chelativorans	0.34%
Selenomonas	0.34%	Belnapia	0.34%
Spirillospora	0.34%	Jiangella	0.34%
Caulobacter	0.34%	Desulfothermus	0.34%

BCG6

Genus	Relative abundance	Genus	Relative abundance
Acetivibrio	0.37%	Labrys	0.37%
Acetobacterium	0.75%	Laceyella	0.37%
Acetomicrobium	0.37%	Legionella	1.87%
Acholeplasma	0.37%	Leisingera	0.37%
Acidiphilium	0.37%	Levilinea	1.12%
Acidisoma	0.37%	Lewinella	0.37%
Aciditerrimonas	0.37%	Luedemannella	0.37%
Acidothermus	0.37%	Luteimonas	0.37%
Actinocorallia	0.75%	Luteolibacter	0.37%
Aeromicrobium	0.37%	Lysobacter	0.37%
Alloactinosynnema	0.37%	Martelella	0.37%
Alterococcus	1.87%	Massilia	0.37%
Ammonifex	0.37%	Meniscus	0.37%
Anaeromyxobacter	1.12%	Methylocapsa	0.37%
Anaerospora	0.37%	Methylopila	0.37%
Archangium	1.12%	Mogibacterium	0.37%
Arthrobacter	0.37%	Mycobacterium	0.37%
Aurantimonas	0.75%	Mycoplasma	0.37%
Azoarcus	0.75%	Neokomagataea	0.37%
Bacillus	1.87%	Niastella	1.12%
Bauldia	0.75%	Nitrosomonas	0.37%
Bdellovibrio	0.75%	Nitrosospira	0.37%
Bellilinea	2.24%	Nitrospira	0.75%
Blastopirellula	0.37%	Opitutus	0.37%
Bordetella	0.37%	Paenibacillus	1.49%
Burkholderia	0.37%	Paludibacter	0.37%
Byssovorax	0.75%	Paraliobacillus	0.37%
Caldalkalibacillus	0.75%	Parasegetibacter	0.37%
Caldimonas	0.37%	Parvibaculum	0.37%
Cellulosimicrobium	0.37%	Pasteuria	0.37%
Cellvibrio	0.37%	Pedomicrobium	0.37%
Chitinophaga	0.75%	Pelobacter	0.75%
Chloroflexus	0.37%	Pelomonas	0.37%

Chondromyces	1.49%	Pelotomaculum	1.12%
Chryseobacterium	0.37%	Phaselicystis	0.75%
Cohnella	0.37%	Phenylobacterium	0.37%
Conexibacter	0.75%	Phycisphaera	0.37%
Crossiella	0.37%	Planctomyces	1.49%
Cryptosporangium	0.37%	Plantactinospora	0.37%
Cupriavidus	0.37%	Porphyrobacter	0.37%
Cystobacter	0.37%	Propionivibrio	0.37%
Cytophaga	1.12%	Prosthecobacter	5.97%
Dactylosporangium	0.37%	Pseudochrobactrum	0.37%
Dechloromonas	0.37%	Pseudolabrys	0.37%
Denitratisoma	0.37%	Pseudomonas	0.37%
Derxia	0.37%	Pyramidobacter	0.75%
Desulfobacca	0.37%	Pyxidicoccus	0.75%
Desulfobotulus	1.12%	Rhodocytophaga	0.37%
Desulfococcus	0.37%	Rhodomicrobium	0.37%
Desulfomonile	0.75%	Rhodopila	0.37%
Desulfonema	0.75%	Rhodoplanes	0.37%
Desulfosalsimonas	0.75%	Rubrobacter	0.75%
Desulfovibrio	0.75%	Rudaea	0.75%
Dethiobacter	0.75%	Rugosimonospora	0.37%
Dethiosulfovibrio	0.37%	Ruminococcus	0.75%
Dictyoglomus	0.37%	Salicola	0.37%
Dokdonella	0.37%	Serratia	0.37%
Dongia	0.75%	Simkania	0.37%
Dyella	0.37%	Skermanella	0.75%
Ectothiorhodosinus	0.37%	Solirubrobacter	0.75%
Ectothiorhodospira	0.37%	Solitalea	0.37%
Ferruginibacter	0.37%	Sphingobium	0.37%
Flavisolibacter	0.37%	Sphingomonas	0.75%
Gemmata	0.37%	Spirochaeta	0.37%
Gemmatimonas	1.87%	Steroidobacter	0.37%
Geobacter	2.24%	Streptomyces	0.37%
Geothrix	1.12%	Sulfobacillus	0.37%
Gluconobacter	0.37%	Sulfuricella	0.37%
Gracilibacter	0.37%	Sulfurihydrogenibium	0.37%
Haematobacter	0.37%	Sulfuritalea	0.75%
Haliangium	2.24%	Syntrophothermus	0.37%
Haloferula	0.37%	Terrimonas	1.49%
Halothiobacillus	0.37%	Thauera	0.37%
Holophaga	0.37%	Thioalkalispira	0.37%
Hyphomicrobium	0.37%	Tistrella	0.37%
Kibdelosporangium	0.37%	Vampirovibrio	0.37%

Kofleria	0.75%	Verrucomicrobium	1.87%
Kushneria	0.37%	Zavarzinella	2.24%

BCG7

Genus	Relative abundance	Genus	Relative abundance
Acetomicrobium	0.88%	Longilinea	0.29%
Acidisoma	0.29%	Luteolibacter	1.18%
Acidisphaera	0.29%	Marinilabilia	0.29%
Aciditerrimonas	0.59%	Marinobacter	0.29%
Acidithiobacillus	0.29%	Marinomonas	0.29%
Acidothermus	0.29%	Methylophilus	0.29%
Adhaeribacter	0.29%	Mucilaginibacter	0.29%
Aeromonas	0.29%	Naxibacter	0.29%
Alterococcus	1.76%	Niabella	2.35%
Aquabacterium	0.29%	Niastella	0.59%
Aquimonas	0.29%	Nisaea	0.29%
Archangium	0.29%	Nitrosococcus	0.59%
Arenimonas	0.88%	Nitrosospira	0.88%
Asanoa	0.88%	Novosphingobium	0.59%
Azoarcus	0.29%	Oceanibaculum	0.29%
Azospirillum	0.59%	Opitutus	1.47%
Bacillus	0.29%	Paenibacillus	0.29%
Bacteriovorax	0.29%	Paludibacter	1.18%
Bdellovibrio	0.29%	Pandoraea	0.29%
Bellilinea	0.59%	Paracoccus	0.29%
Blastopirellula	0.29%	Parasegetibacter	0.59%
Brooklawnia	0.29%	Pasteuria	0.59%
Burkholderia	0.59%	Patulibacter	0.29%
Byssovorax	1.76%	Pedobacter	0.88%
Caldimicrobium	0.29%	Pelagicoccus	0.29%
Caloramator	0.29%	Pelobacter	0.59%
Caulobacter	0.29%	Phaselicystis	1.76%
Cellvibrio	0.59%	Phenylobacterium	0.59%
Cesiribacter	1.18%	Piscinibacter	0.29%
Chitinophaga	1.47%	Polaromonas	0.29%
Chlorobium	0.29%	Porphyrobacter	0.29%
Chondromyces	3.24%	Prosthecobacter	1.47%
Chryseobacterium	0.29%	Proteiniclasticum	0.29%
Corallococcus	0.29%	Pseudochrobactrum	0.29%
Crabtreella	0.29%	Pseudomonas	1.18%

Craurococcus	0.59%	Pseudorhodoferax	0.29%
Curvibacter	0.29%	Pyramidobacter	0.29%
Cytophaga	3.53%	Rhizobacter	0.29%
Desertibacter	0.29%	Rhizobium	0.88%
Desulfitobacterium	0.29%	Rhodobacter	0.88%
Desulfobacterium	0.29%	Rhodococcus	0.29%
Desulfobotulus	0.29%	Rhodoplanes	0.29%
Desulfomonile	1.76%	Roseivirga	0.29%
Desulfonatronum	0.29%	Rubrivivax	0.29%
Desulfonema	0.29%	Rudaea	0.88%
Desulfotomaculum	0.29%	Salicola	0.29%
Desulfurispora	0.29%	Sediminibacterium	2.35%
Dethiosulfovibrio	0.59%	Segetibacter	0.29%
Dokdonella	0.29%	Singulisphaera	2.94%
Dongia	0.59%	Siphonobacter	0.29%
Duganella	0.29%	Solitalea	4.12%
Dyadobacter	0.29%	Sphingomonas	0.88%
Dyella	0.29%	Spirochaeta	0.29%
Ferruginibacter	3.82%	Sporocytophaga	0.29%
Filimonas	0.29%	Steroidobacter	0.29%
Flavisolibacter	0.29%	Streptomyces	0.29%
Flavobacterium	1.47%	Sulfuritalea	0.59%
Flectobacillus	0.59%	Syntrophothermus	0.29%
Fulvivirga	0.29%	Tepidimicrobium	0.29%
Gemmatimonas	3.24%	Terrimonas	2.65%
Glycomyces	0.29%	Thauera	0.88%
Haliangium	3.24%	Thermoanaerobacter	0.29%
Haliea	0.29%	Thermosporothrix	0.29%
Haliscomenobacter	0.29%	Thiomonas	0.29%
Hirschia	0.59%	Thiorhodococcus	0.29%
Hydrogenophaga	0.29%	Turneriella	0.29%
Hydrotalea	1.18%	Uliginosibacterium	0.29%
Iamia	0.29%	Vasilyevaea	0.29%
Ilumatobacter	0.59%	Verrucomicrobium	0.29%
Kushneria	0.29%	Vogesella	0.29%
Legionella	0.29%	Zavarzinella	0.29%
Lentisphaera	0.29%	Zoogloea	0.59%
Leptospira	0.29%	Zooshikella	0.29%
Lewinella	1.76%		

BCG8

Genus	Relative abundance	Genus	Relative abundance
Acetobacterium	0.34%	Isosphaera	1.03%
Acetomicrobium	0.34%	Jeotgalibacillus	0.34%
Acidisoma	0.34%	Ktedonobacter	0.34%
Acidithiobacillus	1.03%	Levilinea	0.34%
Adhaeribacter	0.34%	Luteimonas	0.34%
Algicola	0.68%	Mesorhizobium	0.34%
Alterococcus	0.68%	Mucilaginibacter	0.34%
Amorphus	0.34%	Mycobacterium	0.34%
Amphibacillus	0.34%	Nannocystis	0.34%
Anderseniella	0.34%	Naxibacter	0.34%
Azospirillum	0.68%	Niastella	0.34%
Bacillus	0.34%	Nitrobacter	0.34%
Bauldia	0.34%	Opitutus	0.68%
Bellilinea	2.05%	Paraliobacillus	0.34%
Blastopirellula	3.08%	Parvibaculum	2.05%
Branchiibius	0.34%	Pasteuria	2.74%
Byssovorax	0.68%	Patulibacter	0.34%
Caldalkalibacillus	1.71%	Pedobacter	0.34%
Caldicellulosiruptor	0.68%	Pedomicrobium	0.34%
Caldilinea	0.34%	Pelagibius	0.68%
Caloramator	0.34%	Phenylobacterium	0.34%
Cellulosilyticum	0.34%	Phycisphaera	0.68%
Cellvibrio	0.68%	Planctomyces	7.19%
Chitinophaga	0.68%	Planifilum	0.34%
Chloroflexus	0.34%	Plesiomonas	0.34%
Chondromyces	0.34%	Prosthecobacter	1.71%
Conexibacter	1.37%	Prosthecochloris	0.34%
Cytophaga	0.34%	Psychrobacter	0.34%
Desertibacter	0.34%	Rhodocista	0.34%
Desulfobacca	0.34%	Rhodococcus	2.40%
Desulfobotulus	0.68%	Roseibacillus	0.34%
Desulfobulbus	0.34%	Rudaea	0.34%
Desulfococcus	1.71%	Rugosimonospora	0.34%
Desulfoluna	0.34%	Ruminobacter	0.34%
Desulfomonile	1.03%	Salinispora	0.34%
Desulfonema	0.68%	Salmonella	0.68%
Desulfosalsimonas	0.34%	Shewanella	0.34%
Desulfotomaculum	2.74%	Singulisphaera	0.34%
Desulfovibrio	1.71%	Solitalea	0.34%

Dethiobacter	0.34%	Sphaerobacter	1.71%
Dongia	2.74%	Sphingomonas	1.03%
Dyella	0.34%	Spirochaeta	0.34%
Edwardsiella	0.34%	Sporolituus	1.37%
Ferruginibacter	0.34%	Sporomusa	0.34%
Fibrobacter	0.34%	Stenotrophomonas	0.34%
Flavihumibacter	0.34%	Streptomyces	0.34%
Flavitalea	0.34%	Syntrophothermus	0.68%
Flexibacter	0.34%	Tepidimicrobium	0.34%
Fodinicurvata	0.34%	Thermacetogenium	4.11%
Gelria	0.34%	Thermoanaerobacter	0.68%
Gemmata	3.08%	Thermodesulfobium	0.34%
Gemmatimonas	1.03%	Thermodesulfovibrio	0.68%
Geobacillus	0.34%	Thermoflavimicrobium	0.68%
Geobacter	0.34%	Thermovenabulum	0.34%
Granulosicoccus	0.34%	Thiodictyon	0.68%
Haliangium	0.34%	Thioflavicoccus	2.05%
Halochromatium	0.34%	Tistlia	0.34%
Haloferula	0.34%	Verrucomicrobium	0.68%
Halorhodospira	0.68%	Verrucosispora	0.34%
Hyphomicrobium	0.34%	Zavarzinella	9.59%

BCG9

Genus	Relative abundance	Genus	Relative abundance
Aciditerrimonas	1.54%	Nitrobacter	3.08%
Afipia	1.54%	Pedomicrobium	3.08%
Aquisalibacillus	1.54%	Pelobacter	1.54%
Azospirillum	4.62%	Peptostreptococcaceae_incertae_sedis	1.54%
Bacillus	3.08%	Phaselicystis	3.08%
Brachyspira	1.54%	Pilimelia	1.54%
Bradyrhizobium	6.15%	Prosthecobacter	4.62%
Chitinophaga	1.54%	Pseudomonas	1.54%
Chondromyces	1.54%	Pyramidobacter	1.54%
Conexibacter	3.08%	Pyxidicoccus	1.54%
Craurococcus	1.54%	Rhodoplanes	6.15%
Desulfomonile	1.54%	Singulisphaera	1.54%
Desulfotomaculum	1.54%	Skermanella	1.54%
Dictyoglomus	1.54%	Solirubrobacter	1.54%
Gemmata	1.54%	Spirochaeta	1.54%

Halochromatium	1.54%	Sporolituus	4.62%
Halothiobacillus	1.54%	Terrimonas	1.54%
Hyphomicrobium	7.69%	Thermoanaerobacter	1.54%
Kaistia	1.54%	Thermodesulfovibrio	1.54%
Legionella	1.54%	Thiocystis	1.54%
Micromonospora	1.54%	Variovorax	1.54%
Niastella	1.54%	Zavarzinella	1.54%

BCG10

Genus	Relative abundance	Genus	Relative abundance
Acetomicrobium	1.71%	Leptospirillum	1.71%
Actinoallomurus	1.71%	Lutispora	0.85%
Actinomadura	0.85%	Methylocystis	0.85%
Actinoplanes	0.85%	Methylosoma	0.85%
Alterococcus	1.71%	Naxibacter	0.85%
Anaeromyxobacter	2.56%	Nitrosospira	0.85%
Azospirillum	3.42%	Nitrospira	0.85%
Blastomonas	0.85%	Pasteuria	0.85%
Burkholderia	0.85%	Planctomyces	2.56%
Caldalkalibacillus	1.71%	Pleomorphomonas	0.85%
Chondromyces	0.85%	Prosthecobacter	5.98%
Chthonomonas/Armatimonadetes_gp3	0.85%	Pseudolabrys	0.85%
Conexibacter	0.85%	Pyramidobacter	0.85%
Desertibacter	0.85%	Rhodocista	1.71%
Desulfobotulus	2.56%	Rhodococcus	0.85%
Desulfomonile	1.71%	Rhodomicrobium	0.85%
Desulfonema	0.85%	Rhodoplanes	0.85%
Desulfotomaculum	0.85%	Roseomonas	0.85%
Dethiosulfovibrio	0.85%	Rubrobacter	0.85%
Ethanoligenens	0.85%	Ruminococcus	0.85%
Ferrimicrobium	0.85%	Singulisphaera	0.85%
Ferruginibacter	0.85%	Skermanella	2.56%
Gemmata	11.97%	Sphingomonas	0.85%
Gemmatimonas	4.27%	Stella	2.56%
Haliangium	1.71%	Steroidobacter	0.85%
Halorhodospira	0.85%	Sulfuritalea	0.85%
Isosphaera	0.85%	Thermacetogenium	0.85%
Kangiella	0.85%	Thermoanaerobacter	0.85%
Kofleria	0.85%	Thermodesulforhabdus	0.85%

Ktedonobacter	2.56%	Thermogemmatispora	0.85%
Leptonema	0.85%	Zavarzinella	8.55%

BCG11

Genus	Relative abundance	Genus	Relative abundance
Archangium	0.91%	Moorella	0.91%
Bacillus	0.91%	Mycoplasma	0.91%
Bradyrhizobium	0.91%	Parvibaculum	0.91%
Brevundimonas	0.91%	Pasteuria	0.91%
Caldalkalibacillus	0.91%	Patulibacter	0.91%
Caldicellulosiruptor	0.91%	Pedobacter	1.82%
Chitinophaga	0.91%	Pelagibius	0.91%
Chloroflexus	0.91%	Persephonella	0.91%
Chondromyces	0.91%	Phycisphaera	1.82%
Coprothermobacter	0.91%	Planctomyces	6.36%
Corynebacterium	0.91%	Prosthecobacter	0.91%
Desulfotomaculum	0.91%	Pseudochrobactrum	0.91%
Desulfovibrio	0.91%	Pseudomonas	0.91%
Escherichia/Shigella	0.91%	Rhodococcus	0.91%
Fervidicoccus	0.91%	Rubrobacter	0.91%
Filimonas	0.91%	Rudaea	0.91%
Gemmata	7.27%	Salicola	0.91%
Gemmatimonas	0.91%	Sediminibacterium	0.91%
Halarsenatibacter	0.91%	Segetibacter	0.91%
Haliangium	0.91%	Simkania	0.91%
Haliscomenobacter	0.91%	Singulisphaera	14.55%
Halospina	0.91%	Sneathiella	0.91%
Hydrotalea	0.91%	Solitalea	3.64%
Kushneria	0.91%	Sphingomonas	0.91%
Lactobacillus	0.91%	Terrimonas	2.73%
Legionella	1.82%	Thermacetogenium	2.73%
Leptospira	0.91%	Thermodesulfobium	0.91%
Leptospirillum	0.91%	Thermogymnomonas	0.91%
Lewinella	0.91%	Thiocystis	0.91%
Massilia	0.91%	Zavarzinella	11.82%