Ecological and Evolutionary Perspectives on Bacterial Resource Use

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA BY

Pamela Weisenhorn

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Sarah Hobbie

August 2014

Acknowledgements

I would like to thank my adviser Sarah Hobbie for her enthusiasm, patience, and support of both me and my project, even as it drifted farther and farther from her area of research. Thanks also to Chris Henry, for taking up the role of adviser as my project progressed and for helping me to develop my knowledge and skills in metabolic modeling. Thanks to Peter Minchin (University of Illinois Edwardsville) for statistical advice and discussions of how to approach comparisons of large genomic datasets that proved crucial to my dissertation work. Thanks to my committee, Mike Travisano, Tom Schmidt (University of Michigan), and Peter Reich, for being generous with their time, their wise feedback, and their gentle corrections. And, mostly, for challenging me to think more deeply and broadly about the work I was doing, its context, and its relevance. With your help, I was able to expand the scope of my research and gain a new perspective on ways to examine organisms' interactions with their environment. My graduate experience was much richer for this and I am a better researcher for your efforts.

I am also thankful to many faculty and researchers at LSU for helping me develop the foundation on which this dissertation work was built. Many thanks to Loretta Battaglia (now at Southern Illinois Carbondale) for my trial by fire induction into the world of research. Your strong spirit and determination to coax (and pry when necessary) my thoughts and ideas into the open helped me to find my intellectual voice and take the very first steps towards independence. Many thanks to my advising trinity for the majority of my time at LSU: Irv Mendelssohn, Marc Cohn, and Dave Longstreth. For years the three of you took turns bearing the brunt of my questions and doubts. You gave me the space and support I needed to being developing my skills in experimental design, critical evaluation of literature and ideas, and general research practices. My thanks to Kyle Harms and CET Paine (Sterling University, UK) for giving me the research opportunity of my dreams in Peru, and for reminding me to always reach for the stars. And, lastly, to Kevin Carman (now at University of Nevada Reno) for believing in me and always giving me a place to land and recover when life got in the way of research.

Finally, I would like to thank Vinnie for his support and love these past few years. His encouragement in the end was what made completing this dissertation possible. Well, that and all the cooking and cleaning he took care of in the final push to completion. Thank you to my friends and support network (you know who you are) for commiserating with me about the trials, and triumphs, of the graduate experience. I would also like to thank my family for their love and support, in spite of my crazy decision to risk life and limb living somewhere white stuff falls from the sky.

Abstract

Bacterial metabolism mediates many biochemical transformations important to the stability and health of a diverse range of ecosystem types. In my dissertation, I examine the evolutionary and ecological context of a subset of bacterial metabolic pathways related to energy and metabolic precursor production that are crucial for bacterial growth. Specifically, I examine whether these pathways are conserved across a large, phylogenetically diverse set of organisms, whether related organisms respond similarly to differences in resource inputs, and whether knowledge of these pathways or phylogenetic relatedness can aid in the prediction of bacterial growth rates across a wide range of C substrates. While I found only a weak phylogenetic signal in the presence or absence of these pathways, there was strong evidence that constraints have limited the number of observed combinations of these pathways. Only 265 (6.5%) of the 4096 potential pathway combinations were found in this dataset of 8178 genomes. I propose this may suggest strong environmental selection acting to rapidly change pathway presence or absence, regardless of past evolutionary history. In order for this suggestion to be feasible, organisms must respond to their environment in a phylogeny-independent manner. To address this, I compared taxa response using 16S amplicon libraries from plots with substantial variation in C and N availability resulting from plant species identity in a long-term field experiment. I found an inconsistent response of soil bacteria at higher taxonomic levels to resource variation, in agreement with organisms responding to environment in a phylogeny-independent manner. I then cultured 56 bacterial isolates from these plots to examine the relative strength of phylogeny versus metabolic pathways in explaining growth responses of isolates across a range of substrates. Phylogenetic relatedness and similarities in energy metabolism each explained about 30% of the observed variation in patterns of bacterial growth, with about 50% overlap between the two approaches. Both phylogeny and energy metabolism are important in determining bacterial growth; however, environmental selection may lead to convergence towards a small number of ecotypes within a system despite high levels of phylogenetic diversity. The strength and consequences of such environmental optimization of metabolism warrant further study.

Table of Contents

List of Tables iv

List of Figures v

List of Supplemental Tables vii

Introduction 1

Chapter 1 4

Chapter 2 11

Chapter 3 20

Illustrations 30

Bibliography 56

List of Tables

Table 1-1. Measures of dispersion for all MC types across the phylogenetic tree was Brownian or near-Brownian. 30

Table 1-2. Measures of dispersion for the twelve individual metabolic pathways comprising MC categorization across the phylogenetic tree was Brownian or near-Brownian. 31

Table 1-3. Measures of dispersion for all ETC types across the phylogenetic tree was Brownian (fermentative) or near-Brownian (all others). 32

Table 2-1. Relationship of resource quantity and chemistry on changes in the abundance of highly variable and highly abundant taxa 40

Table 3-1. Detailed specification of individual pathway presence and absence for each abundant MC type (A-L) 47

Table 3-2. Average positive growth rate ($APGR_{CC}$, change in log(OD600)/hour) on individual substrate classes by ES type 48

Table 3-3. Average positive growth rate ($APGR_{CC}$, change in log(OD600)/hour) on individual substrate classes by MC type. 49

Table 3-4. Maximum observed growth rate (MGR, change in log(OD600)/hour), average positive growth rate (APGR, change in log(OD600)/hour), and metabolic flexibility for isolates containing (+) or lacking (-) individual metabolic pathways. 50

Table 3-5. Average positive growth rate ($APGR_{CC}$, change in log(OD600)/hour) on individual substrate classes by PC type. 51

List of Figures

Figure 1-1. Schematic pathway map showing the major sugar processing pathways and fermentation pathways included in the metabolic chassis. 33

Figure 1-2. Flow chart of the CMM development pipeline..... 34

Figure 1-3. Schematic demonstrating variation in the degree of phylogenetic conservation of a hypothetical trait with states A and B. 35

Figure 1-4. Frequency distribution of CMMs by MC type. 36

Figure 1-5. Distribution of three of the sugar degradation pathways (glycolysis, gluconeogenesis, and ED) on a 16S phylogeny. 37

Figure 1-6. Distribution of ETC type on a 16S phylogeny 38

Figure 1-7. Nonmetric multidimensional scaling ordinations showing variation in the distribution of MC types..... 39

Figure 2-1. Rarefaction curves for A) OTU_{98} richness and B) Shannon's diversity..... 41

Figure 2-2. The four resources that best captured variation in resource landscape among plant species were chosen out of the 35 resources examined via nonmetric multidimensional scaling ordinations..... 42

Figure 2-3. Effects of belowground resource quantity and chemistry on OTU_{98} community composition. 43

Figure 2-4. Effects of resource availability on soi bacterial community composition at the A) family and B) phyla level. 44

Figure 2-5. Effects of resource availability on community composition of the 16 most abundant taxa. 45

Figure 2-6. Relationship between ecosystem process rates and OTU_{98} community composition. 46

Figure 3-1. Correlation between metabolic flexibility and both A) the maximum observed growth rate and B) average positive growth rates. 52

Figure 3-2. Differences among metabolic chassis types in A) maximum observed growth rate (MGR, change in log(OD600)/hour), B) average positive growth rate (APGR, change in log(OD600)/hour), C) metabolic flexibility, and D) genome size. 53

Figure 3-3. Differences among isolates containing A) glycolysis, B) ethanol fermentation, and C) ED in average positive growth rate (APGR, change in log(OD600)/hour). 54

Figure 3-4. Differences among isolates from different families (B=Burkholderiaceae, M=Micrococcaceae, and X=Xanthomonadaceae) in A) maximum observed growth rate (MGR), B) average positive growth rate (APGR), C) metabolic flexibility, and D) genome size. 55

List of Supplemental Tables

Supplemental Table 1. List of PubSEED IDs and taxonomic classification of genomes for which CMMs were built. All genomes are publicly available. Table available online.

Supplemental Table 2. Electron transport chain additions to the GPR of the core model template. These additional reactions were added (by J. Edirisinghe) to support the variety of electron transport chains found in phylogenetically diverse microbes. Table available online.

Supplemental Table 3. Core model template gene-protein reaction network containing all reactions relevant for pathways in the metabolic chassis and electron transport chains. This set of reactions was manually curated by J. Edirisinghe. Table available online.

Supplemental Table 4. Individual pathway presence or absence for each of the 265 MC types observed. N is the number of genomes with that MC type. Table available online.

Introduction

Microorganisms are key mediators of biogeochemical processes and microbial metabolism is directly tied into ecologically important biochemical transformations. Microorganisms have been implicated in everything from the oxidation of the atmosphere (Kopp et al., 2005) to the bulk of nitrogen fixation and denitrification (Firestone and Davidson, 1989, Gubry-Rangin, 2010) to cave formation via weathering (Northup and Lavoie, 2001). Bacteria are important decomposers and are responsible for processing the vast majority of carbon fixed via plant primary production (Cebrian, 1999, Cyr and Pace, 1993). Each of these processes is mediated directly through bacterial metabolism; however, the ecological and evolutionary constraints that influence the metabolic capabilities of communities remains unknown. Knowledge of what controls bacterial community metabolic potential is crucial for developing approaches to manage and manipulate these processes in order to support beneficial processes while decreasing undesirable or harmful ones.

In my dissertation, I examine the evolutionary and ecological context of a subset of twelve bacterial metabolic pathways related to energy generation and metabolic precursor production that are crucial for bacterial growth. These twelve pathways include four primary sugar pathways of central C metabolism (glycolysis, gluconeogenesis, pentose phosphate (PP), Entner-Doudoroff (ED)) and eight fermentation reactions (lactate fermentation, formate fermentation, ethanol fermentation, acetate fermentation, acetone fermentation, butyrate fermentation, butanol fermentation, and 2,3 butanediol fermentation (BDOH)). These pathways were chosen because they are well studied, consistently annotated in automatically annotated genomes, and serve as a primary structure providing key components for bacterial growth and survival.

Specifically, I address the following questions:

1) *Are energy metabolic pathways conserved across a phylogenetically diverse set of microoorganisms?*

2) *Do related organisms respond similarly to differences in plant C and N inputs?*

3) *Are energy metabolic pathways or phylogenetic relatedness informative in regards to bacterial growth rates across a wide range of C substrates?*

I used a phylogenetically diverse dataset containing all 8178 publicly available genomes from the PubSEED to ask whether, at a broad scale, energy metabolic pathways are conserved. While there were strong constraints on the combination of these 12 pathways with only 265 (or 6.5%) of 4096 possible combinations of pathways observed in this dataset, phylogenetic dispersion patterns for each pathway, as well as combinations of pathways considered in concert, were apparently Brownian. Taken together, these strong constraints on combinations of pathways with only a weak phylogenetic signal suggest that, far from

being strongly conserved as generally believed, these pathways are under strong selection and organisms are flexible in which pathways they contain at any point in their evolutionary trajectory. Because this genomic dataset contains genomes from organisms cultured from a wide range of environments, I propose this may suggest that strong environmental selection is acting to rapidly shift pathway presence or absence towards global optima of different environments. If this were the case, organisms would respond to differences in environmental conditions in a phylogeny-independent manner.

To address this, I examined whether phylogenetically similar taxa responded similarly to differences in environment caused by differences in the C and N inputs from different plant species. This portion of my work took advantage of a long-term field experiment where plant species identity was manipulated and maintained over 12 years prior to my sampling. I collected soil samples from beneath each of 16 plant species and compared taxa response using 16S amplicon libraries. While I found a response of individual taxa and of overall community composition at the OTU_{98} level to differences in plant resources, these patterns weakened at higher taxonomic levels, with no pattern evident at the phyla level. This lack of response at higher taxonomic levels indicates an inconsistent response of phylogenetically similar soil bacteria to resource variation and is in agreement with organisms responding to environment in a phylogeny-independent manner.

While the previous study examined the response of phylogenetically similar organisms to resources measured at a scale typical of ecosystem-scale experiments, it is also important to understand how bacteria respond to resources at a finer scale. In order to do this, I cultured 56 bacterial isolates from the plots studied in the previous chapter. I then measured the linear growth rates of these isolates on 72 carbon substrates. In this study, I examined the relative relationships between metabolic similarity, phylogenetic similarity, and phenotypic similarity in terms of linear growth across all substrates examined. Across isolates I found that the number of substrates on which an isolate could grow was the best predictor of both its ability to use substrates and its distribution of linear growth rates across substrates. Both metabolic similarity and phylogenetic similarity were related to these substrate utilization profiles and growth rate profiles. Metabolic similarity and phylogenetic similarity each explained about 30% of the observed variations in growth rate profiles, but when considered together could explain 45% of variation. These results suggest that within a single environment, phylogeny is a stronger predictor of metabolism than would be expected based on the weak phylogenetic signal found in the first chapter.

Taken together, these findings support the hypothesis that strong environmental selection can lead to convergence towards a small number of combinations of metabolic pathways within a given environment. Although statistical power was low with only 3-4 isolates per plot, examination of the effects of inoculum source on the distribution of linear growth rates of the 56 isolates examined in chapter 3, I found a marginally significant ($p=0.06$) effect of plot. This further supports the idea that environmental selection on energy metabolism is strong. Soil bacteria are apparently flexible in terms of their energy

metabolism and even local level (within ecosystem) variation in resources may result in shifts in energy metabolism. Ultimately, both phylogeny and energy metabolism are important in determining bacterial growth; however, environmental selection may lead to convergence towards a small number of ecotypes within a system despite high levels of phylogenetic diversity. The strength and consequences of such environmental optimization of metabolism warrant further study.

Contribution Statement: While I am the lead author of all chapters presented here, each chapter was a collaborative effort. The first chapter focuses on my contribution to a much larger body of work conducted by an interdisciplinary team from which I draw much of my data. J. Edirisinghe pulled together over 8,000 annotated genomes, created a model template, made corrections where necessary, and added reactions related to the electron transport chain. He is also responsible for the classification of ETC types and, with R. Overbeek, identification of pathway presence/absence. My work took advantage of a tree generated by F. Xia using the DOE's KBase. N. Conrad assisted with figure design for the phylogenetic trees presented in this chapter. The questions in this chapter were generated primarily by me, but with discussion and contributions from C. Henry and J. Edirisinghe. In chapter 2, I took advantage of an on-going experiment with measurements of the environmental variables that I have included in my analyses. Although I chose which variables to consider, P. Reich and S. Hobbie designed the experiment and provided resource and process measurement data. The main questions addressed in this chapter were developed in discussion with S. Hobbie. The third chapter presents the bulk of my experimental efforts. I was solely responsible for the culturing and isolation of these organisms. Through the generous support of C. Henry and S. Hobbie, I shared the work of measuring their growth under numerous conditions with C. Nguyen, though I was solely responsible for the design and oversight of this work. S. Hobbie, T. Schmidt, M. Travisano, and C. Henry all provided numerous helpful suggestions on the compilation and analyses presented in the third chapter, but as in all chapters the application of those suggestions was my own. For all chapters, I led the writing, analysis, and interpretation of results and am solely responsible for any mistakes therein.

Chapter 1 Does phylogeny impose constraints upon the microbial energy metabolism?

Introduction

Microorganisms are key mediators of biogeochemical processes and microbial metabolism is directly tied into ecologically important biochemical transformations. Indeed the evolution of oxygenic photosynthesis in cyanobacteria is believed to be responsible for the current oxygenic atmosphere on the planet (Kopp et al., 2005). Additionally, microbes are responsible for everything from the bulk of nitrogen fixation and denitrification reactions (Firestone and Davidson, 1989, Gubry-Rangin, 2010) to cave formation through biological weathering (Northup and Lavoie, 2001) to processing nearly 90% of primary productivity (Cebrian, 1999, Cyr and Pace, 1993). It is not surprising; therefore, that inclusion of microbial parameters into global scale biogeochemical models can increase their accuracy (e.g. Xu et al., 2014). However, identifying meaningful parameters requires the identification of constraints on microbial metabolism and this remains a challenge in microbial ecology.

Microorganisms are extremely taxonomically and metabolically diverse. Unlike eukaryotes that are limited to a small number of energy generating reactions, microorganisms are able to derive energy through a diverse set of pathways. These diverse energy production pathways define much of microbial behavior and greatly impact biomass yields and metabolite production. For this reason, cellular energy generation is a crucial aspect of microbial metabolic modeling and depends upon environmental factors such as the availability of a carbon source, the fermentation capability of the organism, and presence of oxygen or other anaerobic electron acceptors as well as specific variations in the electron transport chain.

The importance of the metabolic potential of a microorganism, and its energy generation pathways in particular, has long been recognized. Before the widespread use of sequencing and taxonomic definitions based on the 16S gene, microorganisms were classified and placed into functional groups based on which carbon substrates they could utilize and under which environmental conditions they could grow. This approach to microbial classification led to a direct relationship between classification and ecological function. Since the rise of sequencing, taxonomic classifications based on the evolution of the 16S gene have dominated, and it is unclear whether these taxonomic classifications are directly related to ecological function. There is much anecdotal evidence that taxonomic classifications are informative about ecological function and culturing techniques designed to enrich for certain taxonomic groups have been developed based on this (Atlas, 2005). There is some experimental evidence that phylogenetically similar organisms respond similarly to changes in resources (Livermore et al., 2014, Fierer et al., 2007). Given these lines of evidence and the known importance of energy production to microbial fitness, it is often assumed that energy-related metabolism is strongly conserved and therefore phylogeny can be used to infer the metabolic and ecological potential of unknown microbes (e.g. Langille et al., 2013).

However, there is also evidence to suggest that energy-related metabolism may not be conserved. For example, iron-reducing bacteria are not monophyletic and are distributed across the Bacteria (Lonergan et al., 1996). Evidence of the successful transfer of key metabolic genes among phylogenetically diverse microbes suggests not only that these energy-related metabolic pathways may not be conserved, but also that their evolution may not be entirely vertical (Friedrich, 2002). More recently, pan-genome studies have shown that genomic content may be highly variable between two organisms from within a single OTU_{98} (Tettelin et al., 2005). It is unclear whether these examples are typical and to what extent they impact the overall metabolic and ecological potential of microbes.

Recognizing the importance of energy metabolism to the predictive abilities of whole genome metabolic models, J. Edirisinghe and colleagues developed a template for improving automatic annotation of pathways related to energy metabolism. When considered in concert, these pathways are referred to as a core metabolic model (CMM). Here we take advantage of all 8178 publicly available CMMs, representing 1346 unique bacterial and archaeal OTU_{98}, to examine phylogenetic constraints on energy metabolism. We have subdivided these CMMs into two functionally distinct subcomponents, the metabolic chassis (MC) that includes pathways by which organisms generate energy and metabolic precursors from organic substrates and the electron transport chain (ETC). The MC consists of pathways of eleven substrate-level phosphorylation reactions and gluconeogenesis that are broadly phylogenetically distributed and well studied and thus have consistently reliable annotations (Fig. 2-1) By focusing on these fundamental reactions, CMMs more accurately and quantitatively represent ATP biosynthetic pathways and enable prediction of key metabolic precursor production in central metabolism than whole genome metabolic models (J. Edirisinghe, unpubl). Using the output of these CMMs, we test the following hypotheses:

1) Because energy metabolism is critical to microbial fitness, we hypothesize that *both MC and ETC types will be phylogenetically conserved.*

2) We further hypothesize that the *individual pathways that comprise the MC will be strongly phylogenetically conserved*.

We discuss our findings in the context of variance of MC types broadly as well as the extent of variation in MC within individual phyla.

Methods

CMM construction process

This study takes advantage of individual CMMs created from a curated CMM template (J. Edirisinghe, unpubl.) applied to each of 8178 SEED annotated genomes (Supp. Table 1). These CMMs represent 1346 unique operational taxonomic units (OTU_{98}). These CMMs were automatically generated from SEED annotations (Overbeek et al., 2005) using the model template pipeline in DOE's

KnowledgeBase (http://kbase.us; Fig. 2-2). Briefly, in the initial step of the pipeline, a subset of high-quality SEED annotated genomes are used to generate a gene protein-reaction (GPR) association network, a table that provides a one-to-one association between gene annotations and enzyme production, which is used to generate an initial draft metabolic model. The biochemical reactions and pathways used in MCs (glycolysis, gluconeogenesis, pentose phosphate (PP), Entner-Doudoroff (ED), lactate fermentation, formate fermentation, ethanol fermentation, acetate fermentation, acetone fermentation, butyrate fermentation, butanol fermentation, and 2,3 butanediol fermentation (BDOH)) were selected from a set of well studied, phylogenetically diverse model organisms including: *Escherichia coli*, *Bacillus subtilis, Pseudomonas aeroginosa, Clostridium acetobutalicum* and *Paracococcus denitrificans*. Additional reactions were added and assigned GPR associations as needed to represent observed variation in microbial ETCs across the phylogeny (Supp. Table 2). Although we made every effort for completeness, we recognize that some reducing reactions (iron, chromium, sulfur, and sulfur derivatives) that are important for the metabolism of some organisms are not included due to poor annotation consistency or propagation through the SEED genomes. Similarly, photosynthesis reactions could not be included for this reason although they undoubtedly play a large role in the metabolism of autotrophic microorganisms. Together, this set of selected pathways and electron transport reactions represents all MC pathways and ETC variations upon which the template CMM was built (Supp. Table 3). Use of this curated model template, manually checked for correct reaction directionality, avoidance of multiple reactions for the same enzymatic reaction, accurate fermentation pathways, and ETC variations, increases the accuracy of energy predictions in the CMMs (J. Edirisinghe, unpubl.). CMMs are grouped into 4 distinct classes: those including only an aerobic ETC (aerobic), those including only an anaerobic ETC (anaerobic), those with both (facultative), and those without any respiration pathway (fermentative). The majority of CMMs contain at least one respiration chain.

Data Analysis

We considered whether individual metabolic pathways and electron transport chain types were conserved using the entire dataset, but restricted consideration of conservation of metabolic chassis types to those types representing at least 100 genomes. Although calculations are robust to imbalances in trait distribution, MC types representing less than 100 genomes individually represent <1.2% of the entire sample and necessarily have low resolution so they were excluded. To quantify the strength of phylogenetic signal, we used the caper package in R (R Core Team, 2012; Orme et al., 2012) to calculate Fritz's *D* (Fritz and Purvis, 2010) against a 16S phylogeny obtained from F. Xia (unpubl.) in order to evaluate whether individual ETC types, MC types, or individual pathways comprising the MC were phylogenetically conserved. A 16S phylogeny was used as the sample data are broadly phylogenetically distributed across the Bacteria (and even include some Archaea) and because this gene is used to assign operational taxonomic unit identity and as a common basis for making predictions of an organism's metabolic

capability. Fritz's *D* compares observed phylogenetic patterns in a binary trait to expected patterns generated by each of two null evolutionary models: random dispersion (D=1) and Brownian motion (D=0). Random dispersion distributions are based on random permutations of tip labels such that the number of tips representing a metabolic group is constant, but position on the tree is varied. Traits that are randomly dispersed have no discernable phylogenetic signal. This pattern would more typically be expected for expressed phenotypic traits with a strong gene by environment interaction, but is possible for genomic traits if selection is highly variable through time or if horizontal gene transfer is prevalent relative to vertical inheritance. Brownian motion distributions are generated by converting binomial variables to continuous variables and allowing those values to vary along branches following a random walk, where small changes along the tree are more probable than large ones. These continuous variables are then re-converted to binomial trait values using a cut-off value adjusted such that the number of tips with the trait remains constant to control for original trait distribution. A distribution explained by a Brownian motion model shows a phylogenetic signal, but the strength of the signal is reflective of a trait under neutral evolution (equivalent to a Blomberg's K value of approximately one; Blomberg et al., 2003). We expect that changes in the MC can occur, but those changes frequently have a high fitness cost and therefore changes do not occur as often as would be expected under a Brownian model. Thus we consider an ETC type, MC type, or MC pathway conserved when there is evidence of stabilizing or negative selection; in other words when the phylogenetic signal is not adequately explained by a Brownian model and Fritz's *D* is less than 0 (Fig. 2-3). Fritz's *D* is more sensitive than similar metrics based on unique branch length and which control only for differences in representation across a phylogenetic tree, such as SES_{MPD} (Pearse et al, 2013).

Individual phyla varied widely in the number of genomes present in our dataset. For this reason, it was not possible to directly compare differences in MC type distribution among phyla. However, we generated a distance matrix from the Sokal and Michener simple matching coefficient (Gower and Legendre, 1986) and used the vegan package of R (Oksanen et al., 2012) to generate non-metric multidimensional scaling ordinations to visually examine variance in MC composition (multivariate consideration of the presence or absence of each pathway) at the phyla-level.

Results

Distribution of MC types

Only 265 (6.5%) of the theoretical maximum of 4096 combinations of MC types were found in the 8178 CMMs analyzed. While a sample size of 8178 is not an exhaustive sampling of the diversity of microorganisms and new MC types are likely to be found with additional sampling effort, it is important to note that 25 of the MC types found represented 5489 (67%) of CMMs (Fig. 2-4). There were 654 genomes that lacked all sugar pathways (8%) and 1535 (19%) genomes that contained all sugar pathways. There were 1584 (19%) genomes that contained no fermentation pathways and only 3 (<1%) that contained all

eight fermentation pathways. No genomes contained all 12 pathways. Only seven out of the 265 MC types observed were represented by at least 100 CMMs and these seven MC types contained 2421 (30%) genomes (Supp. Table 4). The majority of observed MC types (92%) contained at least one major sugar degradation pathway (e.g. glycolysis, gluconeogenesis, PP, or ED); however, the presence or absence of these pathways varied throughout the phylogenetic tree and often within a single phylum.

Phylogenetic signal of CMMs

Of the seven MC types represented by at least 100 CMMs, the distribution of six were adequately explained by the Brownian model (i.e. Fritz's *D* was not significantly different from 0), while a single MC type was more broadly distributed with a D value between Brownian and random (*D*=0.175), but not adequately explained by either model (Table 2-1). On the whole, MC type shows only moderate phylogenetic clumping suggesting that energy-related pathways may not be highly conserved and can differ among closely related taxa (a subset of three sugar degrading pathways is presented in Fig. 2-5).

This pattern is even clearer when the pathways comprising the MC are considered individually. The Brownian motion model adequately explained the distribution of six (ED, ethanol fermentation, lactate fermentation, butanol fermentation, acetone fermentation, and isopropanol fermentation) of the twelve pathways included in MC type, indicating variations in these pathways are consistent with the neutral theory of evolution (Kimura, 1968), while gluconeogenesis was only marginally more distributed than expected under the Brownian motion model (*D*= 0.049, p=0.071; Table 2-2). The remaining five pathways (glycolysis, PP, acetate fermentation, BDOH fermentation, and butyrate fermentation) all had *D* values greater than 0 (0.088, 0.067, 0.074, 0.066, and 0.213, respectively), but were not explained by either the Brownian motion or the random model. Gains and losses of these pathways are occurring more, not less, often than would be expected under the neutral theory evolution. However, this effect is not large as these pathways have a Brownian-like distribution (*D* values are near zero).

ETC types were also not strongly conserved as they were either not significantly different from Brownian or were not adequately described by either model with a D value between 0 and 1 (Table 2-3). Fermentative CMMs (i.e. those lacking any respiration pathways) were adequately described by a Brownian model whereas facultative CMMs were marginally more distributed than expected under this model and both aerobic and anaerobic CMMs were more broadly distributed, but still Brownian-like (D= 0.099 and 0.21, respectively; Fig. 2-6).

Phyla-level variability in MC types

Individual phyla, however, show qualitative variation in the distribution of MC types (Fig. 2-7A-D). While all phyla had variation in MC types, some phyla showed less variation than others with similar representation in the dataset. For example, the Tenericutes (n=149) had 7 MC types while the Spirochaetes (n=202) had 17 MC types. All phyla that were highly represented in this dataset, such as the

Gammaproteobacteria (n=2450) and the Firmicutes (n=2312), showed a great deal of variation in MC types and it is possible that with increased sampling effort phyla with lower representation currently, such as the Tenericutes and Spirochaetes, would have broader distributions of MC types as well.

Discussion

Overall, our findings suggest there are strong constraints on the structure of the MC with only 265 (or 6.5%) of 4096 possible combinations of pathways observed in this phylogenetically diverse set of 8178 genomes. Even fewer of these combinations were highly abundant, with only seven combinations represented by at least 100 genomes, while the majority of observed MC types had very few representatives. Despite these constraints, however, each of the components of CMMs examined here (MC type, individual metabolic pathways, and ETC type) had phylogenetic dispersion patterns that were either Brownian or Brownian-like suggesting that they may have evolved following the neutral theory of evolution. While some phyla seemed qualitatively to be more constrained in MC types than others, it isn't possible to distinguish this apparent pattern from a sampling bias as those phyla that appeared to be more conserved also had fewer representatives in the dataset. Despite the critical role these pathways play in energy generation, and thus microbial fitness, these components of CMMs do not show a strong phylogenetic signal.

The weak phylogenetic signal paired with the skewed distribution of MC types that we observed may, however, be the result of strong environmental selection on energy metabolism in the absence of a strong phylogenetic signal of environment. Because this dataset considers organisms from multiple environments simultaneously, the common MC types examined in detail here may be evidence of strong selection towards the respective global fitness optima for organisms from different environments. Mid-abundance MC types may represent either local optima of common environments or global optima for less common environments, while the low abundance MC types may represent local optima for environments with low representation or short-lived non-optimal combinations of pathways. It is important to remember that our database consists primarily of genome sequences from cultured organisms which represents only a fraction of all known organisms. Thus, a methodological bias in culturing could have resulted in the over or under representation of particular MC types. If culturing conditions have led to such filtering of MC types, that bias itself may be a result of, rather than contradictory to, strong environmental selection on energy metabolism. If indeed that is the case, one would expect that organisms cultured under a single set of conditions would show remarkable consistency in MC types as would organisms cultured under identical conditions for prolonged time periods regardless of their degree of relatedness in either case.

While the lack of a strong phylogenetic signal in energy metabolism may not be intuitive given its potential effects on fitness, it may be attributable to the metabolic context of these energy pathways. Metabolic pathways do not act in isolation but as part of an integrated network where the products from

one reaction are the substrate for one or more other reactions. Any network-based redundancy in potential energy generation pathways may decrease the effects of negative mutations while still allowing a fitness benefit from positive mutations. This would allow greater flexibility in individual pathway acquisition, loss, and modification which could lead to Brownian or near-Brownian distributions of individual pathways as well as pathway combinations. For example, an organism that contains two fermentation pathways growing in a natural environment containing mixed substrates will not be as adversely impacted by a mutation in either of those pathways as an organism reliant upon a single fermentation pathway experiencing a mutation; therefore, despite the importance of maintaining energy production at the organismal level, the failure of a single pathway to generate energy may not have a high fitness cost as long as energy can be generated at the organismal level. Although not all CMMs contained multiple pathways, the majority did, including all of those representing at least 100 isolates, suggesting there may be selection to incorporate redundancy in energy generation pathways because of its importance to maintaining energy production at the organismal level in the case of a failure of a single pathway to generate energy. Given the interrelated network structure at the scale of whole organism metabolism, metabolic pathways beyond the CMM may also be only weakly phylogenetically conserved. However, since the importance of individual metabolic pathways not related to energy metabolism to overall fitness may be lower, it is also possible that there has not been selection for redundancy which might result in a counterintuitive strengthening of the phylogenetic signal in metabolic pathways beyond the CMM.

The weak phylogenetic signal associated with Brownian-like dispersion patterns may lead to observable similarities among closely related taxa. A 16S phylogeny approximates the true evolutionary history of these organisms and even the best automated annotations are imperfect, but given the size of the dataset and the calculated values of Fritz's D, even with perfect datasets the conclusions are unlikely to be substantially different. Indeed, even though this signal was weaker than expected when considered across this broad phylogenetic dataset, the existence of a phylogenetic signal is consistent with anecdotal evidence of limited phylogenetic similarities. However, given the weakness of the phylogenetic signal in comparison to the strong constraints we observed, caution is advised when using phylogeny to predict the metabolic potential and ecological functions of microorganisms, especially across environmental conditions. These traits are not strongly conserved and are more variable than commonly believed. Although it remains unclear what mechanisms underlie the observed constraints on distribution of MC types, it is clear that real constraints are acting to limit realized combinations of the pathways examined here. Focused evolutionary and physiological studies are needed to develop an understanding of the mechanisms underlying these constraints as well as the influence of these constraints on whole organism metabolism. Such mechanistic understanding would be useful to improve probabilistic annotation approaches to improve future automatic annotations and could yield great insights into the basic questions of how and when natural selection has shaped bacterial metabolism and its subsequent effects on patterns in ecosystem processes.

Chapter 2 Effects of differences in plant inputs on soil bacterial communities in a Minnesota prairie

Introduction

Primary productivity accounts for the largest input of resources to soil organisms, which are responsible for the recycling of up to 90% of plant biomass (Cebrian, 1999, Cyr and Pace, 1993). Through land use, humans are having a widespread effect on plant species productivity, community composition, and biodiversity, and are thereby influencing both the amount and chemistry of resources available to soil bacteria. Soil bacteria are key mediators of nutrient recycling; therefore, changes in their composition can feed back to affect the relative rates of key ecosystem processes such as nitrogen (N) mineralization and CO_2 efflux from the soil. Both resource quantity and chemistry can influence soil bacterial richness, diversity, and community composition and these aspects of bacterial communities may feed back to influence ecosystem processes. Increased understanding of the drivers of these microbial feedbacks is critical to improve current ecosystem modeling, but it is currently unknown which aspects of bacterial communities are important to advance this understanding (Wieder, et al., 2013, Treseder et al., 2012).

Alterations in both the quantity and chemistry of resource inputs can directly affect soil microbial communities. As described for plants by Tilman (1986), increases in nutrient supply can lead to decreases in species diversity as faster growing organisms that are unable to acquire limiting resources when those resources are in low concentration increase in dominance. Modeling work has shown a theoretical switch between resource quantity (e.g. total available energy) and nutrient limitation of bacterial growth at a C:N of 40 (Waring et al., 2013). Comparative studies have shown changes in bacterial community composition along nutrient gradients (e.g. de Vries et al., 2012, Allison et al., 2007) as well as under experimental conditions of altered resource supply (e.g. He et al, 2010, Ramirez et al., 2010, Castro et al., 2010). Similarly, additions of C substrates have been shown to alter bacterial community composition (Eilers et al., 2010, Fierer et al., 2007). These community composition shifts may even be observable at the phyla level, with phyla responding differentially to the C addition (Fierer et al., 2007).

Changes in diversity or community composition in response to altered resource inputs can feed back to affect ecosystem processes relating to resource availability (Xu et al., 2014, Treseder, 2008). Increases in biodiversity can lead to increases in ecosystem functioning and services (Hooper et al., 2005, Cardinale et al., 2011, Balvanera et al., 2006) with greater diversity required to support more diverse ecosystem functions (Isbell et al., 2011). Wagg and colleagues (2014) have shown that soil biodiversity loss, specifically, can decrease the rates of several ecosystem processes including rates of nutrient cycling and retention. Further, experiments directly manipulating culturable bacterial diversity showed a decelerating relationship between bacterial diversity and rates of community respiration (Bell et al., 2005).

Shifts in community composition may be as important as changes in biodiversity in influencing rates of ecosystem processes as taxa may have differential influence over ecosystem processes. Changes in taxa abundance may further affect species interactions thus spreading to affect the entire ecological network. These synergistic network-level effects have been related to changes in ecosystem processes, as well (Zhou et al., 2011).

Most experimental studies investigating the effects of nutrients on the biodiversity-ecosystem function relationship have focused on the relationship between plant species and primary productivity. This study examines the influence of altered plant C and N inputs caused by variation in the identity of plant species in monoculture on soil bacterial richness, diversity, and community composition. Specifically, we hypothesize:

1. Higher levels of C or N availability due to differences in plant species identity will result in lower bacterial diversity.
2. Higher levels of available resources (C or N) due to differences in plant species identity will induce shifts in bacterial community composition, with differential responses among phyla.
3. Higher levels of bacterial diversity will correlate with higher rates of soil C efflux and net N mineralization and shifts in bacterial community composition will be related to ecosystem process rates.

Methods

Soil sampling

This project takes advantage of an ongoing grassland global change field experiment that includes plots with wide variation in the stoichiometry of organic matter inputs to soils, the BioCON (Biodiversity, CO_2, and Nitrogen) experiment (Reich et al., 2001; Reich et al., 2004). The BioCON experiment was established at the Cedar Creek Ecosystem Science Reserve in 1997 and manipulates plant species richness and composition, atmospheric CO_2 levels, and N inputs, with CO_2 and N treatments beginning in 1998. Cedar Creek is characterized by sandy glacial outwash soils (Typic Udipsamments, Grigal, 1974) and net primary production at the site is limited by the supply of N (Tilman, 1994). For this experiment, only monoculture species plots at ambient CO_2 and N conditions were sampled. Across the plots, the C:N ratios of plant inputs to soil ranged from 8-79 aboveground and from 10-194 belowground across the 12 years prior to soil collection and from 17-61 and 14-105, respectively, in the year prior (Reich, unpubl). Variations among monoculture species treatments in C:N ratios of plant biomass have been mostly consistent since treatment initiation in 1998 (data not shown). Site preparation included removal of all plant species and application of methyl bromide to prevent growth from the seed bank. Following methyl bromide application, soils were inoculated with a common inoculum from nearby old fields (Chung et al.,

2007). Previous work in this system has shown variation in both microbial community C:N ratios (range: 6-22) and rates of C and N cycling (Dijkstra et al., 2006).

Soil cores were taken from each of two replicate monoculture plots maintained with each of 16 plant species (Andropogon gerardii, Aesclepias tuberosa, Amorpha canescens, Bouteloua gracilis, Schizachyrium scoparium, Sorghastrum nutans, Solidago rigida, Anemone cylindrica, Koeleria cristata, Lespedeza capitata, Lupinus perennis, Petalostemum villosum, Achillea millefolium, Agropyron repens, Bromus inermis, and Poa pratensis) in April 2010 at the start of the growing season and prior to that year's burn (plots are burned biannually). These 16 plant species belong to 4 plant functional groups (C3 grasses, C4 grasses, legumes, and forbs) which are known to vary in productivity and C:N (Reich et al., 2001b). Soils from replicate monoculture plots were homogenized into a single sample per plant species treatment. Soils were stored at -70C until prepared for DNA extraction. During DNA extraction, soils were kept refrigerated for up to 2 days during processing.

DNA extraction and sequencing

DNA was sent to the Institute for Genomics and Systems Biology's Next Generation Sequencing facility at Argonne National Lab where it was extracted following the protocol created for the Earth Microbiome Project (Caporaso et al., 2012). Paired end Illumina sequencing using the MiSeq platform was used to generate 253 bp reads of the V3 to V4 region of the 16S gene using the 515F and Golay barcoded 806R primers. These primers were created with an additional pad region to decrease primer-dimer formation. Sample complexity was increased using PhiX and barcoded samples were run on a single lane. Resulting reads were truncated at their first low quality base (A or B quality score). Reads without a perfect barcode match and reads of less than 75bp were removed prior to pairing reads. Paired reads of less than 253 bp were excluded from analysis.

Bacterial community composition, richness, and diversity calculations

Sequencing data were processed using the QIIME pipeline (Caporaso et al., 2010). Pre-processing of the data to pair reads and remove low quality paired reads was accomplished using scripts provided by Daniel Smith. These scripts are publicly available at: https://www.dropbox.com/s/hk33ovypzmev938/fastq-barcode.pl?dl=1. Sequences were clustered using UCLUST and operational taxonomic units (OTUs) were defined using a 98% identity cut-off. PyNAST was used to align sequences and create phylogenetic trees and the Ribosomal Database Project (RDP) classification was used for taxa identification. Shannon diversity indices were calculated automatically for each plant species treatment using QIIME. Rarefaction curves of clustered OTUs showed that Shannon's diversity had reached an asymptote before 35,000 sequences per sample, while OTU richness had just begun approaching an asymptote although differences in trajectories were evident (Fig. 1-1). Because richness had not fully reached an asymptote, analyses were conducted on both Shannon's diversity and OTU_{98} richness data.

Selection of individual taxa

Selection of individual taxa for examination of taxa-specific responses to resources targeted two groups of taxa: highly abundant taxa and highly variable taxa, those showing the greatest variability in abundance across plots. An abundance threshold of 5,000 16S sequences from a single OTU pooled across all plots identified 16 highly abundant taxa. For highly variable taxa, a standard deviation threshold set at 200 identified 11 highly variable taxa. Of these highly variable taxa, ten were present in all treatments while a single taxon was found in only two treatments and was therefore removed. Six of the remaining ten variable taxa were also identified as highly abundant, resulting in a selection of 20 unique taxa (10 abundant, 4 variable, 6 both) for examination of individual response to altered resources.

Resource availability and environmental data

Data on these plots was acquired from the publicly accessible Cedar Creek Long-term ecological research site data catalog (http://www.cbs.umn.edu/explore/field-stations/cedarcreek/research/data). Data collection methods are described in detail in Reich et al., 2001, Reich et al., 2001b, and <http://www.cbs.umn.edu/explore/field-stations/cedarcreek/research/data/methods?e141=>. Briefly, C and N measurements were taken annually on aboveground and belowground (0-20cm) biomass combined from each of two annual harvests. Aboveground biomass was sorted into live and litter fractions, while belowground biomass was sorted into: coarse, crown, and fine root fractions. KCl extractable N was measured annually and in situ net N mineralization was measured once during a month-long incubation in the middle of the growing season using closed-capped PVC tubes. Extra soil from N mineralization analysis is used to measure pH using the wet slurry method. Whole-plot plant C and N pools were calculated as whole-plant biomass C or N concentration multiplied by whole-plant biomass. Soil moisture (0-20cm) measurements were taken periodically throughout the growing season using TDR. Root ingrowth cores (to 20cm depth) are collected annually. Total soil C and N measurements were taken on finely ground soil every 5 years. Soil CO2 efflux measurements were taken with a LiCor 6400 from PVC rings installed in the plots. Change in soil C and N since experiment initiation (as absolute and percent measures) was calculated. In total we examined 35 plant and environmental characteristics to identify a subset which best described resource quantity and chemistry variation among these treatments.

Data analysis

Shannon's diversity values, an OTU_{98} richness table, and sequence count data were exported from QIIME. All samples except those beneath *Anemone cylindrica* had >35,000 high quality reads post-processing, while *A. cylindrica* had <1,000 reads; therefore data from soil beneath *A. cylindrica* were excluded from all analyses. It is, however, interesting to note that *A. cylindrica* had consistently low biomass throughout the duration of the experiment and likely did not support high bacterial biomass resulting in the low number of reads. For all remaining samples, calculations of Shannon's diversity and

OTU richness were averaged across 10 bootstrapped estimates based on a random subsampling of 35,000 sequences to control for unequal sequencing effort across samples. For community composition analyses, all sequence count data were included. Sequence counts for the 16 most abundant OTU were considered as a subset of data in examining the response of community composition to resource quantity and chemistry. Additionally, sequence counts were used to examine the direct response of individual taxa to resource quantity and chemistry.

All analyses examining the effects of resources used resource data collected in 2009 or the most recent year available for characteristics not measured annually. Non-metric multidimensional scaling (NMDS) ordinations were used to visually examine variation among plant species in this resource space (Fig. 1-2). Based on visual inspection of vectors which explained significant variation ($p<0.005$; $R>0.6$), a subset of four variables (aboveground litter, total plant N, root biomass, and belowground C:N) were chosen to examine community response to environmental variation. The chosen variables represent two indices of the total quantity of organic matter available to bacteria (aboveground and belowground biomass inputs to soil) and two indices of nitrogen availability to bacteria (the total pool of plant N in each treatment and the inverse of N in belowground biomass, expressed as the C:N ratio). Neither physical variable examined, pH or soil moisture, was significantly correlated with differences in resources among treatments ($p=0.175$ and 0.142, respectively), likely due to the small level of variation in these resources over this spatial scale.

Regression analysis was used to describe the relationship between Shannon's diversity, OTU richness, and individual taxa count data both with selected plant variables and with rates of ecosystem processes related to C and N cycling, specifically growing season cumulative soil CO_2 efflux, net nitrification, net ammonification, and net N mineralization. While these latter analyses were aimed at discerning effects of bacterial community composition and diversity on rates of C and N cycling, we recognize that such an interpretation is challenging based on observational data alone. First, total organic C and N inputs could be correlated with both bacterial community metrics and with C and N processing rates, leading to incidental (non-causal) relationships between bacterial communities and process rates. To control for this, we expressed N cycling rates (ammonification, nitrification, net N mineralization) relative to the total plant N pools and growing season soil CO_2 efflux relative to total plant biomass. Second, cause and effect cannot be inferred from regression analyses. However, it is more likely that soil CO_2 efflux was influenced by bacterial community characteristics than vice versa. Rates of N cycling are more ambiguous – relationships could be caused by the influence of bacteral community composition on these rates, or by the influence of inorganic N supply on bacterial communities. Keeping those caveats in mind, to evaluate the influence of plant resources on bacterial community composition and composition of the most abundant OTUs, resource vectors were fit to NMDS ordinations of community composition. NMDS ordinations were constructed from Bray-Curtis dissimilarity matrices after Wisconsin double standardization (Minchin,

1987). These analsyses were done at the OTU_{98}, family, and phyla levels. Similarly, ecosystem process vectors were fit onto NMDS ordinations of bacterial community composition at the OTU_{98} level to examine the relationship between community composition and rates of ecosystem processes. All statistical comparisons were completed using R 2.15.1 (R Core Team, 2012) using the vegan package (Oksanen et al., 2012).

Results

Effects of resources on OTU_{98} richness and Shannon's diversity

Despite greater than 10-fold variation in total plot plant N pools and nearly 3-fold variation in plant C:N, there was no significant relationship between either OTU_{98} richness or Shannon's diversity and either measure of N availability. Further, there was no relationship between OTU_{98} richness or Shannon's diversity and resource quantity (aboveground or belowground biomass inputs). This may be due, in part, to low variation in both OTU_{98} richness and Shannon's diversity estimates across all treatments (5402-6383.4 and 9.7-10.4, respectively). This suggests an insensitivity of bacterial diversity to differences in C and N availability. Further, there were significant differences in OTU_{98} richness ($p=0.0038$) and marginally significant differences in Shannon's diversity ($p=0.099$) among plant functional groups, suggesting that low variation did not entirely preclude discovery of an effect over this wide range of resource conditions (data not shown).

Effects of resources on community composition at the OTU_{98}, family, and phylum level

Shifts in bacterial community composition at the OTU_{98} level were related to aspects of both resource quantity and chemistry, as shown by significant vector fits for belowground plant C:N ratio ($p<0.01$) and root biomass ($p=0.04$; Fig. 1-3). However, changes in root biomass were more strongly related to shifts in community composition than changes in belowground C:N (beta-coefficients of 0.37 and 0.25, respectively). Bacterial community composition was not significantly related to either aboveground litter or biomass N, suggesting that belowground inputs have a stronger effect at this scale. At the family level, bacterial community composition was related to the belowground plant C:N ratio ($p=0.04$), but not to root biomass ($p>0.05$). At the family level less variance could be explained than at the OTU_{98} level (52% vs. 39%, respectively; Fig. 1-4A). At the phyla level, community composition was not related to any resource quantity or chemistry variable (Fig. 1-4B).

Responses of specific taxa to variation in resources

Although overall OTU_{98} community compositional shifts were significantly related to resources, this pattern may have been driven by changes in a relatively small subset of taxa as many taxa did not respond to variation in resources and those individual taxa that did responded differentially. Of the 20 highly abundant or variable taxa examined, only six (30%) showed a significant response to at least one

measure of resources at $p<0.05$, while eleven (55%) showed a significant response at $p<0.1$ (Table 1-1). Only two taxa had a significant response to multiple resources at $p<0.1$ (Table 1-1). Not all taxa that responded showed a response to the same resource and not all responses were in the same direction. A difference in bacterial community composition driven by changes in a subset of taxa was observed in examining differences in the composition of the sixteen most abundant OTU across resources. Composition of these taxa was related to resource chemistry with the belowground C:N of plants explaining 51% of observed variation ($p=0.02$; Fig. 1-5). However, this relationship was primarily driven by changes in three taxa, with two OTU showing increases in abundance with increasing belowground C:N (both $p<0.01$) and a single OTU showing a negative relationship ($p=0.04$) with belowground C:N.

Relationship of OTU_{98} richness, Shannon's diversity, and community composition with ecosystem processes

Neither OTU_{98} richness nor Shannon's diversity were significantly related to any ecosystem process considered here. Despite this, ecosystem-level C and N cycling rates were both related to changes in community composition; however, these effects were only marginally significant ($p=0.07$). Shifts in community composition showed relationships with growing season CO_2 efflux rates (beta-coefficient -0.21) and net N mineralization rates (beta-coefficient=0.22) of similar strength, but in opposite directions (Fig.1-6). The relationship between community compositional changes and net N mineralization was driven primarily by a significant relationship with net nitrification ($p=0.027$), while there was no relationship with net ammonification rates ($p=0.9$; Fig.1-6). In all cases, however, community composition was a worse predictor of ecosystem process rate than was resource availability.

When process rates were considered in relation to resource levels, however, a different picture emerged with nitrification / total plant N ($p=0.0034$), net N mineralization / total plant N ($p=0.0043$), and growing season CO_2 efflux / total plant biomass ($p=0.0077$) all significantly related to community composition. This suggests that while resource availability may be a primary control on rates of these ecosystem processes, the composition of the bacterial community may additionally influence realized process rates. Composition of the most abundant taxa was not significantly related to any ecosystem process considered here; although abundance data for two of the twenty taxa examined individually were related to both nitrogen cycling processes ($p<0.05$).

Discussion

Although no measure of resource quantity or chemistry had a significant effect on OTU_{98} richness or Shannon's diversity, bacterial community composition was affected by both factors. A previous study in grasslands and agricultural systems similarly found no effect of differences in N on richness or diversity (Ramirez et al., 2010). While pH and soil moisture are important drivers of soil bacterial diversity at larger geographic scales (Fierer and Jackson, 2006), neither was related to OTU_{98} richness nor diversity at this site. The lack of effect of soil resources on bacterial diversity may suggest key differences between bacteria

and plants. Plants commonly exhibit decreases in diversity in response to the addition of a limiting resource, a pattern we did not observe in bacteria. It is unlikely that the lack of response by either measure of bacterial diversity was a result of either a high rate of dispersal among plots or a high background of dormant, and therefore unresponsive, bacteria as differences in OTU_{98} richness and Shannon's diversity could be found among different plant functional groups. Thus, although both C and N availability vary among plant functional groups, neither is the primary driver of OTU_{98} richness or Shannon's diversity.

Although there was no effect on bacterial diversity measures, variation in belowground inputs may influence OTU_{98} community composition. Belowground C:N had the largest absolute effect on OTU_{98} community composition; however, root biomass had a larger standardized effect. Together these results suggest that although primary productivity in this system is strongly N limited, the availability of C also has an important effect on members of the soil bacterial community. When considered individually, however, only 30-55% of taxa responded to considerable variation in each of the resources examined here ($p<0.05$ and <0.1, respectively). Of those taxa that did respond, not all taxa responded similarly suggesting that neither C nor N limitation at this site is a strong primary driver of community composition.

Further, examination of community composition at higher taxonomic levels showed that response to resources was inconsistent within phyla. This was demonstrated by a weakening of the effect of individual resources on community composition at higher taxonomic levels. This inconsistent response of taxa within a phylum suggests that shared evolutionary history alone may not be a good predictor of what is limiting to particular taxa at this site. While some studies (e.g. Fierer et al., 2007) have demonstrated consistent shifts in phyla abundance with addition of specific resources, those studies did not look at the effects on individual taxa to see if this response was caused by a strong response of a few taxa or a consistent response across many taxa. The results presented here suggest that the phyla signal was likely due to the former, and not the latter. It follows that tools such as PICRUSt, which make predictions of organismal characteristics based on a 16S phylogeny, should be used with caution (Langille et al., 2013).

Bacterial community composition was also related to ecosystem process rates; however, resource availability was a stronger control of process rates than OTU_{98} community composition. That resource availability was a stronger predictor is not a surprising result, as resources must be available in order for them to be cycled within an ecosystem. However, OTU_{98} community composition was also related to differences in C and N cycling rates when they were examined relative to the inputs of plant C or N, respectively. Functional differences among community members may be partially responsible for these effects on both absolute rates and rates of these processes relative to total plant C or N inputs. It is impossible, however, to rule out the possibility that factors not considered here (e.g. soil aggregation) may affect both community composition and ecosystem rates simultaneously. If this is the case, these factors may be responsible for the apparent relationship between community composition and ecosystem process rates. The observed relationship of community composition with relative rates of the broad biogeochemical

process of C cycling was weaker than that with the more narrow biogeochemical process of N cycling. This has been suggested elsewhere (Schimel and Schaeffer, 2012), although in this particular case it may be due to the measurement of CO_2 efflux including both soil and root CO_2.

There is some evidence, however, that key community members may influence rates directly. Abundances of two individual OTUs (10% of those examined) were correlated with process rates. One of these was in the Bradyrhizobiaceae, a bacterial family directly associated with the N cycling process, which had a negative relationship with net N mineralization and net nitrification relative to total plant N inputs. This is interesting to note as the abundance of rhizobia nodules are also known to decrease with N addition (e.g. Clayton et al., 2004). The other OTU was in the Spartobacteriaceae within the Verrumicrobia, a common phylum of soil bacteria. It showed a positive relationship with the relative rates of both N processes. Organisms in the Verrucomicrobia have remained a challenge to culture and thus little is known of their physiology or potential functional roles in the environment. Functional differentiation among individual OTUs may influence ecosystem rates, even if the effect of variation in their abundance on differences in OTU_{98} community composition is negligible (Hooper et al, 2005).

The lack of relationship between bacterial biodiversity and resource availability paired with the disparate responses of individual OTUs to changes in resource inputs suggest that bacterial co-occurrence may not arise from functional differences in the ability to acquire resources. This may be explained by the much lower amount of structural complexity and morphological differentiation in bacteria. In other systems, such as epiphytes on algae and in the human gut microbiome, bacterial communities have high functional similarity despite low phylogenetic similarity (Burke et al, 2011, Human Microbiome Project Consortium, 2012). On the other hand, our measures of resource availability and chemistry were coarse and likely did not characterize variation in resources that bacteria perceive (e.g., variation in the types of organic compounds in plant tissues) and that have been shown to correlate with soil processes (Meier and Bowman, 2008).

In order to understand the mechanisms allowing co-occurrence we must first begin to understand the physiological trade-offs that underlie meaningful functional differentiation of soil bacteria. These functional differences, in turn, may be drivers of ecosystem process rates (Hooper et al., 2005). Further, functional differences may better enable examination of the patterns seen here between community structure and relative rates of C and N cycling. Evolutionary history does not reveal functional differences in how taxa respond to changes in resources suggesting that this trait may be under environmental selection. Thus, other physiologic trade-offs and constraints may play a stronger role in identifying meaningful functional distinctions among soil bacteria.

Chapter 3 Growth characteristics of soil bacterial isolates classified by taxonomy, metabolism, and ecological strategy

Introduction

Soil bacteria and other decomposers are responsible for processing 90% of plant primary productivity (Cebrian, 1999, Cyr and Pace, 1993) and soil bacteria also mediate key steps of other biogeochemical cycles (e.g. Firestone and Davidson, 1989). Given their importance in ecosystem functioning, developing understanding of the role of bacterial community structure in influencing ecosystem processes has been a major motivation in microbial ecology (Treseder et al., 2012). Ideally, understanding the relationship between community structure and function could be directly linked to studies of individual isolate physiology as well as inter-isolate interactions. However, this has not been feasible for several reasons. First, soil bacteria are difficult to study because their study necessarily disrupts their growth environment. Second, it is technically challenging to measure resources on the scale(s) important to bacteria (e.g. Meier and Bowman, 2008). Third, soil bacteria exhibit extreme diversity and only a small fraction, estimated at 1-10%, are culturable using modern culturing approaches. Therefore, attempts to link patterns in community composition to ecosystem processes must necessarily rely upon generalizations about interactions and metabolic potential to affect biogeochemical processes of interest.

As microbial ecology seeks theoretical frameworks from which to advance (Jackson et al, 2007; Gudelj et al, 2010), multiple approaches to relate community structure to function have been presented. Studying metagenomics has been informative for understanding community-function relationships in the case of certain very specific biogeochemical processes that are mediated by a small number of organisms and for which biochemical reactions, and the genes encoding the enzymes mediating these reactions, are relatively well studied. For example, metagenomic studies of denitrification have been successful in elucidating relationships between microbial community structure and process rates because the process can be directly linked to a small number of genes encoding key enzymes (Wallenstein et al., 2006). For broader biogeochemical processes such as decomposition or carbon (C) cycling, that involve hundreds of enzymes and thousands of taxa, however, this approach must be replaced with one reliant on meaningful categorizations of the high diversity of bacteria in order to compensate for the inability to culture a majority of organisms involved and to better tease out the inherent metabolic complexity driving the process. It is clear that while no categorization framework is ideal under all circumstances, in regards to biogeochemical functioning, a best-suited framework is one in which categories are closely linked with the metabolic activity responsible for the function. Three such frameworks are currently emerging: phylogenetic based classification approaches (PC), classical generalist-specialist ecological strategist based approaches (ES), and a new approach, presented here, based on classification by metabolic chassis defined as the presence or absence of central metabolic and fermentation pathways (MC).

Phylogenetic classification (PC) has been used to generalize metabolic functions of related taxa at the OTU (e.g. Langille et al., 2013) or phylum (Fierer et al., 2007) levels. This framework depends heavily upon the 16S gene as a proxy of shared evolutionary history. This approach is based on the underlying assumption that organisms sharing a greater proportion of their evolutionary history also have more of their genes in common, including metabolic genes, and thus will also be more metabolically similar. There is a great deal of anecdotal and some empirical evidence to support this (e.g. Fierer et al., 2007). However, there are also reasons to question these assumptions as variation in genomic content within a single OTU may be quite high (Tettlin et al., 2005). Even within what would be considered a single OTU, a single base pair variation in the 16S has been linked to large-scale niche differentiation (Eren et al., 2013). The extent to which these observations are relevant to biogeochemical interpretations of taxon relatedness depends, in large part, on how relevant such genetic variation among organisms is to the metabolic potential of an organism, which remains unknown.

The ES framework grew out of classical ecological classifications of organisms along a continuum of generalist to specialist with respect to substrate utilization. This approach has a rich history in microbial ecology as it has been directly related to the similar continuum of copiotrophic (organisms that grow in nutrient dense environments) to oligotrophic (those that grow in nutrient poor environments) bacteria (Lauro et al., 2009, Vieira-Silva and Rocha, 2010). Historically, this classification has been applied either based directly on phenotypic characterization or is inferred from the environmental conditions under which an organism was cultured. Recent efforts, however, have aimed at identifying genomic signatures of these strategies with the aim of predicting phenotypic characteristics from genomic data alone (e.g. Livermore et al., 2014, Lingner et al., 2010, Vieira-Silva and Rocha, 2010, Lauro et al., 2009). Lauro and colleagues (2009) identified multiple genomic signatures of the copiotrophic-oligotrophic lifestyle, including relating increased genome size with increased copiotrophic tendencies. Vieira-Silva and Rocha (2010), meanwhile, directly related predicted maximal growth rate to other genomic signatures. These genomic signatures may then be applied to infer growth and metabolic characteristics of unculturable organisms or metagenomic samples from the environment.

MC categorization is based on grouping organisms by the presence or absence of twelve substrate-level phosphorylation pathways (glycolysis, gluconeogenesis, pentose phosphate, Entner-Doudoroff (ED), lactate fermentation, formate fermentation, ethanol fermentation, acetate fermentation, acetone fermentation, butyrate fermentation, butanol fermentation, and 2,3 butanediol fermentation (BDOH)). These pathways are of fundamental importance because energy production defines a large portion of metabolic behavior and has the greatest ability to predict both biomass and metabolite production yields. Historically, microorganisms have been classified based on which substrates they can utilize for growth and under which conditions they can grow, but such a classification is limited to culturable organisms. When applied to anaerobic organisms, metabolic functional groups have led to great insight in wetland

biogeochemistry. In non-flooded environments, however, where aerobic respiration dominates, MC is able to distinguish among functional groups of organisms because of its focus on variation in metabolic energy pathways rather than on electron transport chains. By examining the ability of organisms to process substrates for energy generation, MC attempts to generate functional metabolic classifications in a culture-independent manner.

To compare the utility of applying these categorization systems to predict growth rate and metabolic flexibility, soil isolate growth was measured under a variety of conditions to test the following specific hypotheses:

1. ES will be related positively to genome size, maximum growth rate (MGR), and average positive growth rate (APGR). In turn, these characteristics will be positively correlated with each other.
2. PC types will differ in genome size, metabolic flexibility, MGR, and APGR. MC types will also differ in these characteristics.
3. Isolates that have more similar metabolic pathway profiles will also be more similar in regards to metabolic flexibility and growth rate distribution. MC will be a better predictor of metabolic flexibility and growth rate across all substrates than PC.

Additionally, we report the relationship between individual metabolic pathways considered within the MC and metabolic flexibility, MGR, APGR. We also explore the relationships of these categorization approaches and average positive growth rates on specific substrate classes ($APGR_{CC}$). The classes considered here are: monosaccharides, sugar alcohols, other sugars, amino acids, acyclic carboxylic acids, other carboxylic acids, and other carbon compounds which could not be easily classified into one of these classes.

Methods

Culturing and selection of isolates

Soils were sampled in April 2010 prior to the start of the growing season. Soil cores were collected from each of two replicate monoculture plots from an ongoing grassland experiment initiated in 1996 and located at the Cedar Creek Long Term Experimental Research site in central Minnesota (Reich et al., 2001). Soils are sandy (Typic Udipsamments, Nymore Series), derived from glacial outwash. Site preparation included removal of all plant species and application of methyl bromide to prevent growth from the seed bank. Following methyl bromide application, soils were inoculated with a common inoculum from nearby old fields (Chung et al., 2007). Plot treatments were established in 1997. Plots sampled here were planted and maintained as monocultures (n=2) with each of 16 plant species (*Andropogon gerardii, Aesclepias tuberosa, Amorpha canescens, Bouteloua gracilis, Schizachyrium scoparium, Sorghastrum*

nutans, Solidago rigida, Anemone cylindrica, Koeleria cristata, Lespedeza capitata, Lupinus perennis, Petalostemum villosum, Achillea millefolium, Agropyron repens, Bromus inermis, and *Poa pratensis*). Soils from two replicate monoculture plots were homogenized into a single inoculum per plant species treatment. One gram of soil inoculum was dissolved in 100 mL of physiological saline (Zuberer, 1994). After thorough mixing, 100μL was sampled and spread onto a petri dish containing a one hundred-fold dilution of soil extract agar (Zuberer, 1994). Petri dishes were incubated at 20°C in the dark. From each plate, up to eight colonies were chosen from those colonies appearing on days 3, 4, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 21, and 23 for a total of 1302 cultures. Three serial isolation streaks were performed on each of these 1302 cultures. We did not control for potential effects of inoculum source as there were not significant differences in which substrates were utilized when isolate growth was considered individually ($p=0.98$); however there was a marginally significant effect of resources when averaged by inoculum prior to statistical testing ($p=0.063$, Mantel $R=0.3098$).

A BioTek microplate reader was used to take OD_{600} measurements for each culture every ten minutes for sixteen hours on a ten-fold dilution of liquid soil extract using three replicate wells. These initial estimates of growth rate (change in OD_{600} per unit time, whether linear or exponential) were used to select a subset of 288 cultures for 16S sequencing. Based on these data, 124 cultures were chosen to maximize variation in growth rates (fastest, slowest, and near median from each inoculum) with secondary consideration of phylogenetic diversity used to select which of the near median growth rate cultures would be sequenced.

Sequencing, genome assembly, and pathway definitions

DNA was extracted from the selected cultures using the MOBIO PowerSoil kit. Frozen DNA was sent to the Institute for Genomics and Systems Biology's Next Generation Sequencing facility at Argonne National Lab for sequencing. DNA sequencing was performed on barcoded samples using six lanes on the HiSeq bench-top sequencing platform providing ~50X coverage. Genomes were automatically assembled using the Kiki pipeline and automatic annotations were generated with the Rapid Annotation using Subsytems Technology based on the SEED Framework (Overbeek et al., 2005). Quality of the annotated genomes was evaluated based on the copy number of ribosomal proteins (0.9-1.2 copies) and annotation subsystem coverage of at least 40%. Fifty-one of the 124 cultures were pure and passed this genome quality analysis. Boolean rules were applied to these genomes to identify the presence or absence of individual pathways (J. Edirisinghe, unpubl), and thus to classify them according to the MC classification scheme (see below). Branching pathways containing alternate reactions were allowed as long as at least one branch was present and the pathway was complete. Any incomplete pathway is considered absent, even if some portion of it may be present and potentially active. Of the 12 pathways examined (glycolysis (Gly), gluconeogenesis (GNG), pentose phosphate (PP), Entner-Doudoroff (ED), lactate (Lac), formate (For), ethanol (Eth), acetate (Act), acetone (Acn), butyrate (Bty), butanol (BOH), and 2,3-butanediol (BDOH)), 4

were not present in any of these genomes (formate, butyrate, butanol, and 2,3-butanediol). The Ribosomal Database Project classifier was used to assign taxonomy to isolates based on the 16S sequence obtained by the genomic sequencing effort (Wang et al., 2007).

Phylogenetic tree creation

The 16S rDNA sequences were extracted from assembled genomes and MAFFT was used to generate both primary and secondary structural alignments (Katoh and Standley, 2013). PartitionFinder (Lanfear et al., 2012) was used to select the best-fitting model of nucleotide evolution from those supported by RAxML using the Akaike information criterion (AIC). The model selected was the generalized time reversible (GTR) model with branch-specific evolutionary rates following a gamma distribution with a proportion of invariant sites (GTR+I+G). We applied this model to generate maximum likelihood trees in RAxML 7.7.2 (Stamatakis, 2014). The ape package in R was used to convert these trees into phylogenetic distance matrices (Paradis et al., 2004). Statistical comparisons were performed using both of these distance matrices; however, there were no statistical or qualitative differences so only analyses using the distance matrix generated from the secondary structural alignments are described.

Growth experiments and growth rate calculations

Isolates were grown on Biolog Gen III media plates. These plates contain 94 culture conditions, but only growth on the 72 carbon substrates is examined here. Biolog plates were inoculated by selection of a single colony grown on a hundred-fold dilution of soil extract agar and added to a nutrient solution containing PO_4, K, NH_4, Mg, Ca, and trace elements. Each well was inoculated with 250µL of this solution. OD_{600} was measured every thirty minutes for a period of 24 hours, and then every 6 hours between 24 and 48 hours of incubation time, with periodic shaking between measurements. During processing, Biolog plates were kept at room temperature in the dark. Because of the high-throughput nature of this approach, isolate growth was not balanced under all conditions; therefore, the growth rates presented here represent the real change in absorbance per unit time, but the time it took for absorbance to double should not be considered doubling time as both linear and exponential growth are represented in this data set.

Growth rate calculations were automated using a multistep process. First, a linear model was used to evaluate if change in OD_{600} over the entire interval was positive. If there was not a significant, positive slope the isolate was considered to have no growth on a substrate. For isolate-conditions with an overall positive slope, a linear model was used to calculate growth along the entire curve using a sliding window of fifteen OD_{600} measurements (modified from Carpenter). For curves with multiple significant positive segments, growth rate for inclusion was decided using the hierarchical consideration of the best R^2 value, followed by maximum observed slope if there were multiple segments with the same R^2. If there were no instances where individual segments had a fit above the R^2 cut-off value, the isolate was considered to have no growth on this substrate regardless of whether there was a significant positive slope overall. All analyses

of linear growth rate distributions were repeated with datasets using each of four R^2 cut-off values (0.60, 0.80, 0.90, and 0.95), but comparisons of these datasets did not vary in either statistical significance or qualitative results, so only results from the most inclusive dataset are presented here.

Category definitions

Metabolic flexibility was defined as the number of substrates on which an isolate showed significant positive growth. While metabolic flexibility was used as a proxy for ES strategy along a continuum, ES was also used as a categorical variable so that it could be compared directly with MC and PC which are inherently categorical. The number of substrates on which an isolate could grow was used to determine its ES categorization. The cutoffs used were < 36 (<50%, specialist), 36-54 (50-75%, intermediate), and >54 (>75%, generalist). This categorization resulted in 26 specialists, 20 intermediate, and 5 generalists. MC categorizations were based on the presence or absence of each metabolic pathway. There were 12 unique MC types (Table 3-1), but only 4 of them had enough replication for statistical tests: A (containing pathways: GNG+PP+ED+Lac+Eth+Act), B (GNG+Gly+PP+Lac+Act), C (GNG+Gly+PP+ED), D (GNG+Gly+PP+ED+Lac) with 9, 7, 7, and 10 representatives, respectively. PC categorization was based on the Ribosomal Database Project classification at the family level (Wang et al., 2007). Overall, isolates were from 4 phyla and 9 families, with three families having multiple representatives (Burkholderiaceae, Micrococcaceae, and Xanthomonadaceae with 15, 10, and 21 isolates, respectively).

Data analysis

Linear models were used to test the hypothesis that ES was positively related to genome size, MGR, and APGR. Analyses of variance (ANOVAs) were used to examine differences in metabolic flexibility, MGR, APGR, $APGR_{CC}$, and genome size by PC and MC categories. Tests of relationships between individual pathway presence or absence and metabolic flexibility, MGR, and APGR were examined using Student's T-tests, while pathway relationships with $APGR_{CC}$ for individual compound classes were tested using Hotelling's (multivariate) t-test. A pairwise distance matrix for ES was generated using Euclidean distances of the actual number of substrates on which there was positive growth. A distance matrix for MC was calculated using the simple matching coefficient. A distance matrix for PC was calculated from pairwise branch lengths using the cophenetic function in the ape package of R. The growth rate distribution response matrix was generated using the Euclidean distance metric while the substrate utilization response matrix used the simple matching coefficient. Mantel tests were used to evaluate the correlations between each categorization type (PC, ES, and MC) and both substrate utilization and growth rate distributions. Additionally, Mantel tests were used to evaluate the degree of covariance among the different categorization approaches. All statistical analyses were performed in R, using the ape and vegan packages (R Core Team, 2012, Paradis et al., 2004, Oksanen et al., 2012).

Results

ES was related to growth rate, as organisms with greater metabolic flexibility also grew more quickly, both in terms of MGR and APGR ($p<0.0001$; $R=0.3$ and 0.36, respectively; Fig 3-1A, B). Additionally, ES type was related to MGR, with generalists and intermediates having a higher maximum growth rate than specialists ($p<0.0001$). ES type was also related to APGR with generalists having higher APGR than intermediate, which had a higher APGR than specialists ($p<0.0001$). This difference in APGR was robust with ES types showing significant differences in average positive growth on each C compound class ($APGR_{CC}$; Table 3-2). In all cases, $APGR_{CC}$ was highest for generalists and lowest for specialists, with the intermediate organisms in between (Table 3-2). Genome size was not related to ES type, metabolic flexibility, MGR, or APGR ($p>0.05$, data not shown).

Metabolic chassis types captured some variation in growth characteristics. MC types varied in MGR ($p<0.001$, Fig 3-2A), APGR ($p<0.0001$, Fig 3-2B), metabolic flexibility ($p=0.03$, Fig 3-2C) and, interestingly given the lack of any direct relationships, in genome size ($p<0.0001$, Fig 3-2D). This variation in growth characteristics was again robust as MC types also differed in $APGR_{CC}$ for every carbon class, with types A and B broadly tending to have faster $APGR_{CC}$ than C or D, regardless of which substrate class was examined (Table 3-3). MC type A lacks glycolysis but contains both ethanol fermentation and ED, while B has glycolysis but lacks ED or ethanol fermentation. Both of the slower growing metabolic types contain both glycolysis and ED, but not ethanol fermentation.

Of the twelve pathways comprising the MC, only eight were present in this sample of soil isolates. Each of three key pathways were significantly related to differences in growth rate distributions across all substrates (glycolysis, ethanol fermentation, and ED with $p=0.001$, 0.001, and 0.002, respectively). Additionally, both acetone fermentation and acetate fermentation pathways were marginally significantly related to differences in growth rate distributions at $p<0.1$. Both glycolysis and ethanol fermentation were directly related to MGR, APGR, and metabolic flexibility. Glycolysis and ethanol fermentation pathways were each related to metabolic flexibility with glycolysis having a negative relationship and ethanol fermentation being positively related ($p<0.001$; Table 3-3). The relationships were the same with MGR ($p<0.01$ and $p=0.034$, respectively, Table 3-4) and also with APGR ($p<0.01$ for both). ED was marginally related to APGR, but not the other growth characteristics ($p=0.066$; Table 3-4). Overall, organisms with a complete glycolysis pathway, and to a lesser extent ED, grew more slowly while those with ethanol fermentation were faster (Fig 3-4). This pattern held regardless of substrate class; however, it was driven by a subset of substrates within each substrate class as there were examples of substrates lacking a clear distinction within each class. There were no examples of substrates, however, on which organisms containing glycolysis or lacking ethanol fermentation performed better than those with ethanol fermentation.

The third categorization approach, PC, was also able to capture variation in these growth characteristics. Metabolic flexibility, MGR, APGR, and, again, genome size all varied across families ($p<0.001$ for all, Fig 3-4). Burkholderiaceae grew on more substrates and at a higher maximum and average rate, while also having the largest genome. Again, this pattern was robust across substrate classes. Families were significantly different in $APGR_{CC}$ across all C classes, with Xanthomonadaceae growing slowest on average across all C classes and Burkholderiaceae having the highest $APGR_{CC}$ on four of seven classes (Table 3-5).

Pairwise phylogenetic distance was similarly related to differences among isolates in substrate utilization profiles ($p=0.04$, $R=0.38$) and growth rate distributions on those substrates ($p=0.001$, $R=0.33$), as was pairwise metabolic distance ($p=0.001$ for both, and $R=0.44$ and 0.32, respectively). Phylogenetic distance and metabolic distance matrices co-varied with about 55% overlap ($p=0.001$). Pairwise ES distances however, were substantially more related to variation in both substrate utilization profiles and growth rate distributions on those substrates ($p=0.001$ for both, and $R=0.70$ and 0.41, respectively).

The covariance of pairwise ES distances with phylogenetic and metabolic distances was lower at 27% for phylogenetic distance and 22% for metabolic distance ($p=0.001$ for both). After controlling for the phylogenetic effect, metabolic distance remained informative with a significant relationship to differences in substrate utilization profiles and growth rate distributions ($p=0.001$, $R=0.24$, 0.30, respectively); although differences in pairwise ES distances were still considerably more related to either ($p=0.001$, $R=0.68$, 0.36, respectively). Although there was overlap among the three categorization approaches, there was not a one-to-one correlation between any pairwise comparison of MC type, ES type, or PC type, suggesting that each approach provides information not captured by the others.

Discussion

As predicted, the number of substrates on which an isolate was able to grow was positively correlated with both MGR and APGR. This suggests that not only are generalists able to grow on more kinds of substrates, they may also be able to take up substrates more quickly allowing for faster growth (Lauro et al., 2009). While Vieira-Silva and Rocha (2010) saw a relationship between indicators of ecological strategy and predicted maximal growth rate, the empirical data here support extension of that to include metabolic flexibility and APGR. The relationships among metabolic flexibility, MGR, and APGR are robust across all three categorization types and all substrate classes. Because the strength of these relationships were evident regardless of the categorization scheme, it may be reasonable to use genomic signatures to infer not only maximal growth rates, as has been proposed, but also to predict the typical growth rate of an organism. It must be remembered, however, that this rate should be considered relative to other organisms with known growth rates and also that this is only a generalization of typical growth and may not hold for any individual substrate when considered individually.

Surprisingly, we did not observe a direct relationship between genome size and MGR or APGR. Other studies have suggested such a relationship exists and it has been proposed that more specialized taxa may have evolved decreased genome size through reduction in the number of uptake proteins as a response to high risk of attack by bacteriophages in natural environments (Livermore et al., 2014). While MC types C and D (both containing Gly+ED, but not Eth) as well as isolates from the Xanthomonadaceae (all types that had traits typical of a specialist, e.g. grew on fewer substrates) did have smaller genomes, there was not a direct relationship between metabolic flexibility (or ES type) and genome size. This suggests that within soil bacteria, while variation in genome size may be apparently correlated with substrate specialization, there is not a direct relationship between genome size and growth rate. Any apparent relationship is likely a by-product of other strategy-related constraints, rather than a direct relationship. Indeed, some isolates with a large genome were only able to grow on an intermediate number of substrates while a subset of isolates with small genomes could grow as generalists. Therefore we advise careful use of genome size as a genomic signature of ecological strategy or growth rates.

Each of the categorization approaches studied here were able to capture variation in the observed growth characteristics of the soil isolates studied. The ES categorization approach showed the tightest correlations with both substrate utilization profiles and growth rate distributions across the substrates in this study. It is important to note that the ES categories used here were defined based on metabolic flexibility from this same set of substrates. Therefore, this increased ability to explain variation in growth characteristics may not only be expected but may also be, at least partially, an artifact of the way ES categories were defined here. In spite of that consideration, the strong relationship between metabolic flexibility and growth means it is likely that ES is a good classifier for MGR, APGR, and substrate utilization profiles. The ability of ES categorizations based on genomic signatures, rather than phenotypic measurements, to accurately capture variation in growth characteristics depends upon the reliability of the signatures chosen. Likewise, the ability of ES categorizations to explain growth characteristics relative to PC or MC is also likely to be dependent on the genomic signatures chosen.

Pairwise phylogenetic distances between isolates were related to differences in both substrate utilization profiles and growth rate distributions. For the substrates examined here, the assumption that isolates with a shared evolutionary history will have more similar substrate utilization profiles and growth rate distributions holds and phylogeny is a reasonable predictor of these characteristics. Pairwise distances in metabolic pathway presence or absence did not perform any better or worse than phylogenetic distance; although metabolic distance also performed worse than ES distance when ES was defined empirically. This was in spite of a mechanistic relationship between the energy pathways comprising MC and organism growth. However, the effects of the presence or absence of three individual pathways (glycolysis, Entner-Doudoroff, and fermentation to ethanol) with substrate utilization profiles and growth rate distributions was remarkably strong. Surprisingly, these relationships also remained across substrate classes. This suggests

that energy production pathways may have an effect beyond the metabolism of immediate compounds feeding into them. In this case the sum of that effect is apparently similar in strength to the effect of shared evolutionary history.

On average, isolates that lacked glycolysis or contained ethanol fermentation pathways grew more quickly and were able to utilize a wider variety of substrates than those containing glycolysis or lacking ethanol fermentation pathways. This counterintuitive result may be the product of a trade-off between growth rate and efficiency. While glycolysis yields more adenosine triphosphate per unit substrate than ethanol fermentation, it contains a greater number of reactions with smaller free energy differences and may therefore be slower. While this pattern was robust across substrate classes, there were some substrates within each class with no clear differences in growth between isolates having and lacking any particular pathway. Pathway assignments were binary (i.e. present, absent); therefore, they may be inaccurate for substrates that enter the pathway after the first (or several) reactions have been completed. This inaccuracy arises because the enzymes necessary for processing those substrates entering the pathway at a later reaction may still be present in isolates that were scored as lacking the pathway as a whole. This imprecision in pathway assignments may explain instances where, for individual compounds, no difference in growth rate was observed between isolates containing and those lacking a particular pathway.

Overall, ES, PC, and MC each provided useful information on substrate utilization profiles and growth rate distributions. While there was substantial covariance (ES-PC: 27%, ES-MC: 22%, and PC-MC: 55%), each categorization method provided meaningful information. Therefore it is recommended to use ES, PC, and MC categorizations in concert when possible, and to consider inclusion of ES whenever feasible. In those cases in which genomic signatures are to be applied in lieu of empirical data, the combined use of genomic signatures for each categorization type is advised. Caution should be taken, however, to ensure that genomic signatures for ES accurately capture metabolic flexibility and it should be remembered that any apparent relationship with genome size might be indirect. Although ES categories, as defined here, had the strongest relationship with growth characteristics, they are also the most difficult to apply as they require cultured organisms to be grown on various substrates. In instances where culturing isolates is not feasible, the combined use of PC and MC is still quite effective and is expected to explain around 45% of the variability in growth rate distributions. Predictions from the use of PC and/or MC may also be used to devise culturing conditions that would favor isolation of cultures of interest for further study.

Illustrations

Table 1-1. Measures of dispersion for all MC types across the phylogenetic tree was Brownian or near-Brownian. There was no evidence for strong phylogenetic conservation of MC type, i.e. D was never both significantly different than 0 and negative.

MC Type	D	p(Brownian)	Interpretation	# Containing
1	0.175	0.003	More distributed than expected under Brownian	256
2	-0.0906	0.881	Described adequately by Brownian model	225
3	-0.107	0.933	Described adequately by Brownian model	217
4	0.0895	0.108	Described adequately by Brownian model	253
5	0.0945	0.121	Described adequately by Brownian model	190
6	-0.124	0.97	Described adequately by Brownian model	262
7	-0.0439	0.856	Described adequately by Brownian model	1018

Table 1-2. Measures of dispersion for the twelve individual metabolic pathways comprising MC categorization across the phylogenetic tree was Brownian or near-Brownian. There was no evidence for strong phylogenetic conservation of any of these metabolic pathways, i.e. D was never both significantly different than 0 and negative.

Pathway	D	p(Brownian)	Interpretation	# Containing
GNG	0.0486	0.071	Marginally more distributed than expected under Brownian.	5288
Glycolysis	0.0883	0.003	More distributed than expected under null Brownian	4286
PP	0.0666	0.01	More distributed than expected under null Brownian	4010
ED	-0.0318	0.849	Described adequately by Brownian model	2257
Acetate	0.0742	0.006	More distributed than expected under null Brownian	4467
Ethanol	-0.00702	0.599	Described adequately by Brownian model	3354
Lactate	0.0218	0.232	Described adequately by Brownian model	2215
BDOH	0.0659	0.008	More distributed than expected under null Brownian	3319
Butyrate	0.213	0.022	More distributed than expected under null Brownian	76
Butanol	-0.0158	0.593	Described adequately by Brownian model	253
Acetone	-0.0553	0.755	Described adequately by Brownian model	199
Isopropanol	-0.0114	0.611	Described adequately by Brownian model	907

Table 1-3. Measures of dispersion for all ETC types across the phylogenetic tree was Brownian (fermentative) or near-Brownian (all others). There was no evidence for strong phylogenetic conservation of any ETC type considered here, i.e. D was never both significantly different than 0 and negative.

ETC type	D	p(Brownian)	Interpretation	# Containing
aerobic	0.099	0.001	More distributed than expected under null Brownian	1416
anaerobic	0.21	0.001	More distributed than expected under null Brownian	188
facultative	0.04	0.075	Marginally more distributed than expected under Brownian.	4104
fermentative	-0.01	0.604	Described adequately by Brownian model	1048

Figure 1-1. Schematic pathway map showing the major sugar processing pathways and fermentation pathways included in the metabolic chassis.

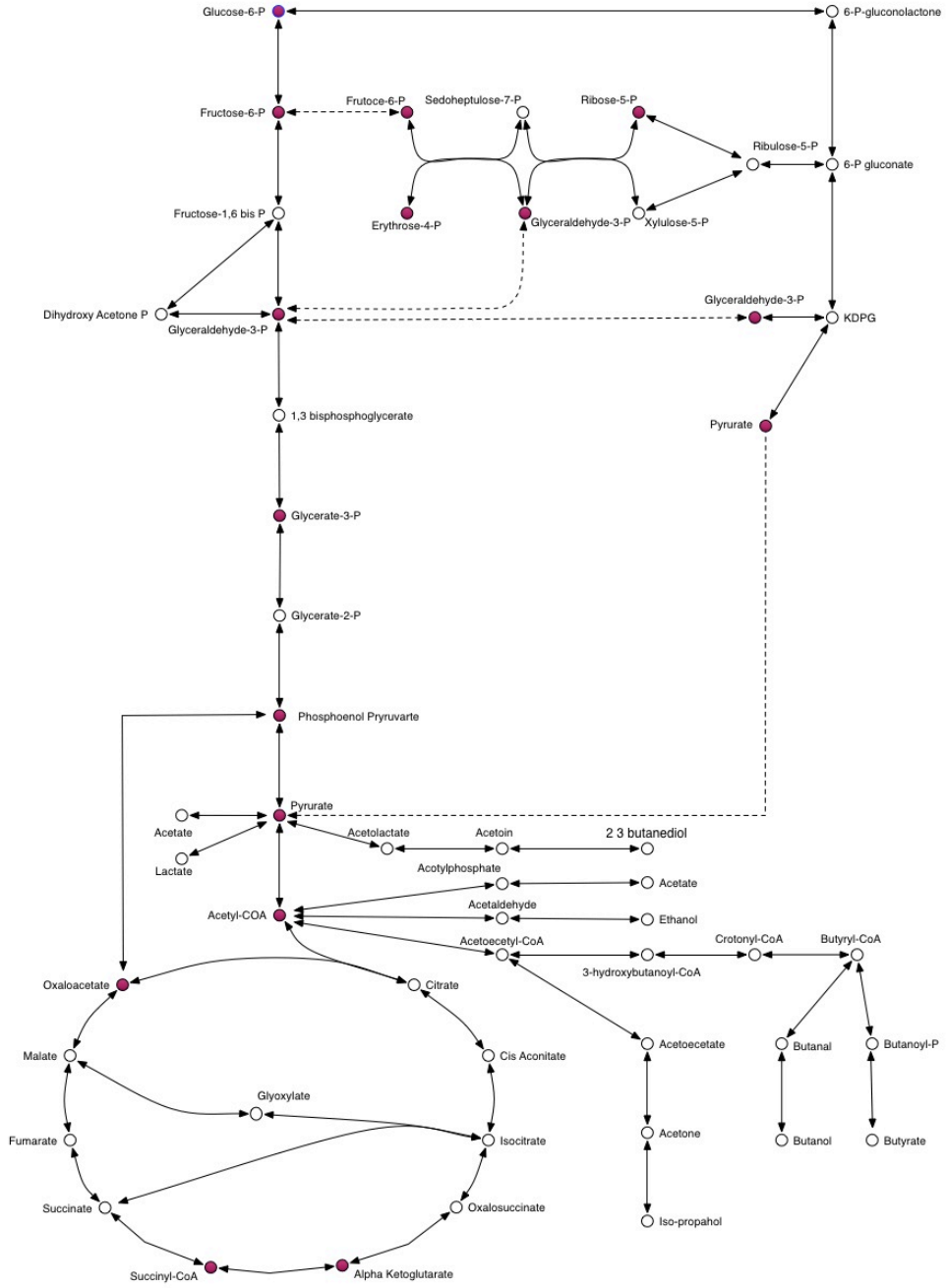


Figure 1-2. Flow chart of the CMM development pipeline.

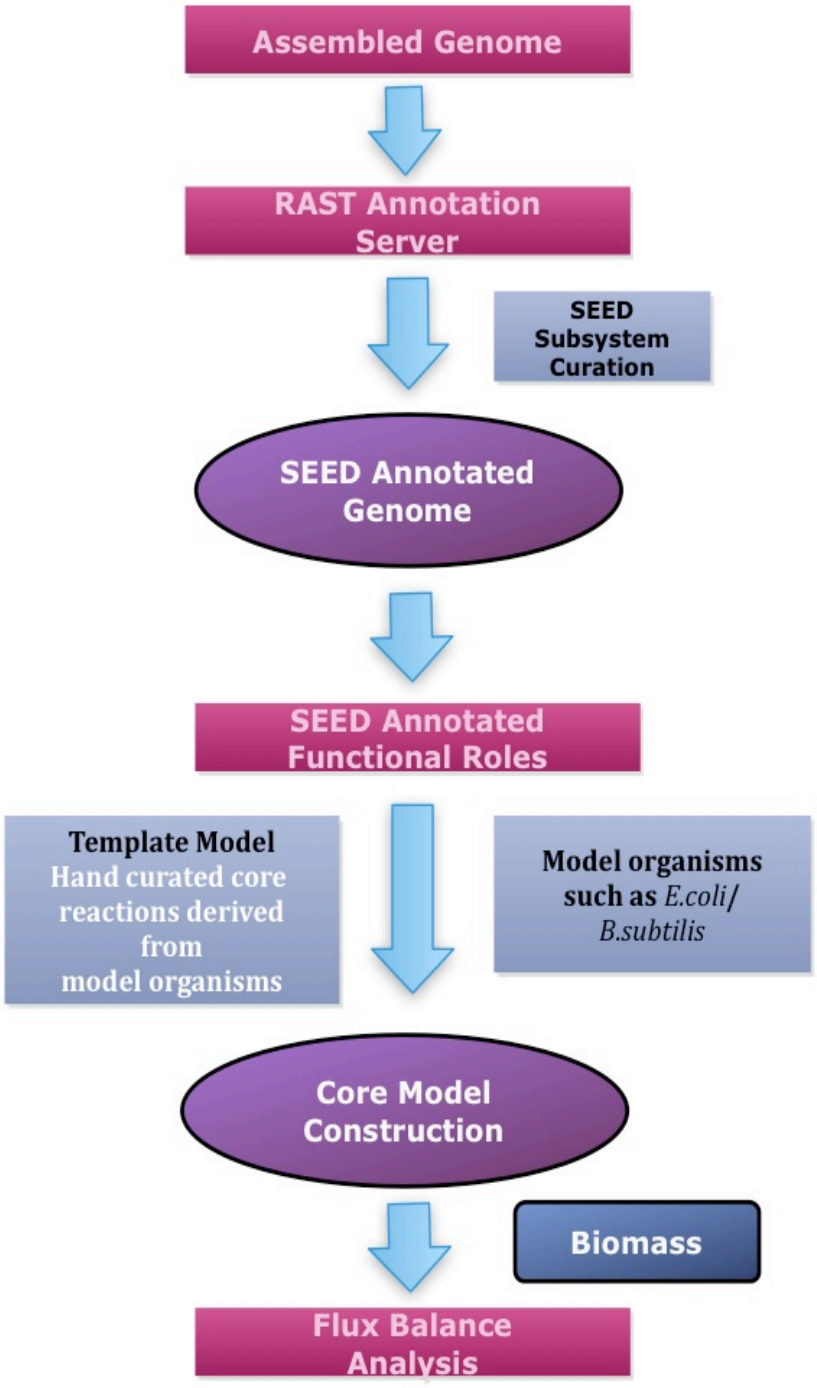


Figure 1-3. Schematic demonstrating variation in the degree of phylogenetic conservation of a hypothetical trait with states A and B. Fritz's D tests two null models: random (D=1) and Brownian (D=0). We expect changes in components of CMMs to be strongly conserved, having D-values that are both negative and significantly different from 0.

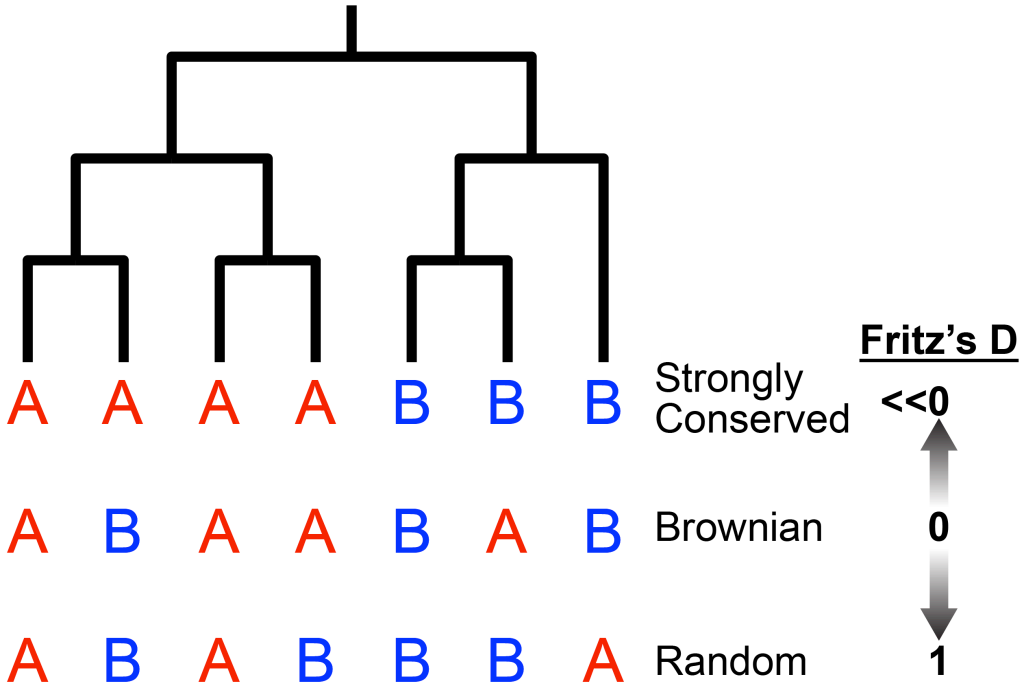


Figure 1-4. Frequency distribution of CMMs by MC type. The majority of CMMs were represented by a small proportion of observed MC types, and observed MC types represented only a small proportion (6.5%) of theoretical MC types.

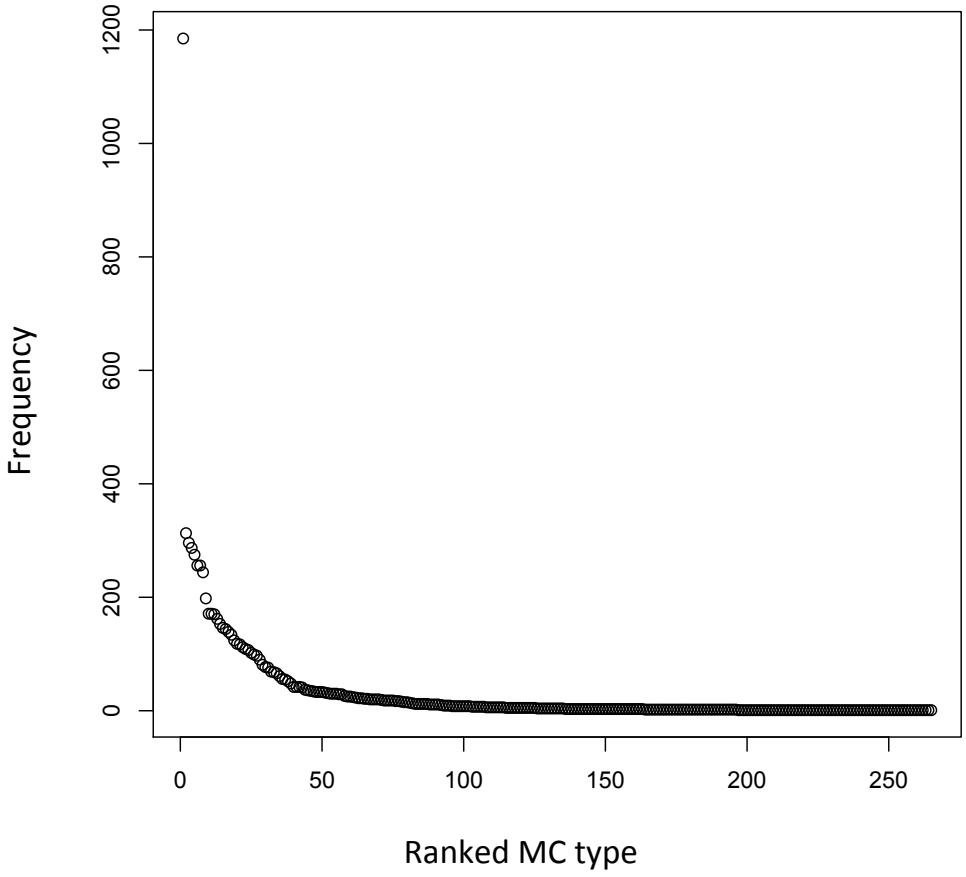


Figure 1-5. Distribution of three of the sugar degradation pathways (glycolysis, gluconeogenesis, and ED) on a 16S phylogeny. Glycolysis was Brownian-like, although it was more distributed than expected under a Brownian model. Gluconeogenesis was marginally more distributed than expected under a Brownian model. ED was not significantly different from a Brownian model.

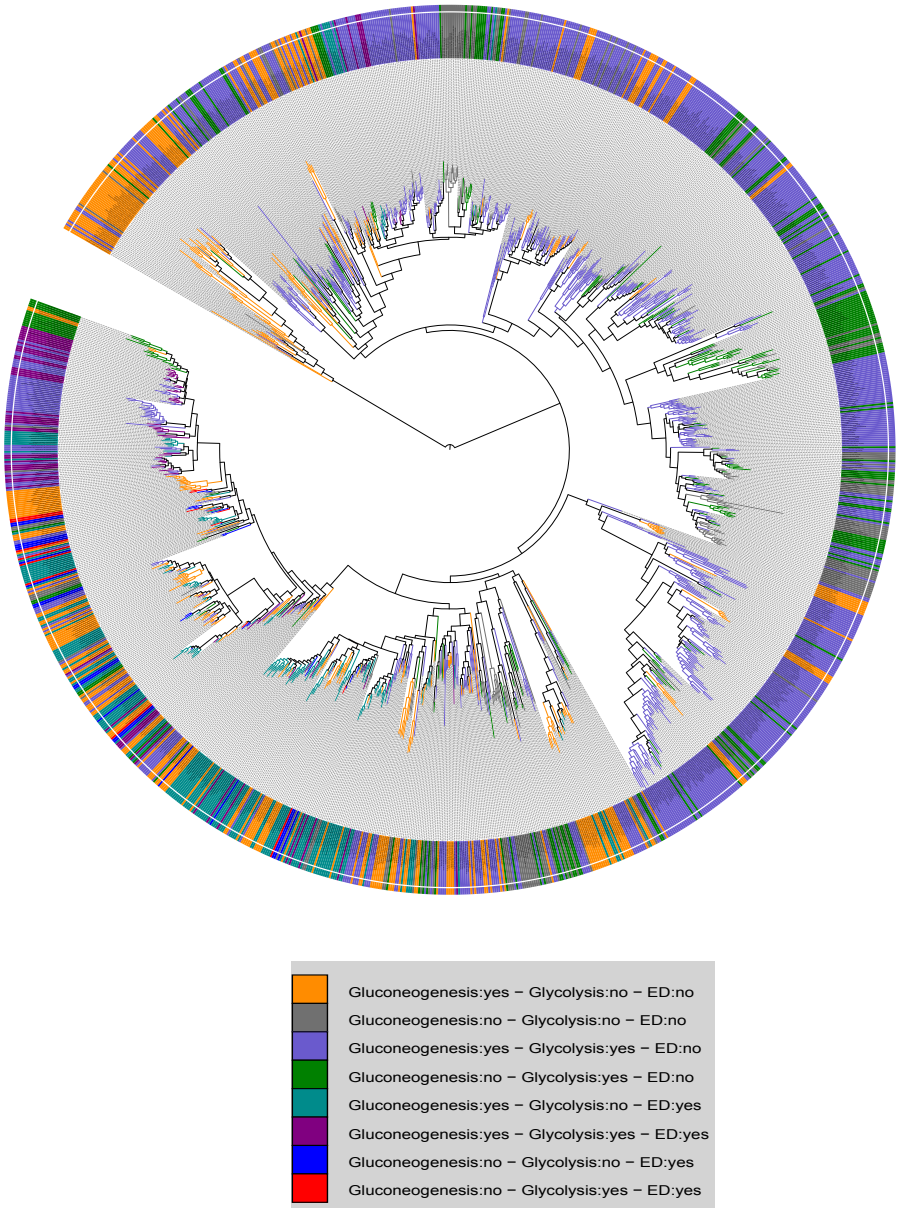


Figure 1-6. Distribution of ETC type on a 16S phylogeny. All ETC types were adequately described by the Brownian model of evolution. While there was a phylogenetic signal in ETC type, ETC type was not strongly conserved.

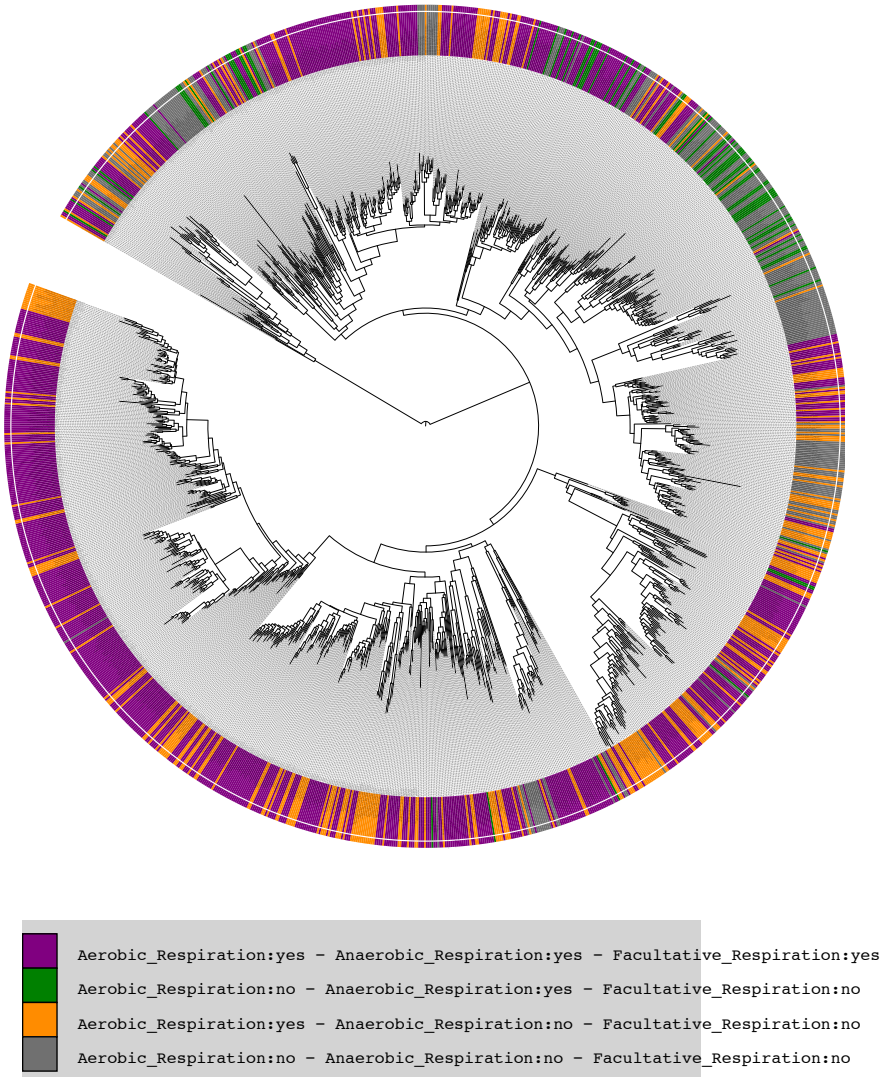


Figure 1-7. Nonmetric multidimensional scaling ordinations showing variation in the distribution of MC types. MC types representing individual phyla are colored in blue. A) Tenericutes (n=149), B) Spirochaetes (n=202), C) Gammaproteobacteria (n=2450), and D) Firmicutes (n=2312). The Tenericutes have a low degree of variation with only 7 MC types while the similarly represented Spirochaetes have 17 MC types. Phyla with higher representation of genomes in our sample dataset, such as the Gammaproteobacteria and the Firmicutes, also have substantially more MC types.

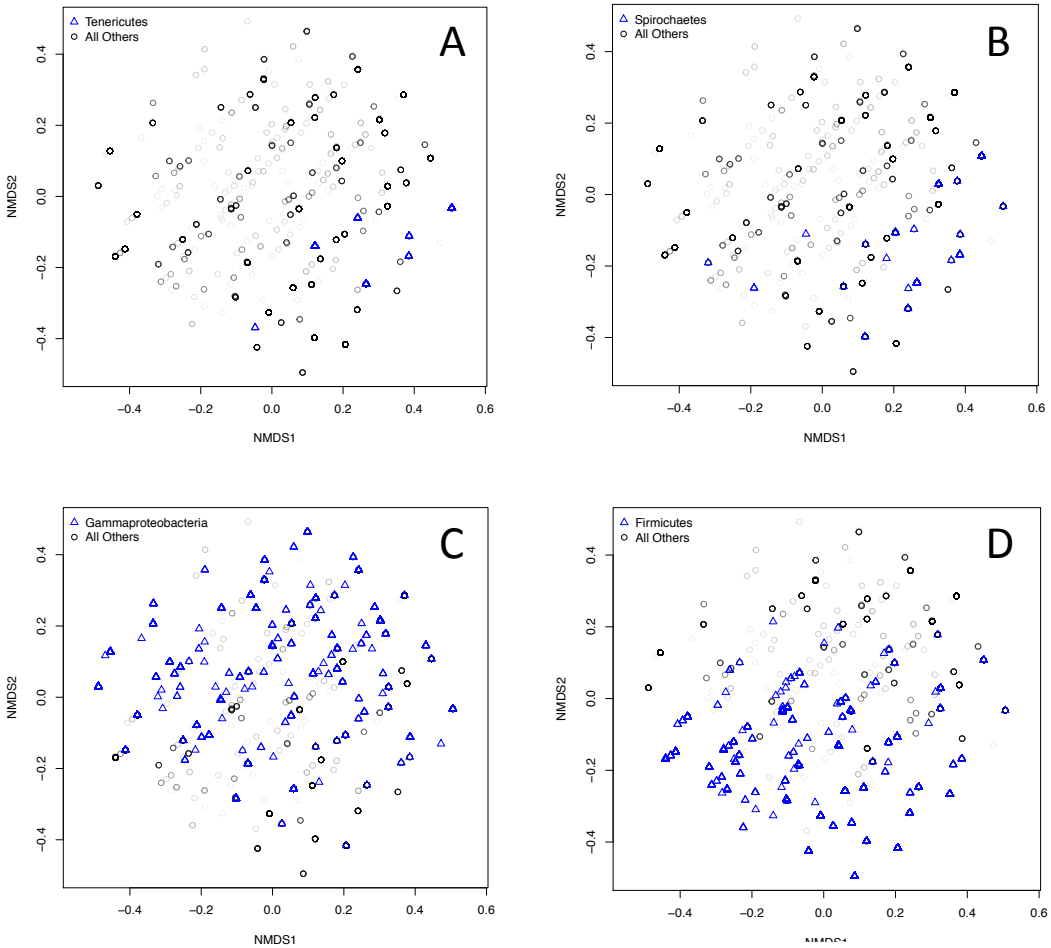


Table 2-1. Relationship of resource quantity and chemistry on changes in the abundance of highly variable and highly abundant taxa. All significant relationships presented as: p-value; R^2 (direction of response). Taxa are classified as abundant (>5,000 copies of the 16S pooled across all plots; n=7), variable (SD of copy number across all plots >200, n=4), or both (n=7).

OTU type	OTU ID	# Sequences	Litter	Root biomass	Belowground C:N	Plot-level N
both	10749	15643	n.s.	n.s.	n.s.	n.s.
abu	8337	9287	n.s.	0.09; 0.2043 (-)	n.s.	0.04; 0.2876 (-)
abu	17558	8280	n.s.	n.s.	0.007; 0.4376 (+)	n.s.
both	11863	7847	0.002; 0.5499 (+)	n.s.	n.s.	n.s.
both	798	7535	n.s.	n.s.	n.s.	n.s.
abu	5481	7175	n.s.	0.08; 0.2167 (+)	0.006; 0.4523 (+)	n.s.
abu	14126	6761	n.s.	n.s.	n.s.	n.s.
both	3757	6246	0.06; 0.2445 (+)	n.s.	n.s.	n.s.
both	36770	6024	n.s.	n.s.	n.s.	n.s.
both	6965	5977	n.s.	n.s.	n.s.	n.s.
abu	20403	5912	n.s.	0.07; 0.2284 (-)	n.s.	n.s.
abu	28570	5757	n.s.	n.s.	0.04573; 0.2729 (-)	n.s.
abu	13375	5607	n.s.	n.s.	n.s.	n.s.
abu	18724	5597	0.081; 0.2165 (+)	n.s.	n.s.	n.s.
abu	3550	5294	n.s.	n.s.	n.s.	n.s.
both	8507	5050	n.s.	n.s.	n.s.	n.s.
var	2897	3078	n.s.	n.s.	0.09; 0.2036 (-)	n.s.
var	9204	2725	n.s.	n.s.	0.02; 0.3453 (+)	n.s.
var	13817	1886	n.s.	n.s.	0.08; 0.2170 (-)	n.s.
var	32374	1292	n.s.	n.s.	n.s.	n.s.

Figure 2-1. Rarefaction curves for A) OTU_{98} richness and B) Shannon's diversity. OTU_{98} richness has started to reach an asymptote, while Shannon's diversity is stably at an asymptote by 35,000 sequences.

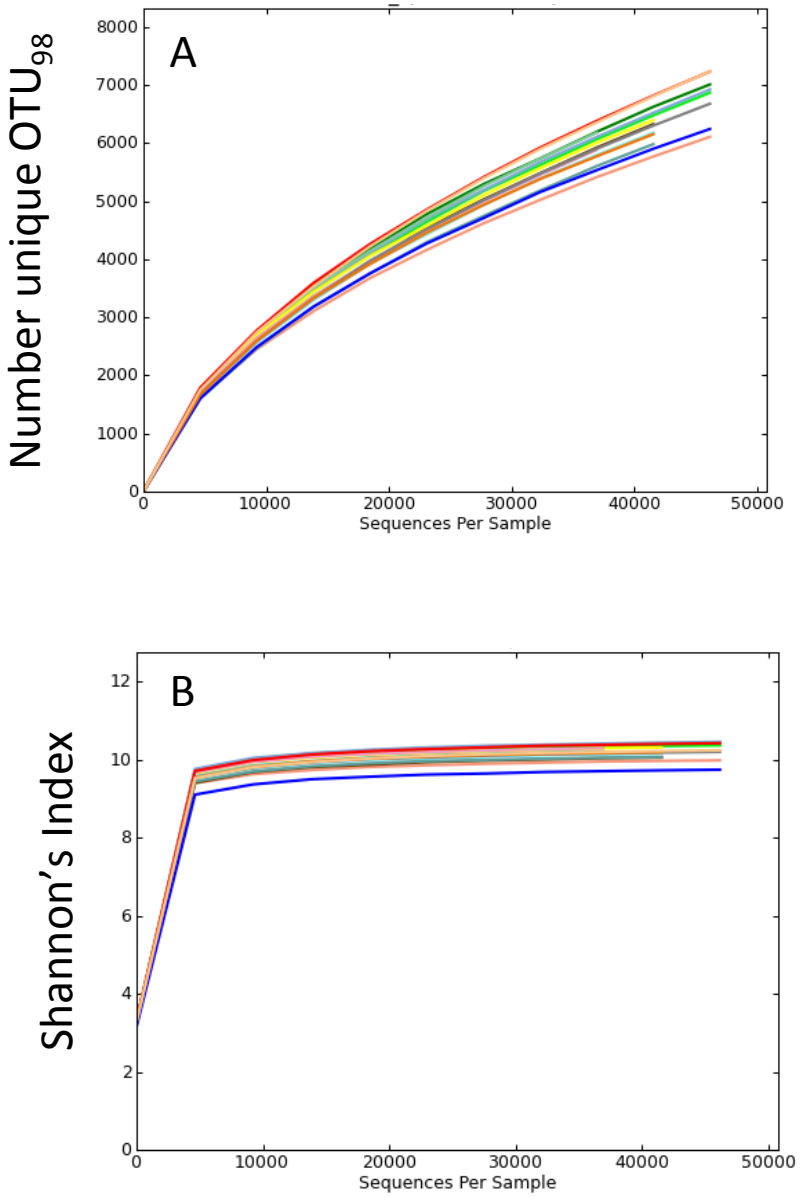


Figure 2-2. The four resources that best captured variation in resource landscape among plant species were chosen out of the 35 resources examined via nonmetric multidimensional scaling ordinations. Two quantity variables (aboveground litter and root biomass) and two chemistry variables (plot-level N availability and belowground C:N ratio) were chosen.

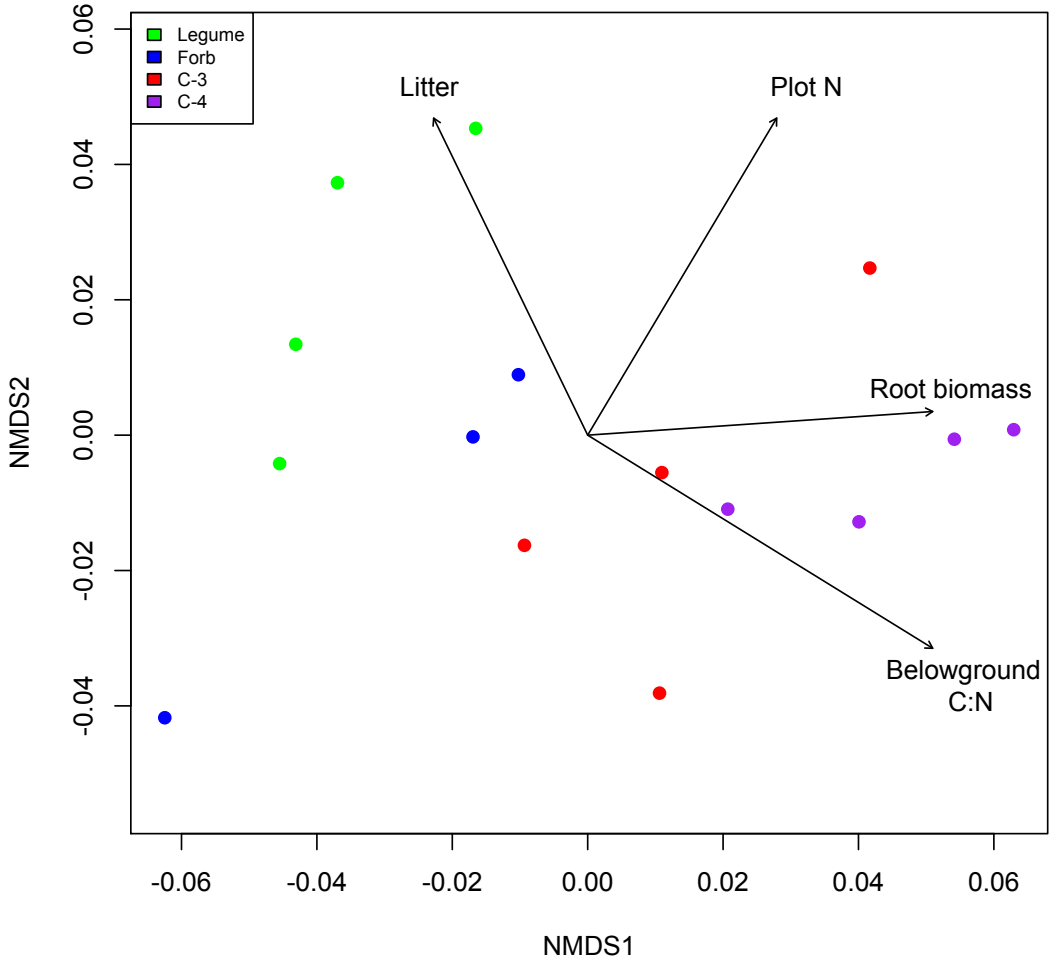


Figure 2-3. Effects of belowground resource quantity and chemistry on OTU_{98} community composition. Considered independently, belowground C:N of plants explained 52% of the variation in community composition while root biomass explained 45%. Neither aboveground litter nor plot-level N was significantly related to changes in community composition of OTU_{98}.

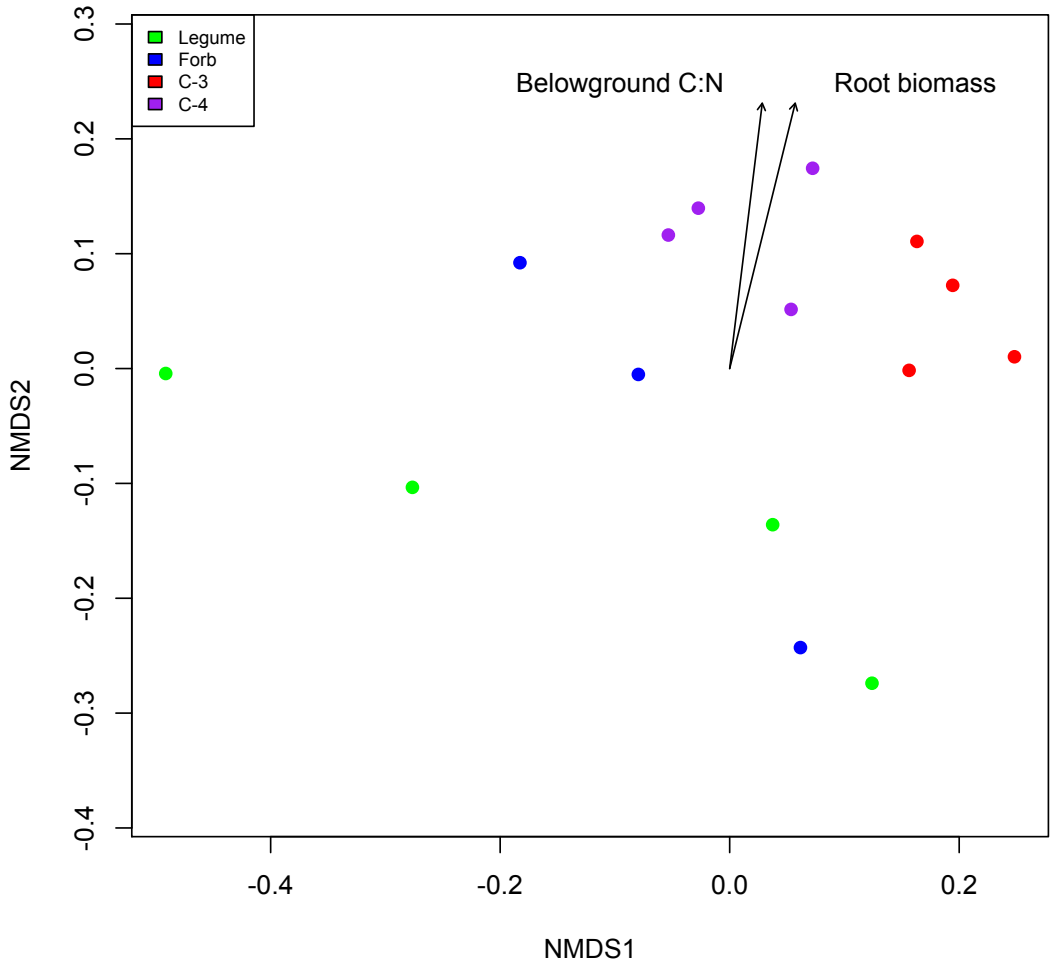


Figure 2-4. Effects of resource availability on soi bacterial community composition at the A) family and B) phyla level. While effects were weaker at the family level, belowground C:N was still related to differences in community composition. At the phyla level, however, community composition was not significantly related to the availability of any resource examined.

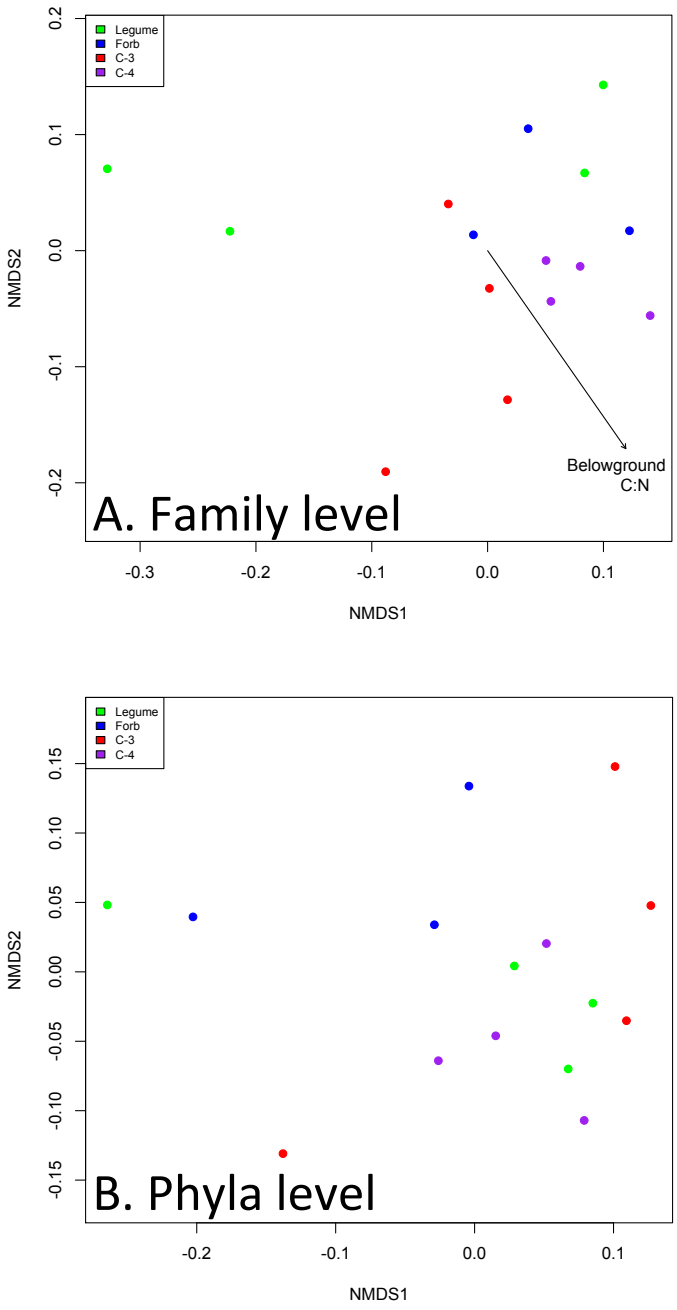


Figure 2-5. Effects of resource availability on community composition of the 16 most abundant taxa. Belowground C:N of plants explained 51% of the variation in community composition of the most abundant taxa. Neither root biomass, aboveground litter, nor plot-level N was significantly related to community composition of these taxa. However, a subset of individual taxa were related to each of these measures of resources.

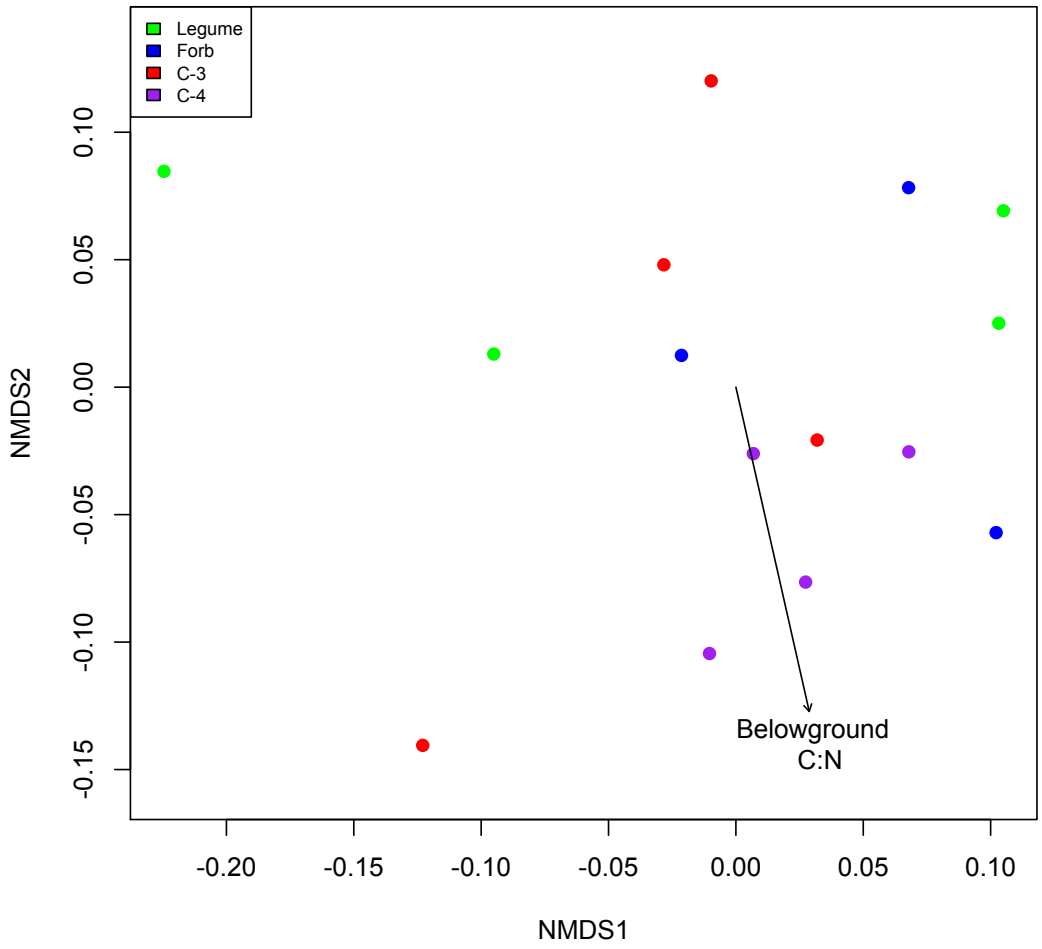


Figure 2-6. Relationship between ecosystem process rates and OTU_{98} community composition. OTU_{98} community composition was correlated with growing season CO_2 efflux rates as well as rates of net nitrification and net N mineralization (p=0.07).

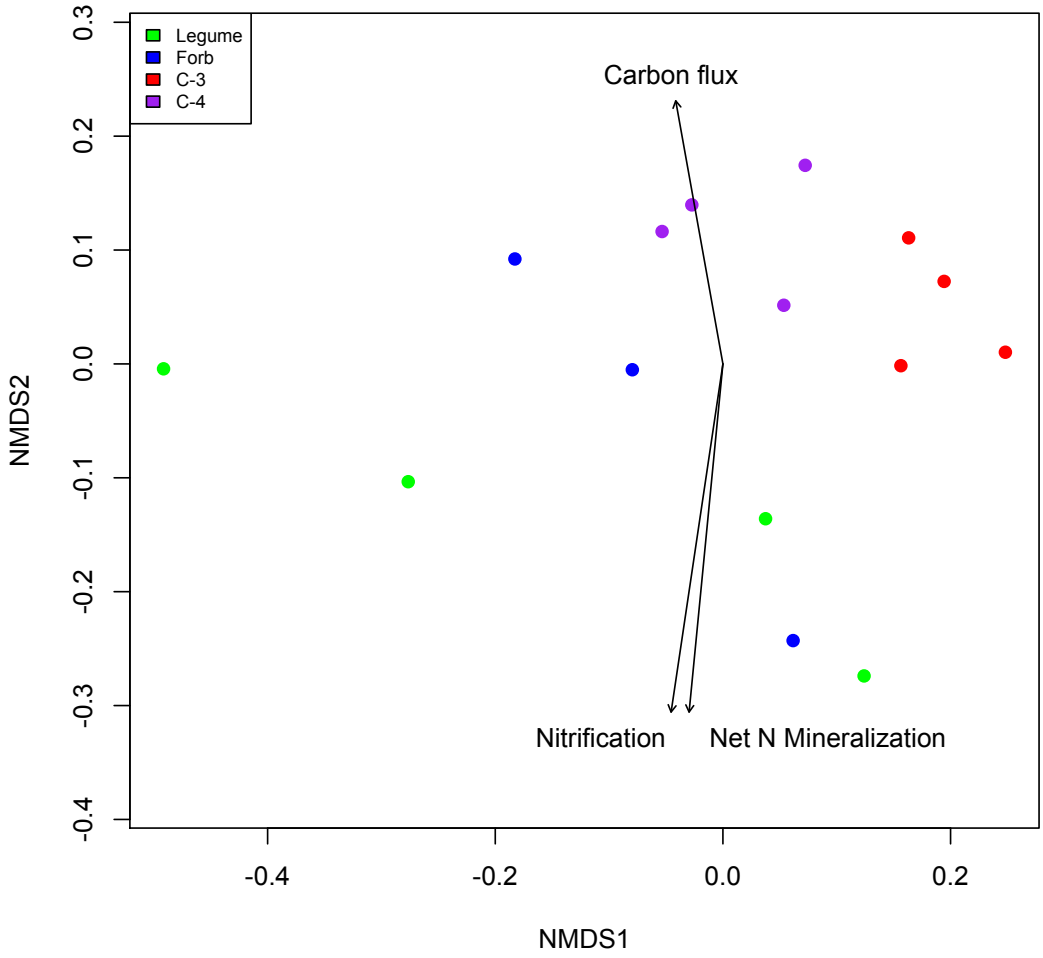


Table 3-1. Detailed specification of individual pathway presence and absence for each abundant MC type (A-L). N is the number of isolates representative of that MC type.

	A	B	C	D	E	F	G	H	I	J	K	L
Gluconeogenesis	1	1	1	1	0	1	1	1	1	1	1	1
Glycolysis	0	1	1	1	0	0	0	0	0	0	1	1
Pentose Phosphate	1	1	1	1	1	1	1	1	1	1	0	1
Entner-Duodoroff	1	0	1	1	0	1	1	1	1	1	0	0
Lactate	1	1	0	1	1	1	1	1	1	1	1	1
Formate	0	0	0	0	0	0	0	0	0	0	0	0
Ethanol	1	0	0	0	0	0	0	1	0	1	0	0
Acetate	1	1	0	0	0	0	1	0	1	1	1	0
Acetone	0	0	0	0	0	0	0	0	1	1	0	0
Butyrate	0	0	0	0	0	0	0	0	0	0	0	0
Butanol	0	0	0	0	0	0	0	0	0	0	0	0
2,3 butanediol	0	0	0	0	0	0	0	0	0	0	0	0
n	9	7	7	10	1	1	1	1	1	2	1	2

Table 3-2. Average positive growth rate ($APGR_{CC}$, change in log(OD600)/hour) on individual substrate classes by ES type. Letters in parenthesis denote significant differences.

	Generalist	Intermediate	Specialist	p-value
Monosaccharides	0.0156 (a)	0.0118 (a)	0.00652 (b)	<0.0001
Sugar alcohols	0.0179 (a)	0.0119 (b)	0.00535 (c)	<0.0001
Other sugars	0.0157 (a)	0.00766 (b)	0.00600 (b)	<0.001
Amino acids	0.0119 (a)	0.00939 (a)	0.00540 (b)	<0.0001
Acyclic carboxylic acids	0.0144 (a)	0.0101 (a)	0.00416 (b)	<0.0001
Other carboxylic acids	0.0124 (a)	0.0102 (a)	0.00555 (b)	<0.01
Other C compounds	0.01282 (a)	0.00685 (b)	0.00594 (b)	<0.01

Table 3-3. Average positive growth rate ($APGR_{CC}$, change in log(OD600)/hour) on individual substrate classes by MC type. Letters in parenthesis denote significant differences.

	A	B	C	D	p-value
Monosaccharides	0.0131 (a)	0.00115 (a)	0.00610 (b)	0.00470 (b)	<0.0001
Sugar alcohols	0.0120 (a)	0.0134 (a)	0.00574 (b)	0.00319 (b)	<0.0001
Other sugars	0.00790 (ab)	0.0131 (a)	0.00485 (b)	0.00517 (b)	<0.001
Amino acids	0.0103 (a)	0.00914 (a)	0.00463 (b)	0.00446 (b)	<0.0001
Acyclic carboxylic acids	0.0118 (a)	0.00584 (b)	0.00309 (b)	0.00325 (b)	<0.0001
Other carboxylic acids	0.0125 (a)	0.00765 (ab)	0.00387 (b)	0.00507 (b)	<0.01
Other C compounds	0.00882 (ab)	0.0130 (a)	0.00379 (c)	0.00556 (bc)	<0.001

Table 3-4. Maximum observed growth rate (MGR, change in log(OD600)/hour), average positive growth rate (APGR, change in log(OD600)/hour), and metabolic flexibility for isolates containing (+) or lacking (-) individual metabolic pathways.

Pathway	MGR	p-value	APGR	p-value	Metabolic Flexibility	p-value
Glycolysis (+)	0.015	0.0038	0.0069	<0.001	30.9	<0.0001
Glycolysis (-)	0.0286		0.0113		46.5	
Ethanol (+)	0.0287	0.034	0.0116	<0.01	46.2	<0.001
Ethanol (-)	0.0167		0.00737		33	
Entner-Doudoroff (+)	0.0194	n.s.	0.00782	p=0.066	37.6	n.s.
Entner-Doudoroff (-)	0.022		0.0107		34	

Table 3-5. Average positive growth rate ($APGR_{CC}$, change in log(OD600)/hour) on individual substrate classes by PC type. Letters in parenthesis denote significant differences.

	Burkholderiaceae	Micrococcaceae	Xanthomonadaceae	p-value
Monosaccharides	0.0133 (a)	0.0124 (a)	0.00526 (b)	<0.0001
Sugar alcohols	0.0125 (a)	0.0165 (a)	0.00447 (b)	<0.0001
Other sugars	0.00649 (a)	0.0152 (b)	0.00530 (a)	<0.0001
Amino acids	0.0105 (a)	0.00975 (a)	0.00451 (b)	<0.0001
Acyclic carboxylic acids	0.0114 (a)	0.00707 (b)	0.00352 (c)	<0.0001
Other carboxylic acids	0.0123 (a)	0.00839 (ab)	0.00447 (b)	<0.0001
Other C compounds	0.00685 (a)	0.0141 (b)	0.00526 (a)	<0.0001

Figure 3-1. Correlation between metabolic flexibility and both A) the maximum observed growth rate and B) average positive growth rates. Growth rates measured as change in log(OD600)/hour.

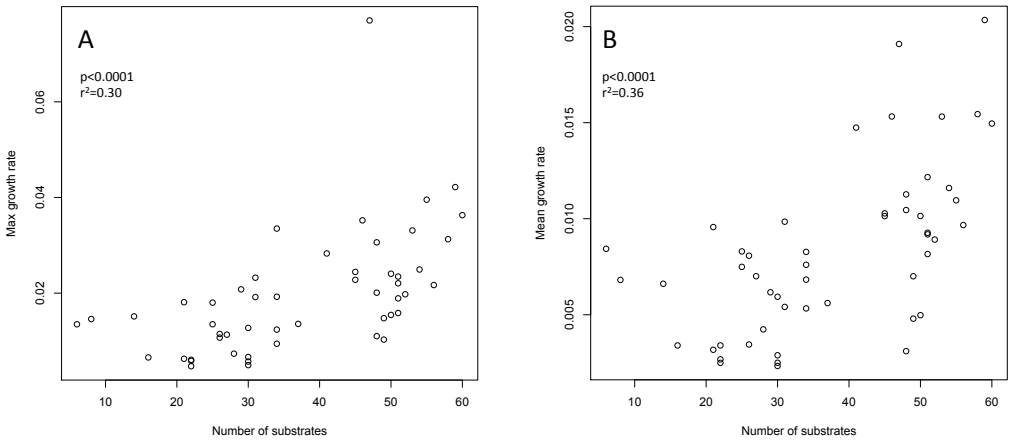


Figure 3-2. Differences among metabolic chassis types in A) maximum observed growth rate (MGR, change in log(OD600)/hour), B) average positive growth rate (APGR, change in log(OD600)/hour), C) metabolic flexibility, and D) genome size. Labels on the bottoms of bars represent the presence or absence of glycolysis, ethanol fermentation, and ED, respectively. Labels at tops of bars indicate significant differences. MC types A and B grew more quickly than either C or D. These patterns in growth rate differences held across substrate classes; although there were some substrates for which no patterns were observed, there were no substrates where the pattern was reversed.

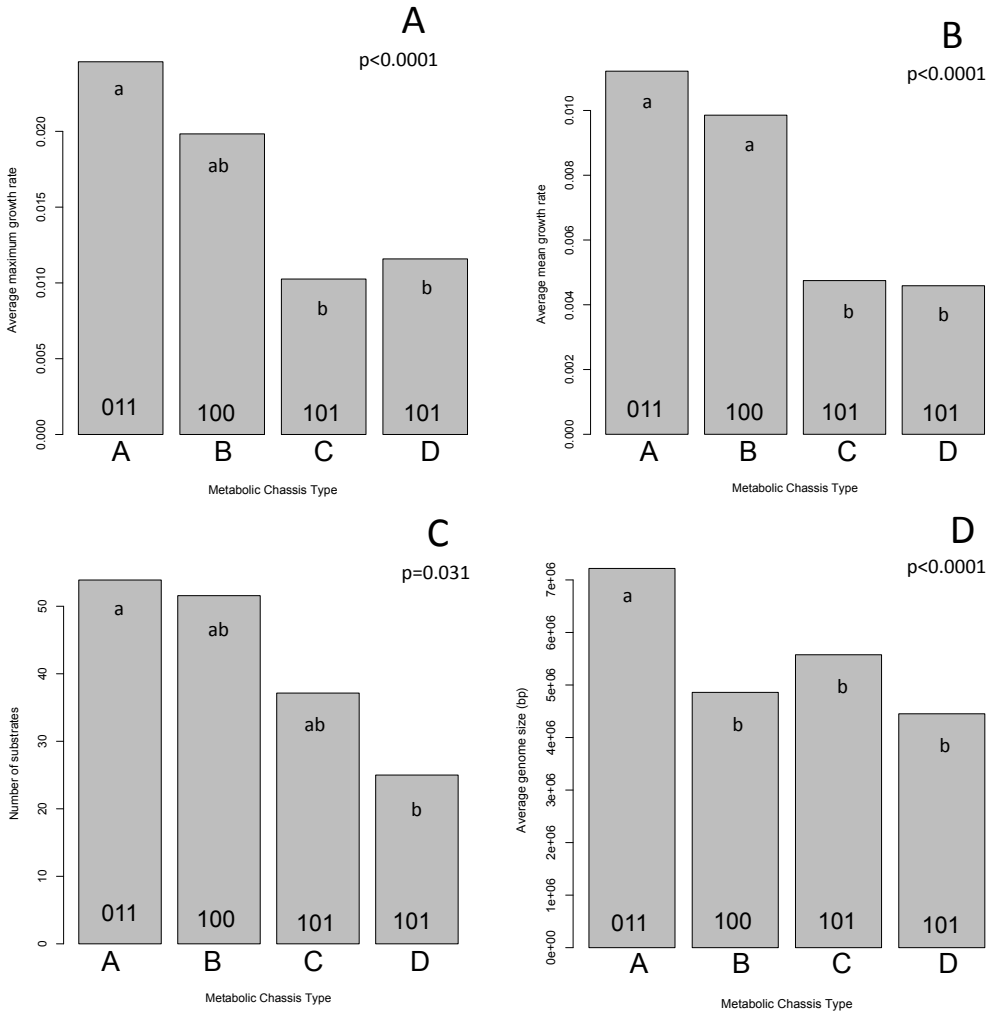


Figure 3-3. Differences among isolates containing A) glycolysis, B) ethanol fermentation, and C) ED in average positive growth rate (APGR, change in log(OD600)/hour). Isolates containing a complete ethanol fermentation pathway grew more quickly, while those containing glycolysis, and to a lesser extent ED, grow more slowly.

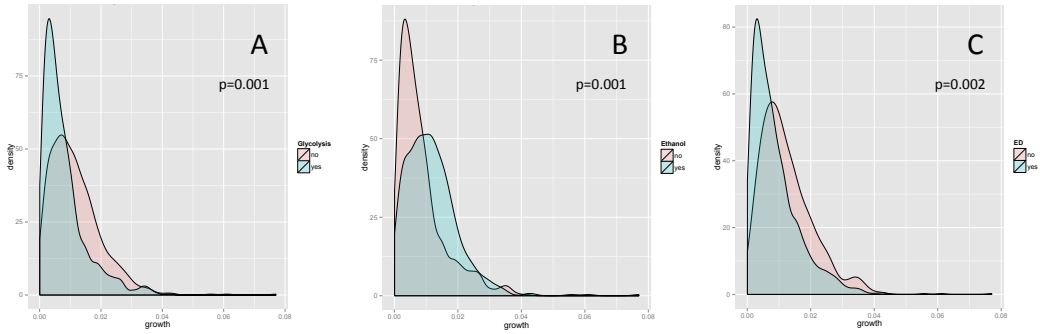
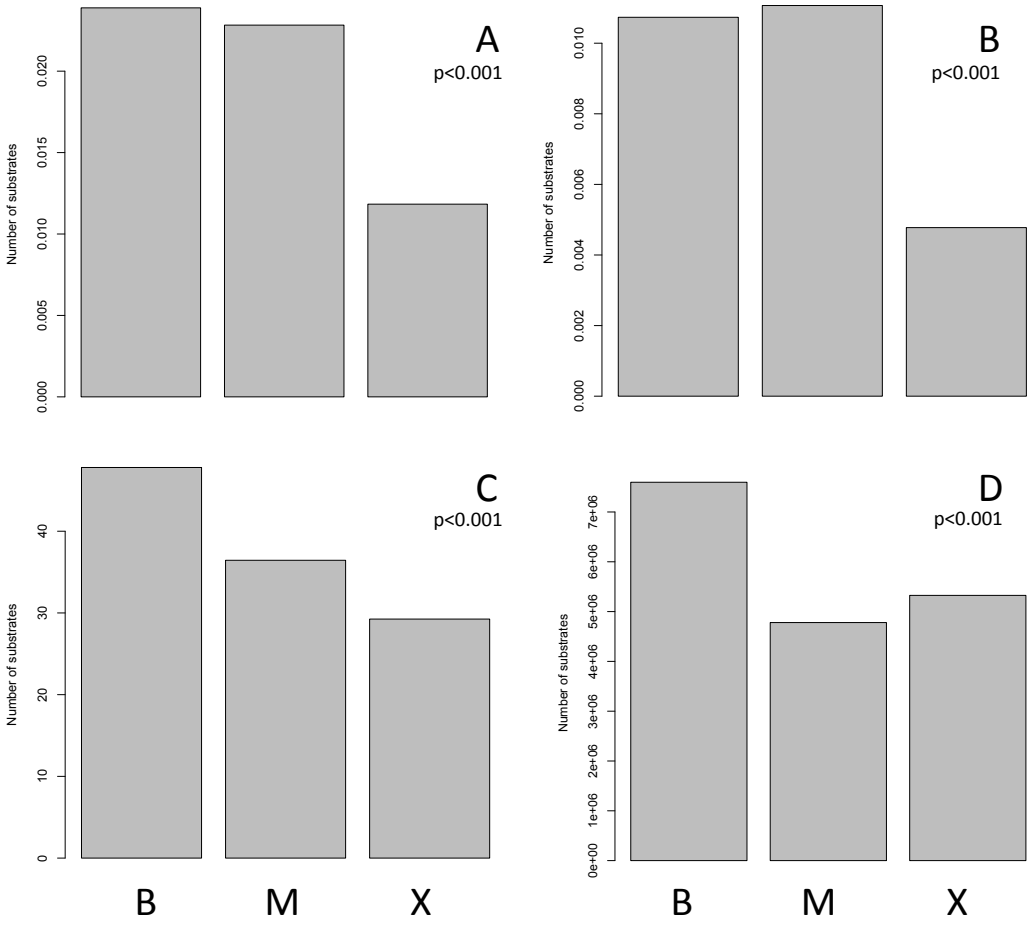


Figure 3-4. Differences among isolates from different families (B=Burkholderiaceae, M=Micrococcaceae, and X=Xanthomonadaceae) in A) maximum observed growth rate (MGR), B) average positive growth rate (APGR), C) metabolic flexibility, and D) genome size. In general, isolates in the Burkholderiacaeae grew faster than those in the Xanthomonadaceae. These patterns in growth rates held across substrate classes; although there are some substrates for which no patterns were observed, there were no substrates where the pattern was reversed.



Bibliography

Allison, V.J., Condron, L.M., Peltzer, D.A., Richardson, S.J., and B.L. Turner. 2007. *Changes in enzyme activities and soil microbial community composition along carbon and nutrient gradients at the Franz Josef chronosequence, New Zealand*. Soil Biology and Biochemistry 39: 1770-1781.

Atlas, R.M. 2005. *Handbook of media for environmnental microbiology, second edition*. CRC Press, Boca Raton, FL. ISBN: 0-8493-3560-4.

Balvanera, P., Pfisterer, A.B., Buchmann, N., He, J.S., Nakashizuka, T., Raffaelli, D., and B. Schmid. 2006. *Quantifying the evidence for biodiversity effects on ecosystem functioning and services*. Ecology Letters 9: 1146-1156.

Bell, T., Newman, J.A., Silverman, B.W., Turner, S.L., and A.K. Lilley. 2005. *The contribution of species richness and composition to bacterial services*. Nature. 436: 1157-1160.

Blomberg, S.P., Garland, T., Jr., and A.R. Ives. 2003. *Testing for phylogenetic signal in comparative data: behavioral traits are more labile.* Evolution. 57:717-745.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. and R. Knight. 2010. *QIIME allows analysis of high-throughput community sequencing data*. Nature Methods. 7: 335-336.
http://www.nature.com/ismej/journal/vaop/ncurrent/full/ismej20128a.html

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., and R. Knight. 2012. *Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms*. The ISME Journal. 6: 1621-1624.

Cardinale, B.J., Matulich, K.L., Hooper, D.U., Byrnes, J.E., Duffy, E., Gamfeldt, L, Balvanera, P., O'Connor, M.I., and A. Gonzalez. 2011. *The functional role of producer diversity in ecosystems*. American Journal of Botany. 98: 572-592.

Carpenter, D. *Growth rate algorithm*. http://www.princeton.edu/genomics/botstein/protocols/Growth-Rate-Using-R.pdf

Castro, H.F., Classen, A.T., Austin, E.E., Norby, R.J., and C.W. Schadt. 2010. *Soil microbial community responses to multiple experimental climate change drivers*. Applied and Environmental Microbiology 76: 999-1007.

Cebrian, J. 1999. *Patterns in the fate of production in plant communities.* The American Naturalist. 154: 449-468.

Chung, H. Zak, D.R., Reich, P.B., and D.S. Ellsworth. 2007. *Plant species richness, elevated CO_2, and atmospheric nitrogen deposition alter soil microbial community composition and function*. Global Change Biology. 13: 980-989.

Clayton, G.W., Rice, W.A., Lupwayi, N.Z., Johnston, A.M., Lafond, G.P., Grant, C.A., and F. Walley. 2004. *Inoculant formulation and fertilizer nitrogen effects on field pea: Nodulation, N2 fixation, and nitrogen partitioning*. Canadian Journal of Plant Science. 84: 79-88.

Cyr, H. and M.L. Pace. 1993. *Allometric theory: extrapolations from individuals to communities.* Ecology. 74: 1234-1245.

Department of Energy Systems Biology Knowledgebase (KBase), http://kbase.us.

Dijkstra, F.A., Hobbie, S.E. and P.B. Reich. 2006. *Soil processes affected by sixteen grassland species grown under different environmental conditions*. Soil Science Society of America Journal. 70: 770-777.

Eilers, K.G., Lauber, C.L., Knight, R., and N. Fierer. 2010. *Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil*. Soil Biology and Biochemistry. 42: 896-903.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., and M.L. Sogin. 2013. *Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data.* Methods in Ecology and Evolution. 4: 1111-1119.

Fierer, N., M.A. Bradford, and R.B. Jackson. 2007. *Toward an ecological classification of soil bacteria.* Ecology. 88: 1354-1364.

Fierer, N. and R.B. Jackson. 2006. *The diversity and biogeography of soil bacterial communities*. Proceedings of the National Academy of Sciences. 103: 626-631.

Firestone, M.K., and E.A. Davidson. 1989. *Microbiological basis of NO and N2O production and consumption in soil.* In Exchange of trace gases between terrestrial ecosystems and the atmosphere, pp 7-21. Andreae, M.O. and D.S. Schimel, eds. ISBN: 0-471-92551-9.

Friedrich, M.W. 2002. *Phylogenetic analysis reveals multiple lateral transfers of adenosine-5'-phosphosulfate reductase genes among sulfate-reducing microorganisms.* Journal of Bacteriology. 184: 278-289.

Fritz, S.A. and A. Purvis. 2010. *Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits.* Conservation Biology. 24: 1042-1051.

Gower, J.C. and P. Legendre. 1986. *Metric and Euclidean properties of dissimilarity coefficients.* Journal of Classification. 3: 5-48.

Gubry-Rangin, C., Nicol, G.W., and J.I. Prosser. 2010. *Archaea rather than bacteria control nitrification in two agricultural acidic soils.* FEMS Microbiology Ecology. 74: 566-574.

Gudelj, I., Weitz, J.S., Ferenci, T., Horner-Devine, M.C., Marx, C.J., Meyer, J.R., and S.E. Forde. 2010. *An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure.* Ecology Letters. 13: 1073-1084.

He, Z., Xu, M., Deng, Y., Kang, S., Kellogg, L., Wu, L., Van Nostrand, J.D., Hobbie, S.E., Reich, P.B., and J. Zhou. 2010. *Metagenomic analysis reveals a marked divergence in the structure of belowground microbial communities at elevated CO_2*. Ecology Letters 13: 564-575.

Hooper, D.U., Chapin, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J.H., Lodge, D.M., Loreau, M., Naeem, S., Schmid, B., Setala, H., Symstad, A.J., Vandermeer, J., and D.A. Wardle. 2005. *Effects of biodiversity on ecosystem functioning: A consensus of current knowledge*. Ecological Monographs 75: 3-35.

Human Microbiome Project Consortium. 2012. *Structure, function, and diversity of the healthy human microbiome*. Nature 486: 207-214.

Isbell, F., Calcagno, V., Hector, A., Connolly, J., Harpole, W.S., Reich, P.B., Scherer-Lorenzen, M., Schmid, B., Tilman, D., van Ruijven, J., Weigelt, A., Wilsey, B.J., Zavaleta, E.S., and M. Loreau. 2011. *High plant diversity is needed to maintain ecosystem services*. Nature 477: 199-202.

Jackson, R.B., Fierer, N., and J.P. Schimel. 2007. *New directions in microbial ecology.* Ecology. 88: 1343-1344.

Katoh, K., and D.M. Standley. 2013. *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Molecular Biology and Evolution. 30: 772-780.

Kimura, M. 1968. *Evolutionary rate at the molecular level.* Nature. 217: 624-626.

Kopp, R.E., Kirschvink, J.L., Hilburn, I.A., and C.Z. Nash. 2005. *The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis.* Proceedings of the National Academy of Sciences. 102: 11131-11136.

Lanfear, R., Calcott, B., Ho, S.Y.W., and S. Guindon. 2012. *PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses.* Molecular Biology and Evolution. 29: 1695-1701.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., a Reyes, J., Clemente, J. C., Burkepile, D.E., Vega Thurber, R. L., Knight, R., Beiko, R. G., and C. Huttenhower. 2013. *Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.* Nature Biotechnology. 31: 814-821.

Lauro, F.M., McDouglad, D., Thomas, T., Williams, T.J., Egan, S., Rice, S., DeMaere, M.Z., Ting, L., Ertan, H., Johnson, J., Ferriera, S., Lapidus, A., Anderson, I., Kyrpides, N., Munk, A.C., Detter, C., Han, C.S., Brown, M.V., Robb, F.T., Kjelleberg, S., and R. Cavicchioli. 2009. *The genomic basis of trophic strategy in marine bacteria*. Proceedings of the National Academy of Sciences. 37: 15527-15533.

Lingner, T., Muhlhausen, S., Gabaldon, T., Notredame, C., and P. Meinicke. 2010. *Predicting phenotypic traits of prokaryotes from protein domain frequencies*. BMC Bioinformatics. 11: 481-492.

Livermore, J.A., Emrich, S.J., Tan, J., and S.E. Jones. 2014. *Freshwater bacterial lifestyles inferred from comparative genomics*. Environmental Microbiology. 16: 746-758.

Lonergan, D.J., Jenter, H.L., Coates, J.D., Phillips, E.J., Schmidt, T.M., and D.R. Lovley. 1996. *Phylogenetic analysis of dissimilatory Fe(III)-reducing bacteria.* Journal of Bacteriology. 178: 2402-2408.

Meier, C.L. and W.D. Bowman. 2008. *Links between plant litter chemistry, species diversity, and below-ground ecosystem function*. Proceedings of the National Academy of Sciences. 105: 19780-19785.

Minchin, P.R. 1987. *An evaluation of the relative robustness of techniques for ecological ordination*. In Theory and models in vegetation science. Springer Verlaug. pgs. 89-107.

Northup, D.E., and K.H. Lavoie. 2001. *Geomicrobiology of caves: a review.* Geomicrobiology Journal. 18: 199-222.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., and H. Wagner. 2012. *vegan: Community Ecology Package*. R package version 2.0-4. URL: <http://CRAN-R-project.org/package=vegan>.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., and W. Pearse. 2012. *Caper: Comparative analyses of phylogenetics and Evolution in R. R package version 0.5.* http://CRAN.R-

project.org/package=caper.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., Lagard, V.C., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S. Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuy, D., Jensen, R., Jamshidi, N., Krause, L., Kuball, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rukert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, Ol, Ye, Y., Zagnitko, O., and V. Vonstein. 2005. *The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.* Nucleic Acids Research 33: 5691-5702.

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

Pearse, W.D., F.A. Jones, and A. Purvis. 2013. *Barro Colorado Island's phylogenetic assemblage structure across fine spatial scales and among clades of different ages.* Ecology 94: 2861-2872.

R Core Team. 2012. *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>.

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

Reich, P.B., Knops, J., Tilman, D., Craine, J., Ellsworth, D., Tjoelker, M., Lee, T., Wedin, D., Naeem, S., Bahauddin, D., Hendrey, G., Jose, S., Wrage, K., Goth, J., and W. Bengston. 2001. *Plant diversity enhances ecosystem responses to elevated CO_2 and nitrogen deposition*. Nature 410: 809-812.

Reich, P.B., Tilman, D., Craine, J., Ellsworth, D., Tjoelker, M.G., Knops, J., Wedin, D., Naeem, S., Bahauddin, D., Goth, J., Bengtson, W., and T.D. Lee. 2001b. *Do species and functional groups differ in acquisition and use of C, N, and water under varying atmospheric CO2 and N availability regimes? A field test with 16 grassland species*. New Phytologist 150: 435-448.

Reich, P.B., Tilman, D., Naeem, S., Ellsworth, D.S., Knops, J., Craine, J., Wedin, D., and J. Trost. 2004. *Species and functional group diversity independently influence biomass accumulation and its response to CO_2 and N*. Proceedings of the National Academy of Sciences USA 101: 10101-10106.

Schimel, J.P., and S.M. Schaeffer. 2012. *Microbial control over carbon cycling in soil*. Frontiers in Microbiology. 3: 1-11.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Ros, I.M., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madof, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappouli, R., and C.M. Fraser. 2005. *Genome analysis of multiple pathogenic isolates of* Streptococcus agalactiae: *Implications for the microbial "pan-genome".* Proceedings of the National Academy of Sciences. 102: 13950-13955.

Tilman, D. 1980. *A graphical-mechanistic approach to competition and predation*. The American Naturalist. 116: 362-393.

Tilman, D. 1994. *Competition and biodiversity in spatially structured habitats*. Ecology 75: 2-16.

Treseder, K.K. 2008. *Nitrogen additions and microbial biomass: a meta-analysis of ecosystem studies*. Ecology Letters 11: 1111-1120.

Treseder, K.K., Balser, T.C., Bradford, M.A., Brodie, E.L., Dubinsky, E.A., Eviner, V.T., Hofmockel, K.S., Lennon, J.T., Levine, U.Y., MacGregor, B.J., Pett-Ridge, J., and M.P. Waldrop. 2012. *Integrating microbial ecology into ecosystem models: challenges and priorities*. Biogeochemistry 109: 7-18.

Vieira-Silva, S. And E.P.C. Rocha. 2010. *The systemic imprint of growth and its uses in ecological (meta) genomics*. Plos Genetics. 6: e1000808.

de Vries, F.T., Manning, P., Tallowin, J.R.B., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Hobbs, P.J., Quirk, H., Shipley, B., Cornelissen, J.H.C., Kattge, J., and R.D. Bardgett. 2012. *Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities*. Ecology Letters 15: 1230-1239.

Wagg, C., Bender, S.F., Widmer, F., and M.G.A. van der Heijden. 2014. *Soil biodiversity and soil community composition determine ecosystem multifunctionality*. Proceedings of the National Academy of Science 111: 5266-5270.

Wallenstein, M.D., Myrold, D.D., Firestone, M., and M. Voytek. 2006. *Environmental controls on denitrifying communities and denitrification rates: insights from molecular methods*. Ecological Applications. 16: 2143-2152.

Wang, Q., Garrity, G.M., Tiedje, J.M., and J.R. Cole. 2007. *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy*. Applied and Environmental Microbiology 73: 5261-5267.

Waring, B.G., Averill, C., and C.V. Hawkes. 2013. *Differences in fungal and bacterial physiology alter soil carbon and nitrogen cycling: insights from meta-analysis and theoretical models*. Ecology Letters 16: 887-894.

Wieder, W.R., Bonan, G.B., and S.D. Allison. 2013. *Global soil carbon projections are improved by modelling microbial processes*. Nature Climate Change 3: 909-912.

Xu, X., Schimel, J.P., Thronton, jP.E., Song, X., Yuan, F., and S. Goswami. 2014. *Substrate and environmental controls on microbial assimilation of soil organic carbon: a framework for Earth system models*. Ecology Letters 17: 547-555.

Zhou, J., Deng, Y., Luo, F., He, Z., and Y. Yang. 2011. *Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2*. MBio 2: e00122-11.

Zuberer. 1994. *Recovery of viable bacteria*. In Methods of Soil Analysis: Part 2 Microbiological and Biochemical Properties. Weaver, R.W., Angle, S., Bottomley, P., Bezdicek, D., Smith, S., Tabatabai, A., and A. Wollum, eds.