UNIVERSITY OF CALIFORNIA,
IRVINE

Disentangling the relationship between bacterial diversity and its functioning: plant litter communities as a model system

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Michaeline Burr Nelson Albright

Dissertation Committee:
Professor Jennifer B.H. Martiny, Chair
Professor Travis E. Huxman
Associate Professor Adam C. Martiny

2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to my advisor and mentor, Jennifer Martiny. You are an inspiring role model both inside and outside of science. Thank you to my committee members, Adam Martiny and Travis Huxman, for your constructive feedback during our meetings. Thank you to Claudia Weihe for being my lab advisor, for teaching me the practical side of research, and for making my day-to-day work life more fun. Thank you to Jessica Pratt, for sharing your enthusiasm for outreach and teaching. Thank you to the J. Martiny lab for being such a great team to work with and for bringing your excitement for microbial ecology to work days and karaoke nights. Thank you to all my parents, Lee and Julie and my brother Lyle for your love and support. And finally, thank you Jason for pushing me to be the best I can be.

I would like to acknowledge the funding sources that made my dissertation research possible. Fellowship support was provided by a US Department of Education Graduate Assistance in Areas of National Need (GAANN) and the DOE Office of Science Graduate Student Research Program. Research support was provided by grants from the NSF Doctoral Dissertation Improvement program, the NSF Division of Environmental Biology, and the Office of Science (BER) DOE program in Microbial Communities and Carbon Cycling.

Chapter 1 was published with the permission of the American Society for Microbiology. The text of this chapter is a reprint of the material as it appears in the journal AEM. Chapter 2 was published with the permission of the National Academy of Sciences. The text of this chapter is a reprint of the material as it appears in the journal PNAS. Chapter 3 was published with the permission of Nature Publishing Group. The text of this chapter is a reprint of the material as it appears in The ISME journal.

CURRICULUM VITAE

Michaeline B.N. Albright
Department of Ecology and Evolutionary Biology
University of California, Irvine • Irvine, CA 92697

EDUCATION

Ph.D. in Biological Sciences: Ecology & Evolutionary Biology, *University of California, Irvine, CA* (2013-2017). Advisor: Dr. Jennifer B.H. Martiny
M.S. Biological Sciences, *University of California, Irvine, CA* (2013-2015)
B.S. Geology-Biology with Honors, *Brown University, Providence, RI* (2007-2011)

RESEARCH EMPLOYMENT

ENVIRONMENTAL MICROBIOLOGY (Graduate Student Researcher), *Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM* (2017)

MICROBIAL ECOLOGY (Graduate Student Researcher), *Ecology and Evolutionary Biology Department, University of California, Irvine, CA* (2013-2017)

MICROBIAL ECOLOGY AND BIOGEOCHEMISTRY (Undergraduate Student Thesis/Research Assistant), *Ecology and Evolutionary Biology Department, Brown University, RI* (2010-2012)

ECO-INFORMATICS SUMMER INSTITUTE (Student Researcher), Research Experience for Undergraduates (REU), *HJ Andrews Experimental Forest and Oregon State University, OR* (Summer 2010)

PALEOCLIMATOLOGY (Research Assistant), *Geological Sciences Department, Brown University, RI* (2009-2010)

COASTAL SYSTEMS (Research Assistant), *Woods Hole Oceanographic Institute*, *MA* (Summer 2009)

PUBLICATIONS (previously published under Nelson)

Peer-reviewed

Albright, MBN, and JBH Martiny. Dispersal alters bacterial diversity and composition in a natural community. In press, *The ISME Journal.*

Nelson MB, Martiny AC, Martiny JBH. 2016. Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil. *PNAS* 113:8033-8040.

Nelson MB, Berlemont R, Martiny AC, Martiny JBH. 2015. Nitrogen Cycling Potential of a Grassland Litter Microbial

Community. *Applied and Environmental Microbiology* 81(20): doi: 10.1128/AEM.02222-15.

Not peer-reviewed

Albright, MBN and JBH Martiny. Is throwing an apple core out the car window littering? – Microbial communities in natural composting. Submitted, *Frontiers for Young Minds.*

Nelson MB, Chase AB, Martiny JBH. 2016, et al. The Microbial Olympics 2016. *Nature Microbiology* 1:16122.

Frankel Z, Rowe G, Schmidt A, Schou N, Weisheit J, **Nelson MB,** Myers A, Weiser K, Anderson N, Nuttall E, Adolphson M. 2012. Crossroads Utah: Utah's Climate Future. Report. http://www.utahrivers.org/wp-content/uploads/2012/10/Crossroads.pdf

HONORS/AWARDS

William F. Holcomb Scholarship, University of California, Irvine, CA $2000 (June 2017)
Honorable Mention, NSF Graduate Research Fellowship Program (2014 & 2015)
Outstanding Research, Academics and Service, *Geological Sciences Department, Brown University*, *RI* (May 2011)

GRANTS/ FELLOWSHIPS

DOE Office of Science Graduate Student Research Program Fellow (SCGSR), (January 2017-September 2017)
Graduate Education Assistance in Areas of National Need (GAANN) Fellowship, (Fall 2016-Summer 2017)
Graduate Education Assistance in Areas of National Need, Travel and Research Grant, (2016)
NSF Doctoral Dissertation Improvement Grant (DDIG) (2016)
Graduate Education Assistance in Areas of National Need, Travel and Research Grant (2016)
Graduate Education Assistance in Areas of National Need (GAANN) Fellowship, (Spring 2016- Summer 2016)
Department Travel Grant, *University of California, Irvine*, *CA* (2016)
Associated Graduate Student Travel Grant, *University of California, Irvine*, *CA* (2016)
Department Travel Grant, *University of California, Irvine*, *CA* (2014)
Travel Grant, *Oregon State University, OR* (2010)

PRESENTATIONS/POSTERS

Invited

Ecological Society of America Annual Meeting (2017) "Quantifying stochastic variation in microbial composition and functioning in the field" Portland, Oregon (Invited Talk)

American Geophysical Union Annual Meeting (2016) "Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil" San Francisco, CA. (Invited Talk)

Contributed
Lake Arrowhead Microbial Genomics Meeting (2016) “Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil” Lake Arrowhead, CA. (Poster) *Received a Poster Award

Associated Graduate Student Symposium, UC Irvine (2016) “Global Biogeography of Microbial Nitrogen-Cycling Traits in Soil” Irvine, CA. (Talk)

Orange County Graduate Women in Science Conference, UC Irvine (2015) “Nitrogen Cycling Potential of a Plant Litter Microbial Community” Irvine, CA. (Talk)

Winter Ecology and Evolutionary Biology Graduate Student Symposium, UC Irvine (2015) “Metagenomic Characterization of Microbial Nitrogen Cycling on Plant Litter” Irvine, CA. (Talk)

International Symposium for Microbial Ecology (2014) “Metagenomic Characterization of Nitrogen Cycling by the Bacterial Community on Plant Litter” Seoul, South Korea. (Poster)

New England Estuarine Research Society Spring Meeting (2012), “Influence of Nitrate and Carbon Loading Rates on Denitrification and Anammox in a Sediment Reactor” Plymouth, MA. (Poster)

American Geophysical Union Annual Meeting (2010), “Diel Fluctuations in Summer Streamflow Depend on Stream Channel Sediment and Valley-Floor Vegetation in the Forested Western Cascades of Oregon” San Francisco, CA. (Poster)

RELATED PROFESSIONAL EXPERIENCE

Teaching
GRADUATE TEACHING ASSISTANT, *University of California, Irvine, CA* (2013- 2016)
(Teaching Assistant) Led discussion sections and lab sections for undergraduate biology courses during 8 quarters of classes at UC Irvine. Courses included (# of quarters taught): DNA to Organisms (1), Restoration Ecology (2), Field Methods in Ecology (1), Freshwater Field Biology (1), Organisms to Ecosystems (2), Physiology Lab (1).

ELEMENTS WILDERNESS, *Huntington, UT* (Spring-Summer 2013)
(Assistant Lead Field Instructor) Worked with high school students struggling with low academic performance, learning differences, and substance abuse, teaching communication and leadership skills. Used the outdoors and outdoors skills as a teaching and learning module for students.

HIGHER GROUND LEARNING, *Salt Lake City, UT* (Fall 2012)
(Tutor) Worked one-on-one with high school students on math and science. Taught SAT/ACT prep classes.

UNDERGRADUATE TEACHING ASSISTANT, *Brown University, RI* (Fall 2010)
(Teaching Assistant) Helped the graduate student teaching assistant with laboratory sections for class and field trips for a Sedimentation and Stratigraphy course.

Outreach
RESEARCH METHODS COURSE (June 2016, & March 2017)
(Guest Speaker). Gave a presentation on my graduate research, specifically highlighting the research process and components of experimental design for a group of community college education students from Mt. SAC. Gave a lab tour.

ORANGE COUNTY SOCIETY FOR CONSERVATION BIOLOGY (January 2015-October 2016) (Secretary, Board of Directors) Organize and host the "Conservation Café" series, a forum designed to encourage informal science-based discussions of conservation issues of importance to Orange County, CA. Develop outreach activity related to local research on microbial communities titled: "Hidden Biodiversity".

CRYSTAL COVE STATE PARK (2015-2017)
(Mentor) Mentor UCI Center for Environmental Biology interns. Develop a learning module on microbial diversity, related to research done in the J. Martiny Lab, for 4^{th} /5^{th} graders coming for school programs at Crystal Cove State Park.

GIRLS ON THE RUN, *Santa Ana, CA* (Fall 2013-Spring 2016)
(Coach) Teach life skills and help youth girls to develop confidence through physical activity. Provide opportunities to low income/minority students (10 week sessions/4 hours per week).

UNDERGRADUATE RESEARCH EXPERIENCE 199 STUDENT, *Irvine, CA* (Spring 2015-Spring 2016)
(Mentor) Mentor upper-level undergraduate students (2) on independent research projects. One student was funded as an Undergraduate Research Opportunity Fellow and presented a research poster at the UC Irvine undergraduate research symposium.

BREWS AND BRAINS, *Newport Beach, CA* (Oct 2015)
(Guest Speaker) Gave a presentation and answered questions on the topic of microbial biogeography related to my research to a general audience in an informal science café setting.

DISCOVERY CUBE ORANGE COUNTY, CA (Spring 2015)
(Outreach) Developed a workshop "Navigating the World" which was run at Inventor's Week at the Discovery Cube. The workshop focused on teaching about scale and map and compass skills.

GRADUATE STUDENT SYMPOSIUM (Winter 2015)
(Event Coordinator) Planned the annual Ecology and Evolutionary Biology Symposium for students to present research talks to the department.

ASK-A-SCIENTIST NIGHT*, Irvine, CA* (Oct, 2013/2014/2015/2016)
(Volunteer) Gave advice to middle school students about their designs for science fair projects.

UTAH RIVERS COUNCIL, *Salt Lake City, UT* (Fall 2012)
(Volunteer) Helped with research and writing for a climate change report.
(Scientific Outreach Writer) Worked on website design and writing.

ECOLIBRI, *Lake Atitlan, Guatemala* (Jan, 2010)
(Volunteer) Spent three-weeks working with an organization focusing on advancing health, education, environment, economy and independence for the community through sustainable projects.

WOMEN IN SCIENCE AND ENGINEERING*, Brown University, RI* (2010- 2011)
(Mentor) Provided first-year and second-year women with academic and community guidance in the sciences.

Other

Canadian Institute for Ecology and Evolution (CIEE) Working Group: The evolution of microbial metabolic and genomic diversity at multiple scales, Invited Participant. (March 2017)

Communications Excellence Certificate Program, UC Irvine Graduate Resource Center (8-week course, Spring 2015)

Mentoring Excellence Certificate Program, UC Irvine Graduate Resource Center (6-week course Winter 2015)

Regional Association of Biology Laboratory Education Meeting/Workshop, UC Irvine (Feb, 2015)

Microbial Genomics and Metagenomics Workshop, *U.S. Department of Energy Joint Genome Institute, CA* (Feb 3-7, 2014)

ABSTRACT OF THE DISSERTATION

Disentangling the relationship between bacterial diversity and its functioning: plant litter communities as a model system

By

Michaeline Burr Nelson Albright

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2017

Professor Jennifer B.H. Martiny, Chair

Microbes are key players in global biogeochemical cycles. Despite their importance, many ecosystem models do not explicitly consider microbial communities and their functions. One reason for this is that we lack a quantitative understanding of the role that microbes play in biogeochemical processes, making their incorporation into models difficult. My dissertation takes a step towards establishing these links between microbial community composition and ecosystem function using two different approaches. The first approach was characterizing the patterns and drivers of a handful of traits associated with a key biogeochemical process, the nitrogen (N) cycle and then asking which taxa were associated with this trait. I tested this approach in one ecosystem and provided a blueprint of the nitrogen cycling potential of a grassland litter microbial community (Chapter 1). I then extended this work to characterize the global biogeography of microbial N cycling traits and investigated what environmental drivers might underlie these patterns (Chapter 2). Moving beyond patterns, understanding the processes driving the distribution of microbial communities presents a further challenge.

Thus, the second approach taken in my dissertation was an experimental approach to investigate the local processes driving variation in bacterial community composition and functioning. More specifically, I focused on disentangling the effects of selection, drift, and dispersal on community assembly. First, I investigated how dispersal influences the assembly of this natural bacterial community using time series data from a field experiment (Chapter 3). I found that changing dispersal rate altered bacterial colonization rates and led to differences in the abundance, richness, evenness, and composition of communities. I then used another field experiment to quantify the role of stochastic processes in shaping microbial communities (Chapter 4). Here, I identified stochastic variation in bacterial community composition even after accounting for measurement error. Furthermore, stochastic variation in community composition translated into variation in functional parameters. Ultimately, the ability to accurately quantify stochastic processes is paramount to determining the predictability of community composition and functioning, whether focused on bacteria that degrade plant litter, microbes in the human gut, or patterns of global biodiversity.

INTRODUCTION

Bacteria drive many key biogeochemical processes, however, it can be difficult to link functioning to changes in bacterial communities under field conditions. One reason for this is that bacterial communities are often highly diverse, especially in terrestrial ecosystems such as plant litter and soils. A single gram of soil is estimated to house ~10^{10} bacterial cells with an estimated species diversity of between 2000 and $8.3x10^{6}$ species (Gans *et al.* 2005; Schloss & Handelsman 2006; Roesch *et al.* 2007). Linking bacterial diversity and its functioning is an important challenge as it can give insight into how current ecosystems work. In addition, this information allows for predictions as to how an ecosystem will look in the future, for example in response to global change (Gubry-Rangin *et al.* 2011; Raes *et al.* 2011; Zhou *et al.* 2013). My dissertation uses a combination of recent advances in sequencing technology (metagenomics) and new experimental approaches to explore links in variation in taxonomic composition to functional metrics.

Building on early studies in plant and animal biogeography, the field of microbial biogeography took off with the ability to characterize and quantify microbial diversity across space in natural environments (Green *et al.* 2008). As with early macroorganism studies, microbial biogeography initially used a taxonomic approach (Martiny *et al.* 2006). Most of these studies investigated how the environment influenced microbial community composition. To do this, they assayed community composition across many different sites and spatial scales ranging from centimeters to kilometers. Variation in taxonomic composition was then associated to variability in the physical template, including parameters such as pH, salinity, organic C, and C:N ratios (Fierer & Jackson

2006; Lozupone & Knight 2007; Rousk *et al.* 2010; Bates *et al.* 2011; Nemergut *et al.* 2011). These studies have demonstrated, that as with macro-organisms such as plants, the composition of microbial communities varies over space and time (Martiny *et al.* 2006; Fierer & Ladau 2012). They have also revealed information on the environmental processes underlying community composition. However, it is still unclear how these communities might contribute to ecosystem functioning.

To address this gap, biogeographic studies are increasingly considering functional traits. This movement has been driven by the goal of predicting changes in terrestrial ecosystem function in response to global changes (Lavorel & Garnier 2002; McGill *et al.* 2006; Green *et al.* 2008; Krause *et al.* 2014). Furthermore, for microorganisms, new technologies such as metagenomic sequencing are rapidly expanding the information available about microbial communities, including their functional traits (Tringe & Rubin 2005; Raes & Bork 2008). Metagenomic data provide information about the taxa present as well as a blueprint of their potential to contribute to functional processes, where underlying information on the genomic structuring of key organismal traits is well known. While it remains a challenge to interpret the large amount of data produced, metagenomes provide an opportunity to consider microbial traits in an environmental context (Barberan *et al.* 2012). Ultimately, incorporating trait-based approaches in microbial biogeography may provide insight into the links between microbial community composition and ecosystem function (Green *et al.* 2008; Raes *et al.* 2011; Louca *et al.* 2016b).

Most previous studies linking microbial community composition and function to environmental parameters have taken the approach of identifying factors that select on

the entire suite of microbial traits (Louca *et al.* 2016a). I reversed this direction of inquiry, instead first characterizing the patterns and drivers of a handful of traits associated with a key biogeochemical process and then asking which taxa are associated with this trait. Taking this approach, I focused on the nitrogen (N) cycle. There are several reasons why N cycling traits are important to consider: 1) N is a key element for all of life, 2) In the environment, microbes are largely responsible for many of the N compound transformations or pathways in the N cycle 3) Humans are altering the N cycle through factory and automobile emissions and the production of inorganic fertilizers. I first characterized N cycling traits representing eight different N cycling pathways in one ecosystem, a southern California grassland, using metagenomic data spanning 2 years and 3 global change treatments (Chapter 1; (Nelson *et al.* 2015)). With this work I showed that seasonal variation impacts the abundance and composition of N cycling traits in this grassland. This study provided a baseline for comparison to other systems. I then extended this approach using 365 publicly available metagenomes to provide the first characterization of the global biogeographic patterns of microbial N cycling traits in soil (Chapter 2; (Nelson *et al.* 2016)).

Biogeographic studies, whether focused on taxonomic or trait composition, demonstrate that the abiotic environment plays a large role in determining microbial composition. However, moving beyond patterns, understanding the processes driving the distribution of microbial communities presents a further challenge. The four processes driving patterns observed in microbial biogeography include, selection, dispersal, drift, and mutation (Hanson *et al.* 2012). Generally, these processes are categorized as either stochastic or deterministic. Deterministic variability in community

composition arises from environmental factors and biotic interactions that select for predictable differences in composition. In contrast, stochastic variability emerges from random differences in replication, death, mutation, and dispersal among individuals in a community (Hubbell 2001; Vellend 2010; Chase & Myers 2011). The role of dispersal is particularly complex, as dispersal may be stochastic or deterministic. Dispersal rates also have the potential to impact ecological drift, stochastic changes in the relative abundance of organisms in a community (Vellend 2010; Nemergut *et al.* 2013a). Specifically, high dispersal rates can limit drift in communities, as migration between locations can homogenize variation in species composition among locations.

Quantifying the dispersal of microorganisms is a challenge due to their small size and high abundance (Nemergut *et al.* 2013a), therefore the relationship between dispersal rates and microbial diversity or composition remains to be tested. Furthermore, these relationships are likely complex and dependent on environmental context (Chase 2007; Louca *et al.* 2016a; Evans *et al.* 2017). More broadly, a solid foundation has accumulated recognizing the importance of both stochastic and deterministic processes in influencing diversity. This foundation has been built on both observational studies and field experiments that track changes in communities across spatial and temporal gradients, while measuring numerous deterministic, potentially selective biotic and abiotic variables (Bell 2010; Ferrenberg *et al.* 2013; Lee *et al.* 2013; Brown & Jumpponen 2014; Hao *et al.* 2016; Louca *et al.* 2016a). Statistical methods, including null-deviation models (Ferrenberg *et al.* 2013; Brown & Jumpponen 2014; Dini-Andreote *et al.* 2015) or variance partitioning (Dumbrell *et al.* 2010; Hanson *et al.* 2012; Langenheder *et al.* 2012), are then used to infer deterministic and stochastic

processes. Still, these studies have been limited to making comparisons on the relative importance of stochastic and deterministic processes. Furthermore, since they often take place across large environmental gradients where variation is likely high, studies may not measure key abiotic and biotic (deterministic) factors driving community composition (Bell 2010). Lastly, they do not quantify measurement error, thereby likely inflating measures of stochasticity (Evans *et al.* 2017).

Thus, my third and fourth dissertation chapters shift from focusing on patterns in the distribution of microbial traits, to using an experimental approach to delve into the local processes driving microbial community composition and functioning. For my third chapter I conducted a field experiment where I manipulated a bacterial community's dispersal rate and tracked responses in the community weekly over five months. In this study I found that bacterial dispersal, like selection by the litter substrate, contributes to the diversity and composition of the bacterial community on grassland litter (Chapter 3; Albright & Martiny, *In press ISME*). My fourth chapter is the first field experiment that directly attempts to disentangle the role of stochastic processes, environmental selection, and dispersal on microbial community composition and functioning. I further provide the first estimate of stochastic variation in microbial community composition and functioning that accounts for measurement error (Chapter 4). Overall, insight gained from mapping current patterns in microbial community distributions and understanding the processes behind those patterns will improve predictions as to how ecosystems will look in the future, for example in response to global change.

CHAPTER 1

Nitrogen cycling potential of a grassland litter microbial community

ABSTRACT

Because microorganisms have different abilities to utilize nitrogen (N) through various assimilatory and dissimilatory pathways, microbial composition and diversity likely influences N cycling in an ecosystem. Terrestrial plant litter decomposition is often limited by N availability; however, little is known about the microorganisms involved in litter N cycling. In this study, we used metagenomics to characterize potential N utilization of microbial communities in grassland plant litter. The overall frequency of sequences associated with eight N cycling pathways differed by several orders of magnitude. Within a pathway, the distribution of these sequences among bacterial orders varied greatly. Many orders within the Actinobacteria and Proteobacteria appeared to be N cycling generalists, carrying genes from most (5 or 6) of the pathways. In contrast, orders from the Bacteriodetes were more specialized and carried genes for fewer (2 or 3) pathways. We also investigated how the abundance and composition of microbial N cycling genes varied over time and in response to two global change manipulations (drought and N addition). For many pathways, the abundance and composition of N cycling taxa varied over time, apparently reflecting precipitation patterns. In contrast to temporal variability, simulated global change had minor effects on N cycling potential. Overall, this study provides a blueprint for the genetic potential of N cycle processes in plant litter and a baseline for comparisons to other ecosystems.

INTRODUCTION

Microorganims play a key role in the decomposition of terrestrial plant litter(Raich & Schlesinger 1992; Aerts 1997; Chapin *et al.* 2002), a process that controls the balance of plant carbon (C) released in the atmosphere as CO_2 versus stored in the soil. Less often considered is the role that litter microorganisms play in nitrogen (N) cycling. N available to microorganisms degrading plant litter comes primarily from several sources. One source is organic N bound in plant tissues and microorganisms. Because the average C:N ratio is much higher in plant litter than in microbial decomposers, N availability is thought to limit litter decomposition (Paul & Clark 1996; Cleveland & Liptzin 2007; Mouginot *et al.* 2014). Fungal hyphae can further translocate N from the soil into plant litter (Boberg *et al.* 2010). And in some ecosystems, atmospheric deposition of inorganic N from human-driven NO_x emissions can also be an important source (Boonpragob & Nash 1990; Fenn & Bytnerowicz 1997; Schimel & Bennett 2004).

Microbes can rapidly alter the forms of N in plant litter through a variety of different N cycle pathways, and these changes in N availability can feedback to influence overall ecosystem functioning (Schimel & Hattenschwiler 2007; Miki *et al.* 2010). During decomposition, bacteria utilize N in both assimilatory and dissimilatory pathways. Assimilatory pathways require energy and lead to the conversion of inorganic N to organic N in microbes (e.g., utilizing N for protein, nucleic acid, and cellular component assembly). Dissimilatory pathways use N compounds to provide energy to microbes. Thus the pathways by which microbes use N impact the fate of N in the ecosystem and specifically, whether it is converted into microbial biomass or converted to new forms and released into the environment. For example, through the ammonia assimilation

pathway, organic N in plant litter may be used by microorganisms for growth (Geisseler *et al.* 2010). Through the denitrification pathway, N may be removed from the plant litter-soil system and lost into the atmosphere as N_2O or N_2 (Brown *et al.* 2012)

Environmental conditions also influence plant litter decomposition and therefore N cycling (Aerts 1997; Gholz *et al.* 2000; Knorr *et al.* 2005; Parton & Silver 2007; Petersen *et al.* 2012). In particular, moisture availability is known to be important to plant litter decomposition rates (Moore 1986; Cisneros-Dozal *et al.* 2007). Climate models predict decreased precipitation in the southwestern U.S. in the next century (Seager *et al.* 2007), a change that may also alter decomposition indirectly via changes in plant composition and litter quality in grasslands ecosystems (Potts *et al.* 2012; Allison *et al.* 2013). In addition, N availability also plays a role in plant litter decomposition (Frey *et al.* 2004; Knorr *et al.* 2005). N loading from anthropogenic sources (in Southern CA estimated between 20-45 kg ha^{-1} yr^{-1}) is expected to continue to increase (Fenn *et al.* 2003) and to affect plant communities and ecosystem functioning (Allen *et al.* 1998; Egerton-Warburton & Allen 2000; Fenn *et al.* 2010).

To investigate the effect of such changes on grasslands, an experiment manipulating nitrogen and precipitation was established in 2007 at Loma Ridge in Irvine, CA (Potts *et al.* 2012; Kimball *et al.* 2014). Previous work indicates that both drought and added N, as well as seasonal and annual climate variation, affects litter microbial composition (Berlemont *et al.* 2014; Matulich *et al.* 2015). Moreover, these shifts in microbial composition have functional consequences (Matulich & Martiny 2015). A reciprocal transplant experiment demonstrated that microbial communities altered by drought had lower rates of plant litter decomposition even under ambient environmental

conditions (Allison *et al.* 2013). Further, taxonomic changes in the litter community were correlated with changes in frequency of glycoside hydrolases, genes responsible for C utilization (Berlemont *et al.* 2014).

Given the intertwined nature of N and C cycling during litter decomposition, we investigated the genetic potential of N utilization in plant litter microbial communities. We analyzed metagenomic samples to identify genes for N cycling in microbial communities from the Loma Ridge experimental plots. We focused our work on prokaryotes because they are the most abundant organisms on the litter (Alster *et al.* 2013), but we also quantified sequences associated with Fungi. Although metagenomic sequences only indicate the functional potential of a community, they provide a holistic description of potential N utilization across many pathways in the N cycle. Specifically, we asked: (1) How does the abundance, taxonomic composition, and diversity of N cycling genes vary among pathways? (2) Within a pathway, do these patterns vary over time? and (3) Does N cycling potential respond to global change manipulations (drought and N addition)?

METHODS

Field Experiment and DNA sequencing

The Loma Ridge global change experiment, Irvine, CA, USA (33°44'N, 117°42'E, 365 m elevation) was established in 2007 with precipitation and N manipulations (Allison *et al.* 2013; Berlemont *et al.* 2014). The precipitation manipulation reduced water by 50%, creating drought plots. Surface soil moisture was significantly lower in drought plots than ambient treatment plots (Matulich *et al.* 2015). The nitrogen addition

plots received 60 kg $CaNO_3$ ha-1 yr-1. Previous studies at the site have shown that litter from nitrogen addition plots contained significantly more nitrogen and lower concentrations of carbon substrates, such as cellulose, hemicellulose, and lignin, than control plots (Allison *et al.* 2013). And furthermore plant-available nitrogen in soil was significantly lower in drought plots, than in control or nitrogen plots (Potts *et al.* 2012). Compared to control plots drought treatments reduced and nitrogen addition plots increased decomposition rates, as measured by mass loss (Allison *et al.* 2013).

The climate in this southern California grassland ecosystem is semi-arid, with mean annual precipitation of 325 mm, most of which occurs between October and April. Beginning in 2010, plant litter samples were taken seasonally for two years in control, drought, and N addition plots. Sampling dates were April 14th, August 20th, and December 17th 2010; February 29th, June 10th, September 21st, and December 14th 2011; and March 12th 2012.

For each of the three treatments (ambient, drought, and N addition), eight plots were sampled at 8 times points (for a total of 192 samples). To balance replication with sequencing costs, we pooled equal concentrations of DNA extracts from 4 plots undergoing the same treatment, where the same plots were pooled on each date (Berlemont *et al.* 2014). Thus, six metagenomic libraries (two replicate libraries per treatment) were sequenced at eight time points for a total of 48 libraries. Although two replicates per treatment is not ideal and limits our statistical power to test for treatment effects, pooling four independent plots into two composite “replicates” provides improved mean estimates (such as the mean abundance of a particular gene) compared to sampling only two plots (Brumelle *et al.* 1984). Further, sampling over 8

time points also gives us additional statistical power to test for treatment effects by reducing the error variance (Sokal & Rohlf 2012).

Metagenomic libraries were prepared using a Truseq(r) library kit (Illumina, Part #15026484 Rev. C - July 2012), sequenced with an Illumina HiSeq2000 (100bp-paired ends), and yielded 107.4 Gbp (passed QC) (Berlemont *et al.* 2014). In total, 46 libraries were analyzed with two libraries excluded due to low sequence counts (April 2010 reduced precipitation (4511045); August 2010 increased N deposition (4511064) (Table S1)). Sequences were uploaded onto the MG-RAST server, where 53% of the sequences were annotated (Berlemont *et al.* 2014).

Taxonomic assignment for the metagenomic libraries was performed by MG-RAST (Meyer *et al.* 2008) using the Kegg database (Kanehisa *et al.* 2010) and downloaded using the MG-RAST API version 3.2 (Meyer *et al.* 2008). Taxonomic annotation was considered for sequences with an e-value of less than or equal to 10^{-5} (Berlemont *et al.* 2014). Each sequence was assigned to the closest related species in the database; however, to be conservative in our taxonomic assignment, we report bacterial taxonomy at the corresponding order level.

N-cycle Pathway Identification

Eight N cycling pathways were defined for this analysis: nitrification (number of genes targeted: n=2), N fixation (n=20), denitrification (n=20), dissimilatory nitrate to nitrite reduction (DNRA($NO_3^- \rightarrow NO_2^-$)) (n=9), dissimilatory nitrite to ammonia reduction (DNRA($NO_2^- \rightarrow NH_4^+$)) (n=4), assimilatory nitrate to nitrite reduction (ANR($NO_3^- \rightarrow NO_2^-$)) (n=2), assimilatory nitrite to ammonia reduction (ANR($NO_2^- \rightarrow NH_4^+$)) (n=2), and

(ammonia assimilation (n=10) (Figure 1.1, Table S2). Although nitrification includes both ammonia oxidation and nitrite oxidation, we combined them here because of their low representation in the samples. Finally, we excluded the anammox pathway from our analyses. The functional genes for this pathway are poorly represented in genome databases (and currently not defined in the KEGG or SEED databases). After some analysis, we were not confident in our ability to distinguish between genes in the annamox pathway and related, non-annamox genes.

To detect genes in the eight pathways, we first identified the corresponding genes in the Kegg (Kanehisa *et al.* 2010) and SEED (Overbeek *et al.* 2005) databases. For the Kegg database, Kegg orthology (KO) numbers (Kanehisa *et al.* 2010) were obtained from the Functional Ontology Assignment for Metagenomes (FOAM) database (Prestat *et al.* 2014). For the SEED database, figfam numbers (FIG) were obtained for the N fixation and denitrification pathways. Next, MD5 IDs for each KO and FIG number sequence were retrieved from the non-redundant M5nr database (Wilke *et al.* 2012). Finally, we searched for the MD5 IDs in our samples annotated by the MG-RAST server. For each pathway, we checked the functional assignments of a subset of sequences using the BLAST algorithm against the MicrobesOnline database. This allowed us to compare the annotations and genome context of those hits in fully sequenced genomes (Dehal *et al.* 2010).

Data Standardization and Statistical Analyses

We first compared the relative abundance of prokaryotic and fungal reads across the different N cycle pathways. To do this, we took the average number of sequences

associated with each pathway across all 46 samples, then divided by the number of different genes searched in the pathway. While we recognize it is common to also standardize by gene length (e.g., (Batmalle *et al.* 2014)), the variation in copy number and gene length for all 80 genes across a wide range of microbial taxa makes this infeasible. Conveniently, our results show that most pathway average abundances differ in orders of magnitude, while gene lengths usually vary to a lesser degree. This suggests that the observed relative differences would likely persist even with such standardization.

All statistical analyses were performed using the 'nlme' and 'vegan' packages in the R software environment (Team 2011; Oksanen *et al.* 2013). To test for differences in abundance of bacterial communities across treatments and sampling date, we used one-way repeated measures ANOVA with plot included as an error term. For this analysis, we standardized the number of read numbers associated with each pathway by the total number of annotated bacterial reads in that library. We standardized fungal read abundances for each sample by the total fungal reads in each library.

Taxonomic diversity of the genes associated with each pathway was quantified using the Shannon evenness index and observed richness of the number of orders. To test whether evenness changed over time, we performed one-way repeated measures ANOVA as described previously.

To assess the effects of treatment and time on bacterial community composition, we performed a permutational multivariate analysis of variance (PERMANOVA) including treatment and sampling date as fixed effects ((Clarke & Warwick 2001; Anderson *et al.* 2008). Taxa were first standardized by their relative frequency within

each pathway, then the analysis was run for each pathway. The analyses were run using partial sums of squares on 999 permutations of residuals under a reduced model. If the model returned non-significant variables, the variables were removed, and the model was tested again. This procedure does not alter the significance of the remaining variables, but reduces the effect of spurious relationships between variables (Harell 2001). To identify the taxa contributing to significant compositional differences, we used similarity percentages analysis (SIMPER) (Clarke & Gorley 2006). Specifically, we tested which taxa accounted for differences between the rainy (winter/spring) versus dry (summer/fall) samples. Lastly, Pearson's correlations were used to test whether the number of sequences attributed to an N cycling pathway were distributed among bacterial orders in proportion to their total abundance in the metagenomes.

Extracellular Enzyme Activity

To assay the functional potential of the litter microbial community, we measured the potential activities of nine extracellular enzymes including, A-Glucosidase (AG), acid phosphatase (AP), B-Glucosidase (BG), B-Xylosidase (BX), cellobiohydrolase (CBH), leucine aminopeptidase (LAP), N-acetyl-B-D-glucosaminidase (NAG), polyphenol oxidase (PPO), and peroxidase (PER), on litter from all treatments at seven sample dates between September 2011 and March 2013. Fluorimetric and oxidative enzyme assays were conducted as described in (Alster *et al.* 2013) and initial results from these analyses are reported in (Matulich *et al.* 2015). For this study we used the same data to calculate the ratio of C and N acquiring enzymes (NAG:C_{enz}) (Stone 2014). The ratio of these two metrics has been proposed to estimate relative allocation to energy versus

nutrient acquisition (Sinsabaugh *et al.* 2009; Stone 2014). The extracellular enzyme N-acetyl-β-D-glucosaminidase (NAG) measures potential chitinase activity, a proxy for the conversion of organic N to ammonium for assimilation. C_{enz} is defined as the sum of four extracellular enzymes involved in C cycling (AG BG, CBH, BX). We used a repeated measures ANOVA to test for differences in NAG:C_{enz} ratios across all treatments and enzyme sampling dates.

To test for correlations between potential enzyme activities and genomic potential, we examined data from three sampling dates when both types of data were collected (September 2011, December 2011, March 2012). Specifically, we tested whether the NAG:Cenz ratio could be predicted by the ratio of the abundance of Assimilatory: Dissimilatory (A:D) N cycling genes, as a genomic index for allocation of energy versus nutrient acquisition. We used an ANCOVA with NAG:C_{enz} activity as the independent variable, the A:D ratio as the dependent variable, and time as a covariate. (We excluded treatment as a covariate because treatment did not significantly affect the NAG:C_{enz} ratio in the test described above.)

RESULTS

Across all the plant litter metagenomic libraries, 59% of the annotated sequences were bacterial (294,674,419 reads). Of the annotated sequences, 0.31% were associated with an N cycling pathway, and of these, 896,943, 197,944, and 3,278 were assigned to Bacteria, Fungi, and Archaea, respectively. The vast majority of these sequences were associated with ammonia assimilation (84%, 75%, and 98%, from

Bacteria, Fungi, and Archaea, respectively). All N cycling fungal sequences were associated with two phyla, Ascomycota (94.5%) and Basidiomycota (5.5%).

Abundance

The total abundance of prokaryotic reads related to different N cycling pathways differed by several orders of magnitude (Figure 1.2). Broadly, assimilatory pathways were much more prevalent (96.5%) than dissimilatory pathways (3.5%). After ammonia assimilation, ANR($NO_3^- \rightarrow NH_4^+$) and DNRA($NO_3^- \rightarrow NO_2^-$) were the next most detected pathways, while the nitrification pathway was the least detected (Table 1.1). All fungal N cycling sequences were associated with assimilatory pathways (Figure 1.2).

The frequency of prokaryotic genes in each pathway varied significantly over time for 5 of the 8 pathways: ammonia assimilation, ANR($NO_2^- \rightarrow NH_4^+$), DNRA($NO_3^- \rightarrow NO_2^-$), N fixation, and denitrification pathways (repeated measures ANOVA; Table 1, Figure 3a). Gene abundances in these pathways tended to covary over time and were lowest in August 2010 and June 2011. This pattern correlated with cumulative precipitation at the site in the two weeks prior to sampling (Figure 1.3b). The only fungal pathway that varied significantly over time was that associated with ANR($NO_2^- \rightarrow NH_4^+$). In contrast to prokaryotic sequences from this pathway, the frequency of fungal sequences was highest in August 2010 and June 2011 (Figure 1.3a).

Composition

The distribution of N cycling potential among prokaryotic taxa differed distinctly by pathway (comparing the columns in Figure 1.4). Genes involved in ammonia

assimilation, were generally detected in proportion to each bacterial order's abundance (R^2 = 0.877, P<.0001), as were genes involved in ANR($NO_2^- \rightarrow NH_4^+$) and DNRA ($NO_3^- \rightarrow NO_2^-$) ($R^2$ = 0.714, 0.657, P<.0001). ANR($NO_3^- \rightarrow NO_2^-$) was the most taxonomically restricted pathway. Even though genes for this process were relatively abundant in the metagenomic libraries (Figure 1.2), they were detected for only five bacterial orders and thus were not well correlated with total abundance of each order (R^2 = 0.459, P<.0001). Finally, genes for DNRA($NO_2^- \rightarrow NH_4^+$), were rare among abundant orders, but common among some rare orders and hence were poorly correlated with overall abundance (R^2 = 0.084, P>0.05; Figure 1.4).

In an intermediate case, denitrification genes were common amongst many abundant bacterial orders (e.g., Rhizobiales, Burkholderiales, and Actinomycetales), but were also found in other less abundant orders including Archaea (Halobacteriales, Methanosarcinales, and Cenarchaeales) and known ammonia-oxidizers (Nitrosomonadales and Nitrosopumilales) (R^2= 0.539, P <.0001, Figure 1.4). This was also true for N fixation (R^2= 0.497, P<.0001), which was common in some the most abundant taxa (e.g., Rhizobiales and Burkholderiales), but absent in the most abundant order (Actinomycetales).

Distinct bacterial taxa appeared to have different N cycling potential in these plant litter communities (comparing rows in Figure 1.4). The most abundant taxa (e.g., Actinomycetales, Rhizobiales, Burkholderiales, and Sphingomonadales) appeared to be N cycling generalists in that they carried genes from most (5-6) of the seven pathways. Other taxa seemed to be more specialized. Notably, the orders from phylum Bacteriodetes (Bacteriodales, Cytophagales, Flavobacteriales, and Sphingobacteriales)

carried genes for only a couple (2-3) pathways, even thought the taxa were relatively abundant (Figure 1.4).

Beyond these average trends, the composition of potential N cycling litter prokaryotes varied among litter samples. Much of this variation could be attributed to temporal variability (Figure S1). In 5 of the 8 examined pathways, time explained between 12% and 45% of the compositional variation among samples (Figure 1.5; Table S3). These trends were largely driven by seasonal differences in a few abundant bacterial orders. For instance, across all N pathways, the relative abundance of Burkholderiales and Sphingomonadales was higher, and Enterobacteriales lower, in the rainy (winter/spring) versus dry (summer/fall) seasons (SIMPER analysis; Table S4). However, the seasonal abundance of at least two orders depended on the N pathway examined. Actinomycetales ANR($NO_2^- \rightarrow NH_4^+$) genes were relatively abundant in winter/spring, but Actinomycetales ANR($NO_3^- \rightarrow NO_2^-$), DNRA($NO_3^- \rightarrow NO_2^-$), and ammonia assimilation genes were higher in summer/fall (Table S4).

In contrast to time, treatment (drought or N addition) had minor effects on the composition of N cycling prokaryotes, explaining only a small percentage of variation in the DNRA($NO_3^- \rightarrow NO_2^-$) (2.9%) and ANR($NO_2^- \rightarrow NH_4^+$) (4.4%) pathways (Figure 1.5). This result was similar to analyses of community composition using 16S, where treatment explained ~3% of estimated variations (Matulich *et al.* 2015). Treatment also interacted with sampling date to account for 11% to 15% of compositional variation in the denitrification, DNRA($NO_3^- \rightarrow NO_2^-$), and ANR($NO_2^- \rightarrow NH_4^+$) pathways (Fig. 5). For instance, N addition altered the bacterial orders carrying ANR($NO_2^- \rightarrow NH_4^+$) pathway genes during the dry season, but not the rainy season (Figure S1).

Diversity

Prokaryotic evenness at the order level was relatively constant over time for all pathways examined (Figure 1.6). The richness and evenness of prokaryotic orders associated with each N pathway were highly correlated; pathways encoded by a higher number of orders also tended to be more evenly distributed across those orders. Generally, evenness was low for the communities with the potential for N fixation, ANR($NO_3^- \rightarrow NO_2$), and DNRA($NO_2^- \rightarrow NH_4^+$) (Figure 1.6). In the case of N fixation and DNRA($NO_2^- \rightarrow NH_4^+$), this reduced diversity may be due to under sampling, as the total number of sequences detected was in the hundreds (Table 1.1). As mentioned above, however, only 5 bacterial orders appeared to carry genes for ANR($NO_3^- \rightarrow NO_2^-$) even though thousands of sequences were sampled. Most (81%) of these genes were attributed to one order, the Actinomycetales.

Extracellular Enzymes

Like the abundance of N cycling genes, the potential activity of nine extracellular enzymes varied over time, but not by treatment in this leaf litter system (Matulich *et al.* 2015). In particular, the ratio of C and N acquiring enzymes (NAG:C_{enz}) varied by month (repeated measures ANOVA; $p<0.0001$), with the lowest NAG:C_{enz} ratios in the dry fall months (September 2012 and 2013). Further, the ratio of A:D gene abundance explained almost half of the variation in NAG:C_{enz} (ANCOVA; $p=0.045$; adjusted $R^2=0.47$)(Fig S2).

DISCUSSION

Our analysis provides a blueprint for the genetic potential of N cycle processes in plant litter without the biases associated with targeting individual genes or a subset of microbial communities. Of course, there are a number of caveats to interpreting metagenomic data (Prosser 2015). Most importantly, for many pathways, gene abundance in a community has not been found to consistently correlate with environmental process rates, which limits our ability to draw conclusions about activity (Gubry-Rangin *et al.* 2010; Attard *et al.* 2011; Graham *et al.* 2014). Still, a powerful feature of metagenomics data is the ability to compare across many functional pathways and taxa at the same time. Indeed, we found that the abundance, composition, and diversity of N cycling genes differed greatly among the eight targeted pathways. Aggregating across pathways, prokaryotes and fungi appear to play an equally important role in N assimilation in this system. Assimilatory pathways were much more prevalent than dissimilatory pathways. And orders within the Actinobacteria and Proteobacteria appeared to be N cycling generalists, carrying genes from most pathways relative to those from the Bacteriodetes, despite also being relatively abundant in the litter.

The metagenomic picture of N cycling potential was further correlated with the functional potential of the community, as assayed by extracellular enzyme activities. Specifically, the ratio of assimilatory to dissimilatory gene abundance explained almost half of the variation in the ratio of C and N acquiring enzymes ($NAG:C_{enz}$), an index of the relative allocation to energy versus nutrient acquisition (Sinsabaugh *et al.* 2009;

Stone 2014). This relationship indicates that the metagenomic variation observed may have direct functional relevance for the leaf litter community.

Comparisons across N cycling pathways

The three most abundant N cycling pathways (i.e., ammonia assimilation, ANR($NO_2^- \rightarrow NH_4^+$), and DNRA($NO_3^- \rightarrow NO_2^-$)) were associated with bacterial orders in proportion to their total abundance, indicating that the pathways are broadly distributed across the Bacteria. For instance, ammonia assimilation is an ability shared by virtually all microorganisms made up the majority of N cycling sequences, similar to that found in other metagenomic studies across various environments (Varin *et al.* 2010; Yu & Zhang 2012; Quinn *et al.* 2014).

The high abundance of ANR pathways suggests that NO_3^- and NO_2^- may be important sources of N on the plant litter. This availability may be due in part to high rates of atmospheric N deposition that occur in the southern California region (Fenn & Bytnerowicz 1997), although microbial ANR appears to be an important process in a wide range of terrestrial systems including undisturbed soils and agricultural fields (Burger & Jackson 2003; Booth *et al.* 2005; Myrold & Posavatz 2007). In terrestrial systems, heterotrophic bacteria seem to prefer ammonium (NH_4^+) over NO_3^- /NO_2^- due to energetic costs (Rice & Tiedje 1989; Recous *et al.* 1990), unless they are limited by NH_4^+, in which case they may also use NO_3^- /NO_2^- (Nishio *et al.* 2001; Booth *et al.* 2005). Thus, the ability to assimilate NO_3^- /NO_2^- may provide an advantage in an N-limited ecosystem (Rediers *et al.* 2009).

The DNRA($NO_3^- \rightarrow NO_2^-$) pathway was also abundant in the plant litter. Relatively little is known about DNRA in terrestrial systems (Rutting *et al.* 2011). While the DNRA pathway in soil bacteria was discovered in the 1930's (Woods), until recently many studies considered denitrification as the only dissimilatory reduction process in soil (Cole 1990). Indeed, DNRA and denitrification are competitive processes, which occur primarily under anaerobic conditions (Rutting *et al.* 2011). DNRA is now recognized as a key process in wetlands and has been observed in moist tropical soils (Silver *et al.* 2001). Modelling studies suggest that DNRA may be important in temperate grassland soils (Muller *et al.* 2004; Muller *et al.* 2007), but its general significance in aerobic upland soils remains unclear (Butterbach-Bahl & Gundersen 2011).

There are several reasons why the plant litter environment may be a conducive for DNRA. First, while the process is most likely to occur under anaerobic conditions, some studies have shown that DNRA is less sensitive to variable O_2 and redox conditions than denitrification (Fazzolari *et al.* 1998; Pett-Ridge *et al.* 2006). Second, DNRA is thought to be favored in high C:N environments (Rutting *et al.* 2011), like that of leaf litter. Finally, oxygen gradients that range from 100 to 0% saturation within a few micrometers have been measured in plant litter layers (van der Lee *et al.* 1999; Reith *et al.* 2002). Thus, we speculate that plant litter could be suitable for DNRA, particularly after rains when the wet, matted down litter may contain anaerobic pockets. Indeed, the highest abundances of DNRA($NO_3^- \rightarrow NO_2^-$) sequences in the litter samples were correlated with increased precipitation at the site. In the future, it would be useful to verify the activity of the pathway at the site by transcriptomics.

While less abundant, denitrification and N fixation genes were also present in the leaf litter metagenomes. Denitrification is known to be an important process in terrestrial grasslands, where soils are a major source of N_2O emissions to the atmosphere (Brown *et al.* 2012; Bouwman *et al.* 2013). Denitrification in soils occurs primarily after precipitation events (Hartley & Schlesinger 2000; Nielsen & Ball 2014). Like DNRA, the denitrification pathway was detected across a wide range of bacterial orders, but it was most common in orders from the phylum Proteobacteria.

N fixation appears to be the rarest of the pathways targeted in the litter community, second only to nitrification. Although N fixation does occur during litter decomposition, estimated rates in the litter/soil layer are low, between <0.5 - 20 kg N ha^{-1} yr^{-1} (Cleveland *et al.* 1999), much smaller than symbiotic N fixation in various agricultural crops (Smil 1999). Like DNRA($NO_3^- \rightarrow NO_2^-$), the genetic potential for N fixation was distributed across a distinct group of bacterial orders, many of which were present at low abundance in the community. It is often assumed that Rhizobiales carry out most N fixation in soils (Resendis-Antonio *et al.* 2011; Wang *et al.* 2013); while 18% of all N fixation genes were classified as Rhizobiales, 82% were from other orders, including Pseudomonadales, Enterobacteriales, and Burkholderiales.

Temporal variation and sensitivity to global change manipulations

In addition to the broad patterns of prokaryotic diversity supporting N cycling in plant litter, we investigated how these communities varied across the seasons and in response to drought and N addition treatments. Within pathways, gene abundance and taxonomic composition varied over time, but little in response to the global change

manipulations. Specifically, the manipulations did not impact N-cycle pathway abundance and only altered composition within the DNRA($NO_3^- \rightarrow NO_2$) and ANR($NO_2^- \rightarrow NH_4^+$) pathways. This limited response of the N cycling pathways is somewhat surprising considering that N content (%N) in the leaf litter increased significantly in the nitrogen addition plots (Allison *et al.* 2013). However, we may have missed some responses because of low statistical power and/or our focus on primarily inorganic pathways.

Previous metagenomic studies have detected changes in potential N cycling in response to environmental perturbations such as N addition (e.g.,(Thurber *et al.* 2009; Fierer *et al.* 2012a)); however, many of these studies consider the N cycle quite broadly, making direct comparisons difficult. A few studies highlight mixed responses by pathway (Luo *et al.* 2014; Mason *et al.* 2014; Cobo-Díaz *et al.* 2015); for instance, burning tended to increase the relative abundance of dissimilatory processes and decrease that on assimilatory processes (Tas *et al.* 2014).

In contrast to the response to global change manipulations, N cycling pathway abundance showed significant temporal variability. Other global change experiments have also observed this pattern of strong temporal versus treatment effects on microbial communities (Cruz-Martinez *et al.* 2009; Yuste *et al.* 2011; Cregger *et al.* 2012; Gutknecht *et al.* 2012). Such a result is perhaps not unexpected; annual mean precipitation at the Loma Ridge site is 30 cm, and almost all of this rainfall falls between November and April (Kimball *et al.* 2014). As a result, microbial biomass on plant litter is reduced to less than 25% of peak levels during the summer dry season (Allison *et al.* 2013). Indeed, the abundance of some N-cycling pathways followed this broad trend. In

these cases, the frequency of the pathway (relative to all prokaryotes) was stable over time (no significant time effect), indicating there was no differential selection for the pathway across seasons above and beyond the fluctuations in total prokaryotic abundance.

However, for the majority of the N pathways, the frequency of gene abundance varied significantly over time, generally increasing during the wetter months. Corresponding with this trend, the NAG:C_{enz} enzyme activity ratio was lowest during the driest months. Thus, in wet conditions, selection may favor traits allowing for rapid N assimilation, whereas dry conditions may select for drought tolerance traits associated with different taxa (Schimel *et al.* 2007).

Similar to the overall taxonomic composition of the litter community (Matulich *et al.* 2015), the diversity and composition of bacterial lineages involved in most N-cycling pathways also varied across seasons. Compositional differences were driven by changes in the relative abundances of the most abundant taxa in the system. For example, some potential NO_2^- assimilators (e.g., Actinomycetales, Enterobacteriales, and Burkholderiales) displayed a seasonal signal whereas others (e.g., Cytophagales and Rhodocyclales) were less affected by seasons. This pattern suggests that this pathway itself may not be selected for by season but instead may be linked to other traits having a distinct distribution in bacterial lineages.

Conclusions

This study provides an overview of microbial N cycling potential in a plant litter system and points to several directions for future research. In particular, the high

abundance of DNRA pathway genes is intriguing and suggests that further work on this process in grassland ecosystems is warranted. We also observed that the degree of N pathway specialization among bacterial orders tended to increase with its abundance in the plant litter, suggesting that N cycling generalists may have an advantage in plant litter. However, it is unclear whether this pattern is specific to this environment or may be a general feature of N limited environments. Indeed, the results described here will be most useful in direct comparison to other ecosystems.

ACKNOWLEDGMENTS

We thank Claudia Weihe for help with sampling and lab work, Steven Allison for general project leadership, Michael Goulden for running the global change experiment, and the J. Martiny lab and Katrine Whiteson for useful comments on earlier versions of the manuscript. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), under Award Number DE-PS02-09ER09-25.

Tables and Figures

Table 1.1 Effect of treatment, sample date (time), and their interaction on the abundance of sequences from each N cycle pathway for prokaryotes and fungi.

N Cycle Pathway	Significance (P value)[a] of effect of: Treatment	Time[b]	Interaction of treatment and time	Total no. of sequences detected
Prokaryotes				
Dissimilatory Nitrate to Nitrite	NS	<.0001	0.07	35,920
Dissimilatory Nitrite to Ammonia	NS	NS	NS	130
Denitrification	NS	<0.01	NS	1,846
Nitrification	--	--	--	17
Assimilatory Nitrate to Nitrite	NS	NS	NS	2,009
Assimilatory Nitrite to Ammonia	NS	<.0001	NS	102,360
Ammonia Assimilation	NS	<.0001	0.05	757,470
Nitrogen Fixation	NS	<0.01	0.02	413
Fungi				
Assimilatory Nitrite to Ammonia	NS	<0.01	0.05	32,172
Assimilatory Nitrate to Nitrite	NS	NS	0.06	17,169
Ammonia Assimilation	NS	NS	NS	148,602

[a] based on repeated measures ANOVA. NS, not significant; --, not tested.
[b] Sample date.

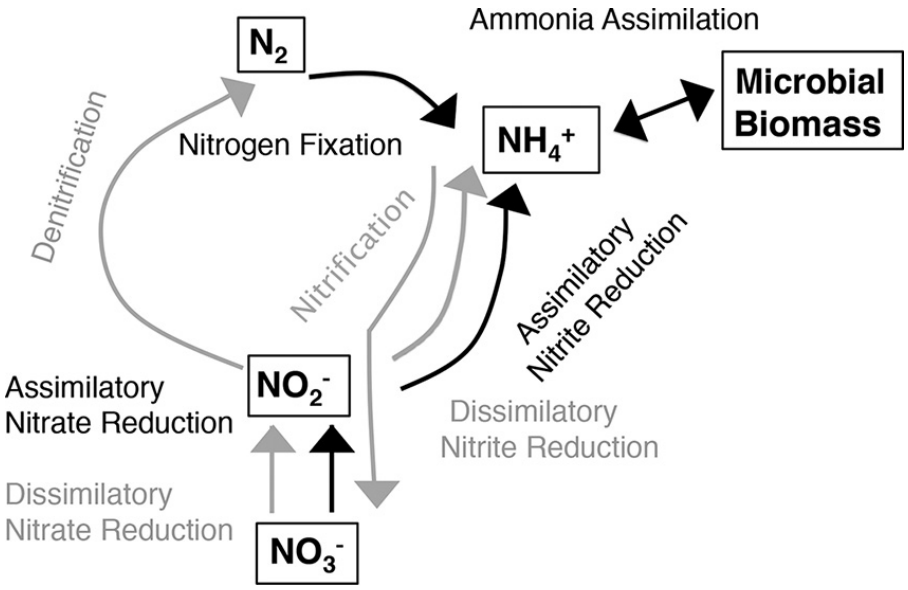


Figure 1.1 Nitrogen cycling pathways considered in this study. Pathways in gray are categorized as dissimilatory, and pathways in black are categorized as assimilatory.

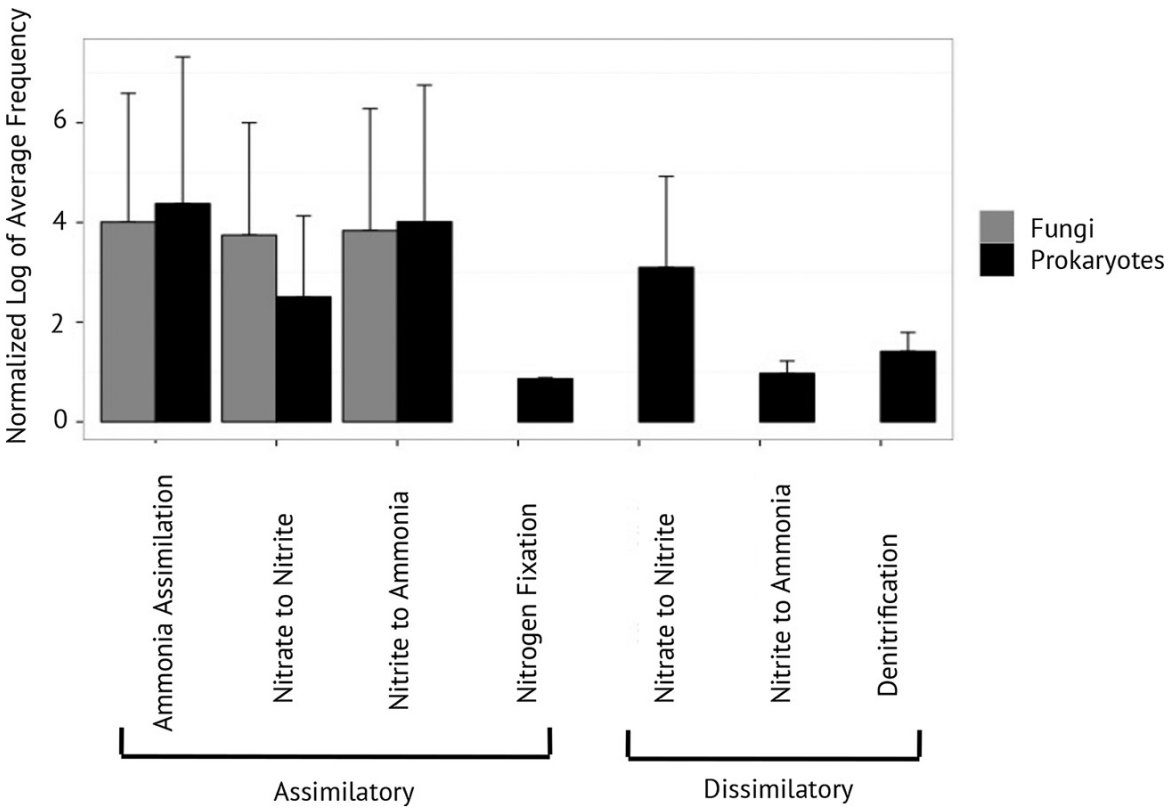


Figure 1.2 Log abundances of prokaryotic and fungal reads for each N cycle pathway. Abundance was calculated as the average across all samples (n=46), standardized by the number of genes targeted in the pathway (see Table S2).

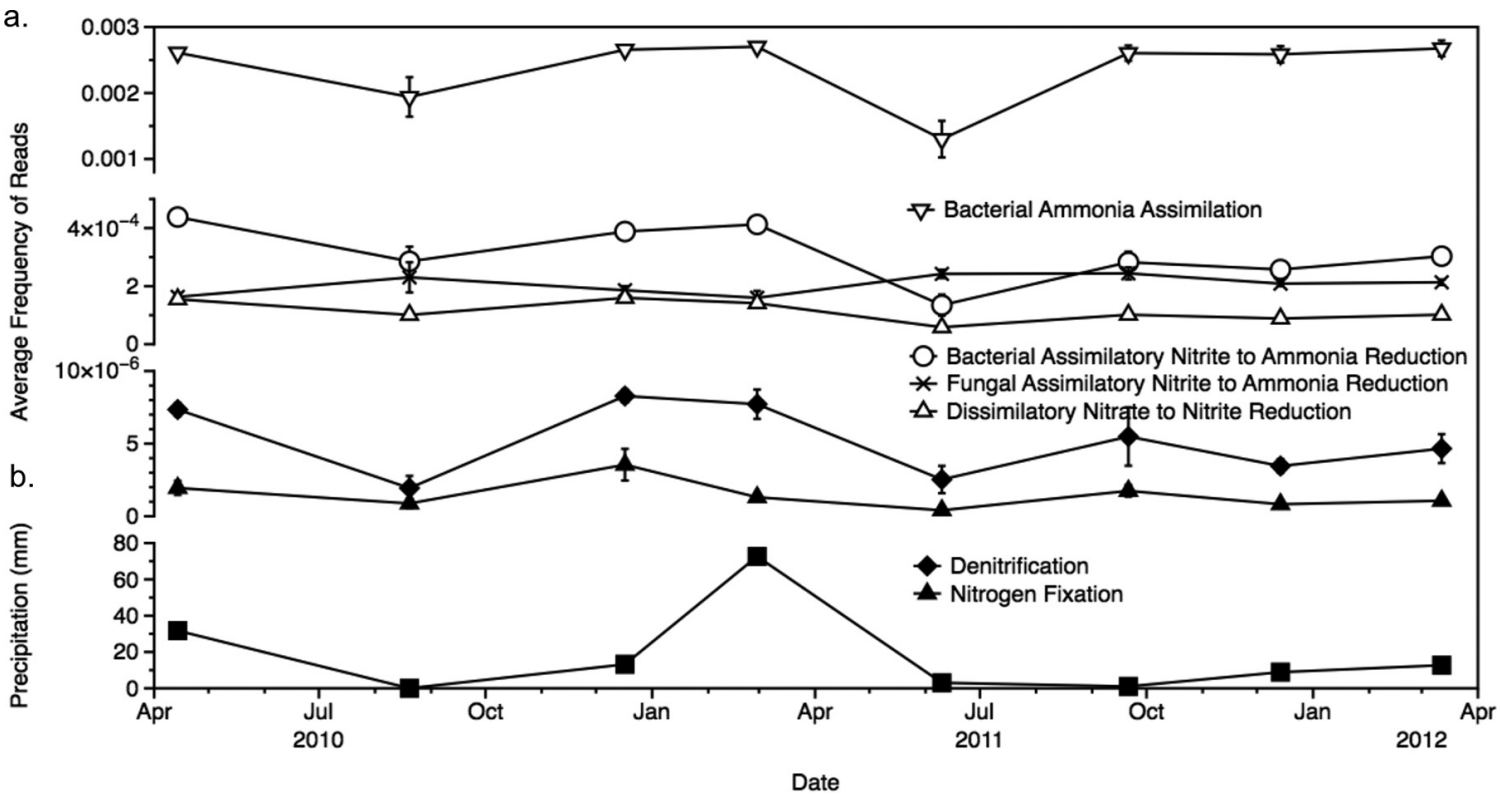


Figure 1.3 (a) Average frequencies of sequences by sample date for those pathways that varied significantly over time, including denitrification, nitrogen fixation, dissimilatory nitrate to nitrite, bacterial assimilatory nitrite to ammonia, fungal assimilatory nitrite to ammonia, and bacterial ammonia assimilation. Error bars are +/- 1 standard error. **(b)** Cumulative precipitation in mm two weeks prior to sampling by sampling date.

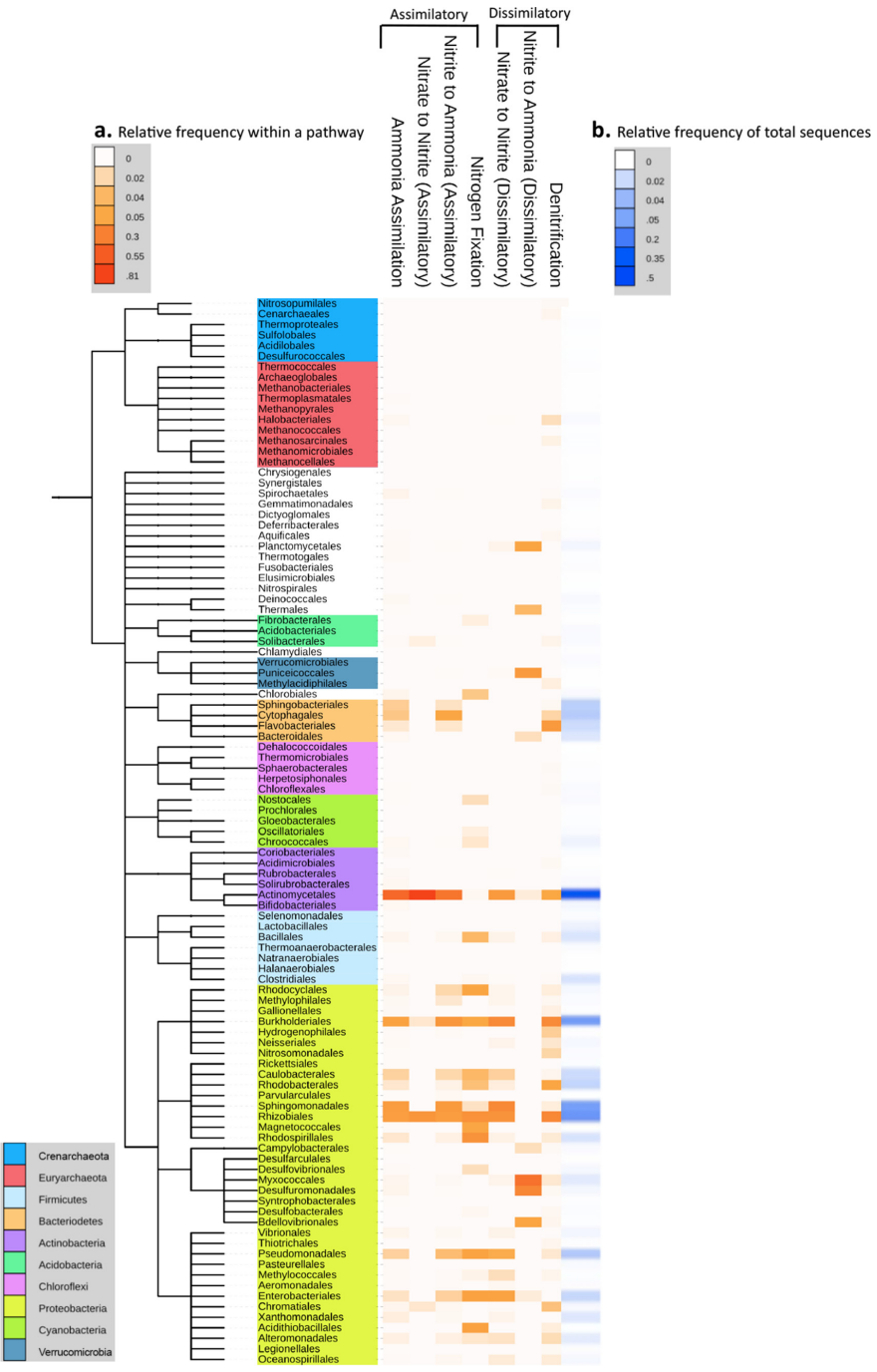


Figure 1.4 Bacterial and archaeal composition at the order level by N pathway. Data are combined across all sampling dates. A coarse phylogeny at left is shown to highlight the major phyla (iToL; (Letunic & Bork)). **(a)** The orange heat map represents the relative distribution of sequence reads by order for each pathway. **(b)** For comparison, the blue heat map shows the relative abundance of all bacterial and archaeal sequences (all predicted proteins and ribosomal RNA genes) by order. (Relative abundance was calculated with all bacterial/archaeal taxa, however only orders with predicted N cycle reads (101/130) are included in the figure).

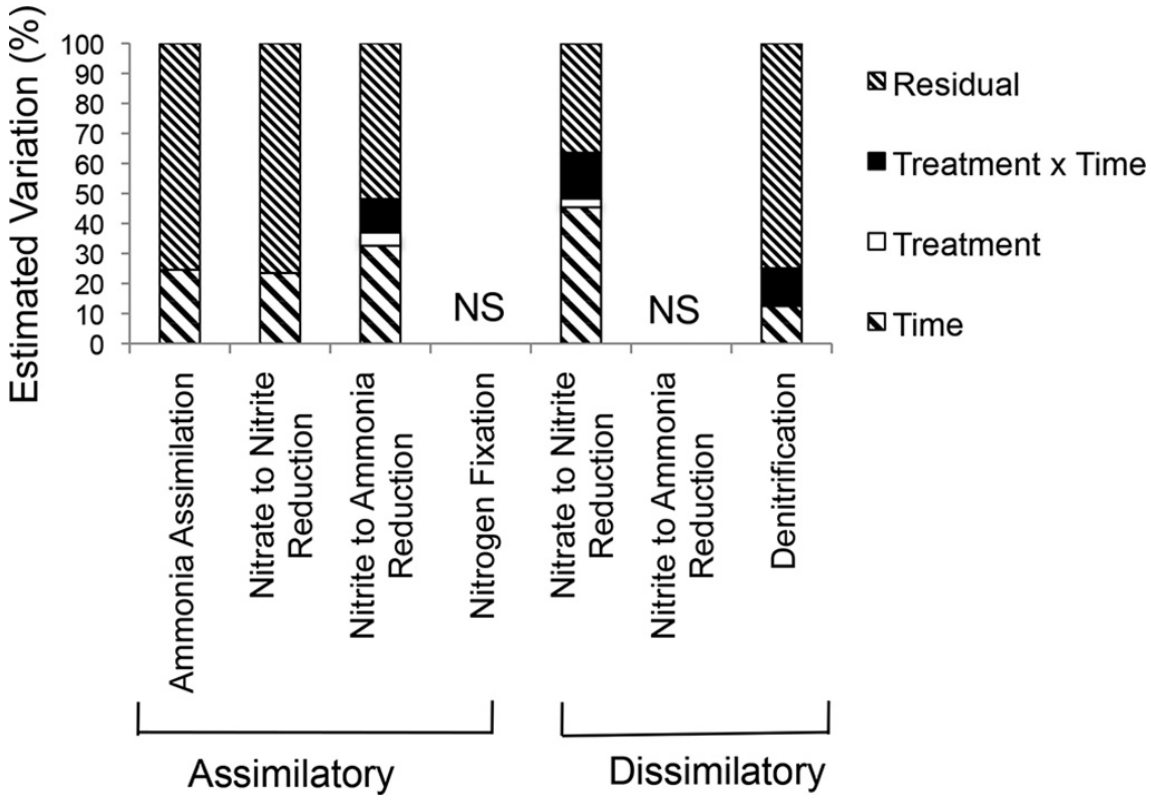


Figure 1.5 Percentages of the estimated variation in prokaryotic community composition explained by time, treatment and their interaction for each N cycle pathway. Estimates were derived from PERMANOVA models, where NS means the test result was not significant (P≥0.1).

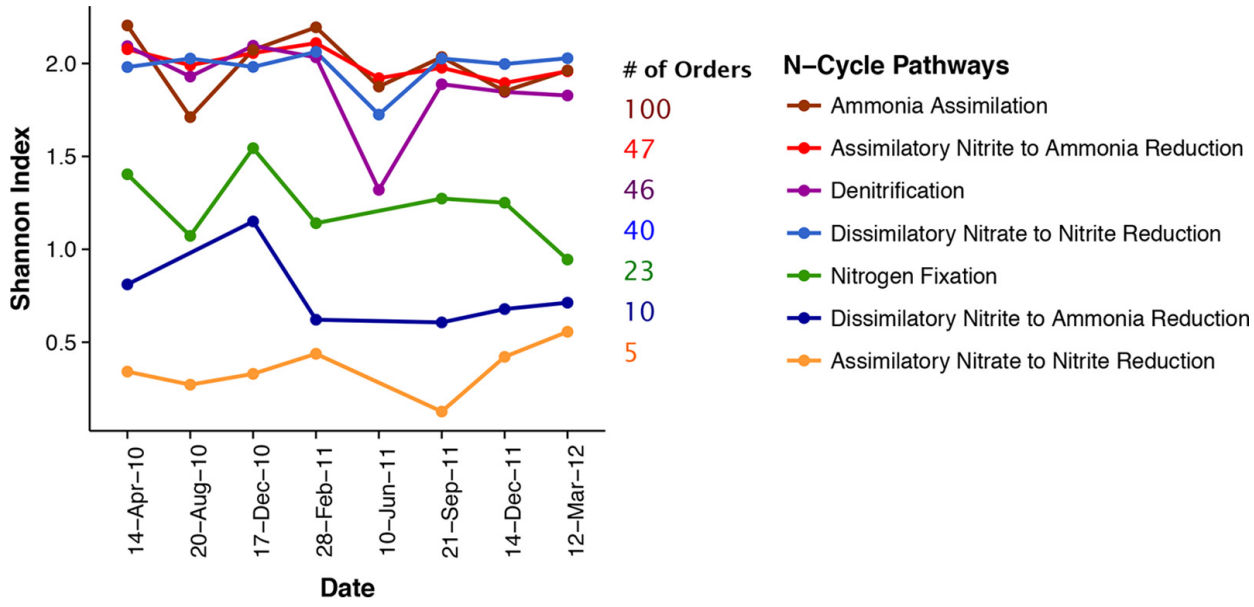


Figure 1.6 Diversity of orders (calculated by the Shannon index) for each N-cycling pathway by collection date. The total number of orders for each pathway is also noted.

Supporting Information

This chapter contains supporting information that can be found online at http://aem.asm.org/content/suppl/2015/09/21/AEM.02222-15.DCSupplemental/zam999116628so1.pdf (doi: 10.1128/AEM.02222-15).

CHAPTER 2

Global biogeography of microbial nitrogen cycling traits in soil

ABSTRACT

Microorganisms drive much of the Earth's nitrogen (N) cycle, but we still lack a global overview of the abundance and composition of the microorganisms carrying out soil N processes. To address this gap, we characterized the biogeography of microbial N traits, defined as eight N cycling pathways, using publically available soil metagenomes. The relative frequency of N pathways varied consistently across soils, such that the frequencies of the individual N pathways were positively correlated across the soil samples. Habitat type, soil carbon, and soil N largely explained the total N pathway frequency in a sample. In contrast, we could not identify major drivers of the taxonomic composition of the N functional groups. Further, the dominant genera encoding a pathway were generally similar among habitat types. The soil samples also revealed an unexpectedly high frequency of Bacteria carrying the pathways required for dissimilatory nitrate reduction to ammonium (DNRA), a little studied N process in soil. Finally, phylogenetic analysis showed that some microbial groups appear to be N-cycling specialists or generalists. For instance, taxa within the delta-Proteobacteria encoded all eight N pathways, whereas those within the Cyanobacteria primarily encoded three pathways. Overall, this trait-based approach provides a baseline for investigating the relationship between microbial diversity and N cycling across global soils.

SIGNFICANCE STATEMENT

Microbes are key players in the Earth's biogeochemical cycles, including the nitrogen (N) cycle. Despite their importance, however, we know little about the abundance and composition of microorganisms responsible for N transformations in soil. We present the first characterization of the global biogeography of soil N cycling microbes by investigating geographic patterns in their abundance and composition. Our analysis reveals the most prominent soil taxa harboring the genetic machinery for these processes. We also discover an unexpectedly high abundance of bacteria encoding dissimilatory nitrate reduction to ammonium, a little studied process in soil. In general, such a focus on the biogeography of microbial traits could improve efforts to connect biodiversity patterns and ecosystem processes.

INTRODUCTION

A grand challenge for this century is to predict how environmental change will alter global biogeochemical cycles. The field of biogeography has an important role to play in this effort (Violle *et al.* 2014). Environmental change is altering the distribution of biodiversity, which in turn, is a key driver of biogeochemical processes (Naeem & Wright 2003; Cardinale *et al.* 2012). Historically, biogeography has viewed biodiversity through a taxonomic lens, primarily resolving species distributions. However, a focus on traits – particularly those involved in ecosystem processes – may offer a clearer link between biodiversity patterns and biogeochemistry (Diaz & Cabido 2001; McGill *et al.* 2006; Reichstein *et al.* 2014).

These ideas are particularly relevant for microorganisms. Microbes catalyze most of the biological transformations of the major elements of life (Falkowski *et al.* 2008), and because of their sheer abundance, they account for a large pool of elements in living matter (Whitman *et al.* 1998). Furthermore, like plants and animals, microbial taxonomic composition varies over space (Martiny *et al.* 2006; Fierer *et al.* 2012b), and this variation can influence ecosystem processes (van der Heijden *et al.* 2008; Strickland *et al.* 2009; Schimel & Schaeffer 2012; Reed & Martiny 2013). Thus, a consideration of microbial traits should improve efforts to connect biogeographic patterns and ecosystem processes (Green *et al.* 2008).

Here, we provide a first characterization of the global biogeographic patterns of microbial nitrogen (N) cycling traits in soil. Microbially-driven transformations regulate biologically available N through exchange with the atmosphere (via N fixation and denitrification) and loss by nitrate leaching. They also influence the forms of N available

for plant uptake. At the same time, human activities have altered, and continue to alter, the nitrogen cycle by increasing the amount of reactive N in the biosphere (Vitousek *et al.* 1997; Fowler *et al.* 2013). At local scales, N addition consistently alters microbial composition in soils and other ecosystems (Allison & Martiny 2008; Ramirez *et al.* 2012). The distribution of microbial traits might therefore be relevant for understanding current and future N cycling.

The taxonomic composition of soil microorganisms is correlated with spatial variation in climate, plant diversity, pH, disturbance, and many other factors (Lauber *et al.* 2009; Philippot *et al.* 2009; Tedersoo *et al.* 2014; Prober *et al.* 2015). These biogeographic patterns help to identify factors that select on the entire suite of microbial traits. In this study, we reverse this direction of inquiry. We first characterize the patterns and drivers of just handful of traits associated with N cycling and then ask which taxa comprise these functional groups.

To quantify the abundance and composition of N-cycling traits, we analyzed ~2.4 billion short-read sequences from 365 soil metagenomes sampled from around the globe. From this dataset, we identified sequences that indicate the potential for a microorganism to perform one of eight N pathways that convert inorganic N to other inorganic forms or microbial biomass. We then quantified the frequency and taxonomic association of microorganisms carrying these pathways in each sample. If a gene from a pathway was detected, we assumed the presence of the entire pathway in the organism. To compare the frequencies among the N pathways, we standardized for the number of genes (2-20) in each pathway. While metagenomic sequences provide a measure of a community's trait diversity (Barberan *et al.* 2012), the presence of a trait

does not indicate how it is being used in the community. Thus, we cannot determine whether genes in the N pathways are expressed or N transformation rates. However, assaying traits based on metagenomic sequences are parallel to other trait metrics used to describe an organism's functional potential, such as nutrient uptake affinity or temperature optimum for growth.

The global soil N trait dataset allowed us to address four main questions. First, what are the overall frequencies of the different N pathways in soil? We expected the frequencies to vary greatly by pathway. Indeed, the ability to perform nitrification is restricted to few microbial taxa, whereas ammonia assimilation is probably present in almost all taxa. Second, what drives variation in the frequencies of N pathways among soil samples? We hypothesized that N pathway frequencies would vary primarily by habitat type, which reflects major differences in plant communities and therefore N inputs into soils. Third, what are the main taxa encoding each N pathway? Surprisingly little is known about the dominant lineages encoding N-cycling traits across global soils. We therefore expected to find previously unrecognized, prominent players, particularly for the less-studied pathways such as dissimilatory nitrate to ammonium (DNRA). Finally, what underlies compositional variation among soil samples in microorganisms encoding N pathways? We hypothesized that the taxa responsible for each pathway would vary greatly by habitat type, because the habitat would select for specialized taxa. We further predicted that soil pH – previously identified as an important driver of soil composition (Rousk *et al.* 2010; Tsiknia *et al.* 2015) – would also influence compositional variation within microorganisms encoding N-cycling traits.

RESULTS

Metagenomic data from surface soil samples were retrieved from the metagenomics analysis server (MG-RAST) (Meyer *et al.* 2008). After curating the samples for sequence and metadata quality, the final 365 samples represented 118 unique locations from 10 distinct habitat types covering natural and human-dominated systems (Figure 2.1; Table S1). Sequencing depth varied greatly among the samples, but was not overtly biased towards any particular habitat type (Figure S1). To standardize for sequencing depth, we report the abundance of each N pathway as its frequency in a sample. The trends observed were similar whether pathway frequency was normalized as the number detected per annotated sequence or per marker gene (based on 30 conserved, single-copy genes) (Figure S1).

Bacteria dominated the metagenomic libraries, comprising 95% of all sequences, followed by 2% for Archaea and only 3% for Fungi. The fraction of fungal sequences in metagenomic libraries is known to be lower than their contribution to soil microbial biomass (Fierer *et al.* 2012b). We therefore concentrate our analyses on Bacteria and Archaea and report only general trends for Fungi. For instance, the proportion of total sequences of Bacteria, Archaea, and Fungi varied across habitat type (G-test of independence; $p<<0.001$)(Figure S2). Archaea ranged from 0.9% to 11% of all sequences by habitat with the highest percentage detected in deserts. The ratio of fungal to bacterial sequences was particularly high in temperate forest soil, as previously observed (Fierer *et al.* 2009).

Frequency of Soil N Pathways

On average, 0.5% of all annotated sequences in a soil sample were associated with one of the eight N pathways (Figure 2.2a), or an average of 3.3 and 4.7 N pathways per marker gene for Bacteria and Archaea, respectively. The frequency of the individual pathways varied by several orders of magnitude (one-way ANOVA $p<0.001$; $F=74.21$, $df=7$) (Figure 2.2b). Bacteria and Archaea displayed similar trends in their relative frequency of N pathways except for the absence of the dissimilatory nitrite reduction to ammonium pathway in Archaea. Fungal sequences were only associated with assimilatory pathways, including ammonia assimilation, assimilatory nitrate to nitrite, and assimilatory nitrite to ammonium.

Across all domains, the most common pathway was ammonia assimilation (Figure 2.2b). For instance, among the Bacteria, an average of 280 ammonia assimilation pathways were detected for every million annotated bacterial sequences. In comparison, nitrification and nitrogen fixation were the least common pathways and detected only 6.1 and 4.6 times per million sequences, respectively. Notably, the relatively unstudied dissimilatory nitrite reduction to ammonium pathway was slightly more common that these two pathways, detected on average 9.3 times per million sequences.

Across all soil samples, N pathway frequencies were overwhelmingly positively correlated for both the Bacteria and Archaea (Figure 2.3a&b). To examine difference in pathways beyond the trends shared by all, we calculated the residuals of the frequency of each pathway regressed against the frequency of all N pathways in a sample. This residual variation was also significantly correlated among many of the N pathways (Figure 2.3c&d). For instance, denitrification was highly positively correlated with

dissimilatory nitrate reduction to nitrite within both Bacteria and Archaea (R^2=0.86 and 0.97, respectively, p=<0.001). This relationship is expected, because dissimilatory nitrate reduction to nitrite is the first step of the complete denitrification process; however, we separated the two steps here, because nitrate reduction to nitrate is also the first step in dissimilatory nitrate reduction to ammonium (DNRA)(Yoon *et al.* 2015). Similarly, we separated DNRA into its two pathways: dissimilatory nitrate reduction to nitrite and dissimilatory nitrite reduction to ammonium (Figure 2.2a). Among Bacteria, the assimilatory nitrite to ammonium pathway residual was negatively correlated with all other pathways. Similarly, the residual frequency of the ammonia assimilation pathway was negatively correlated with all other N pathways in both Bacteria and Archaea. N fixation generally showed weak or no correlation with other pathways.

Drivers of N Pathway Frequencies

The frequency of all N-cycling traits (summing across all pathways) varied greatly among soil samples, and initial analyses revealed broad biogeographic patterns. On average, the highest frequencies of total bacteria N pathways were detected in tropical forest and human-dominated (pasture, lawn, and agriculture) soils, whereas the lowest frequency were observed in cold deserts (Figure S3). Total N pathway frequency also tended to decrease with increasing latitude (R^2=0.22*; p*<0.05; Figure S4).

To disentangle the drivers behind these patterns, we performed a multivariate regression analysis including habitat type and environmental parameters known to influence microbial abundance and composition (Fierer & Jackson 2006; Bru *et al.* 2011). Local measurements were not available for most samples; instead, we estimated

these variables from secondary sources. For Bacteria, the regression model explained a large and significant proportion of the variability in the frequency of total N pathways (R^2=0.58*; p*<<0.001; Table 2.1). Habitat type contributed most to this model, both directly (positively related to total N pathways) and through interactions with soil carbon and N. The regression model for Archaea explained less variability in total N pathway frequency than for Bacteria (R^2=0.43; p<0.001; Table 2.1). An interactive effect between carbon and nitrogen contributed the most to the model, and habitat was only important through an interactive effect with temperature.

We next examined the drivers of individual N pathway frequencies. Due to high covariance between pathways (Figure 2.3a&b), we fitted regression models to the total-frequency-corrected residuals for each pathway. These models varied greatly in their ability to explain this additional variation (Table 2.1). For example, the models for the N fixation pathway explained 80% and 63% of the variation among samples in Bacteria and Archaea, respectively (p<<0.001). In contrast, the model parameters did not explain any variation in the frequency of the dissimilatory nitrite reduction to ammonium pathway in Bacteria.

Among the significant models, habitat type was an important predictor of the individual pathway frequencies (Table 2.1). Habitat also interacted with other factors including precipitation, temperature, and soil nitrogen to influence the frequency of some pathways. For instance, denitrification frequency increased with temperature in deserts, but decreased with temperature in tropical forests. Similarly, ammonia assimilation frequency increased with soil nitrogen in temperate forests, but decreased with soil nitrogen in tropical forests. Soil carbon, which appeared to be a primary driver

of total N pathway frequency, did not explain differences in the frequency of individual pathways in Bacteria. Including estimates of N deposition in these models only improved the denitrification model (R^2 increased from 0.41 to 0.48); denitrification frequency increased with increasing N deposition.

The models for individual pathway frequencies in Archaea generally explained less variation than those for Bacteria, perhaps due to the lower number of sequences per sample (Table S1). However, for the significant models, the individual N pathways were often best explained by the same parameters as the Bacteria. For instance, habitat type and habitat-by-temperature were the most important predictors of N fixation frequency within both domains. Likewise, habitat, habitat-by-precipitation, and habitat-by-temperature contributed to the variation in assimilatory nitrate to nitrite frequency in both Archaea and Bacteria.

Taxonomic and Phylogenetic Distribution of N Pathways

A diverse range of microorganisms, encompassing 402 bacterial and 53 archaeal genera, encoded the N pathways. We first investigated the association of pathways within the same genera (Figure 2.4, Figure S5). All genera for which we detected over 10 sequences carried the ammonia assimilation pathway. Genera carrying the pathway to complete the second half of denitrification also generally carried the first half of the pathway, dissimilatory nitrate to nitrite reduction. The same genera carrying these denitrification pathways sometimes, but not always, carried the dissimilatory nitrite reduction to ammonium pathway, or the second part of the complete DNRA process (Figure 2.4, Figure S5). Some genera within the gamma-, delta-, and epsilon-

Proteobacteria (e.g., *Edwardsiella, Wolinella, Anaeromyxobacter*) contained all three pathways. Indeed, denitrification and DNRA has recently been shown to be present and functional in the same bacteria (Mania *et al.* 2014; Yoon *et al.* 2015). We also detected genera that only carried the dissimilatory nitrite to ammonium pathway (in addition to ammonia assimilation), as was the case for five genera within the phylum Bacteriodetes.

More broadly, soil genera, and the phyla they fall into, varied in their degree of pathway specialization. Genera within the delta-Proteobacteria appeared to be N-cycling generalists, harboring up to six pathways (in addition to ammonia assimilation). Note, however, that these patterns do not distinguish between whether these genera are made up of generalists that encode many pathways or multiple specialists that encode specific pathways. In contrast, genera within the Cyanobacteria seemed to be specialists, carrying primarily the assimilatory nitrite to ammonium and N fixation pathways.

Focusing on each pathway individually revealed the most prominent taxa carrying the pathway across all soil samples. Here we consider two contrasting pathways, both in terms of their taxonomic distribution and the degree to which they have been studied. First, the abundance of the N fixation pathway in the soil samples was distributed broadly among both Archaea and Bacteria (Figure 4, Figure S5). The most abundant N fixers detected were concentrated within the phylum Proteobacteria, with notable exceptions among the Chlorobi, Firmicutes, and Cyanobacteria (Figure 5a). Most sequences were closely related to N-fixing genera that might be predicted to be common in soil, such as *Bradyrhizobium* and *Burkholderia*. Other abundant genera

were less expected. For example, *Azoarcus* is an organism studied for its abilities to degrade soil contaminants (Sun & Cupples 2012), and *Pectobacterium* (gamma-Proteobacteria) is known primarily as a plant pathogen (Ma *et al.* 2007). Indeed, while it is known that *Pectobacterium* encodes the suite of N fixation genes, it remains unclear if they are functional (Toth *et al.* 2015).

Second, the pathway encoding dissimilatory nitrite reduction to ammonium was also broadly distributed across soil bacteria (Figure 2.4), as noted before (Welsh *et al.* 2014). However, the dominant soil taxa were restricted to two phyla, the delta-Proteobacteria and Verrucomicrobia (Figure 2.5b). Verrucomicrobia are known to be abundant in soils, but their ecological role remains unclear (Bergmann *et al.* 2011; Fierer *et al.* 2013). The pathway's most abundant genus*, Anaeromyxobacter* (phylum delta-Proteobacteria), is common in agricultural soil and has recently been shown to carry out a previously unrecognized process of non-denitrifying N_20 reduction to N_2 (Sanford *et al.* 2012). The relative abundances of genera encoding the other six N pathways in the soil samples are reported in Figure S6.

Drivers of Taxonomic Composition by N Pathway

The same environmental variables that explained the overall frequency of the N pathways well, explained much less of the variation in the taxonomic composition of the organisms encoding the pathways. For the 8 pathways, the models only explained 7% to 19% of the composition variation of the individual N pathways (Table S2). However, as for pathway frequency, habitat type was the best predictor of composition, explaining up to 14% of the compositional variation in the assimilatory nitrite to ammonium

pathway. Temperature also explained 11% of the compositional variation for the nitrification pathway. All other predictors, including pH, explained at most 3% of the variation for any pathway.

A closer examination of two pathways confirms weak compositional differences between the habitats. The 15 most abundant genera carrying the N fixation pathway were similarly abundant across all habitats except in cold deserts (Figure 2.5a). The most abundant genera encoding the dissimilatory nitrite reduction to ammonium pathway displayed greater variability among habitats (confirming the model results in Table S2), but of these, only one genus (*Chlorobium*) appeared specialized on a habitat (wetland) (Figure 2.5b).

DISCUSSION

Here, we used metagenomic data to characterize the biogeographic patterns of microbial N cycling traits in soil. The advantage of this approach is that it allows us to identify the traits – and the organisms harboring them – involved in many key functions at once. Specifically, the analysis provides a comprehensive map of the dominant lineages involved in eight N processes in soil. The approach also allowed us to search all known genes in a pathway, while avoiding primer biases towards particular lineages (Myrold *et al.* 2014).

The overall structure of microbial N traits – the relative frequency of the eight pathways – appears to be quite consistent across global soils. This is not unexpected, but had not been previously tested. For instance, the ammonia assimilation pathway was relatively common, and the pathways for nitrogen fixation and nitrification were

relatively rare, as observed previously in soil and other environments (Varin *et al.* 2010; Martiny *et al.* 2013; Quinn *et al.* 2014; Souza *et al.* 2015). Less expected, however, was that N pathway frequencies within a soil sample were overwhelmingly positively correlated (Figure 3). This result suggests that soil communities with relative high numbers of cells able to use one N pathway also generally support higher numbers of cells that can use other N pathways. Greater numbers of metagenomic sequences associated with nutrient cycles have previously been interpreted to be indicative of faster nutrient cycling rates (Fierer *et al.* 2012b). The positive correlations between pathways within the N cycle would seem to support this hypothesis. We also found a high frequency of Bacteria encoding dissimilatory nitrite reduction to ammonium pathway, which leads to recycling of N in soils. The balance between DNRA and denitrification, which leads to the loss of N to the atmosphere, is thought to be key to soil N budgets. Our results confirm previous studies suggesting that this pathway may be more common than previously thought (Rutting *et al.* 2011; Nelson *et al.* 2015), but the taxa encoding the process in soil environments remain to be carefully characterized (Kraft *et al.* 2011).

The frequency of N traits further displayed clear biogeographic patterns. At the broadest scale, N trait frequency in Bacteria tended to decrease at higher latitudes, perhaps reflecting a general trend in N limitation in high latitude ecosystems (Yergeau *et al.* 2007). Beyond latitude, the frequency of N cycling traits in soil communities depended largely on habitat type as well as soil carbon and nitrogen concentrations. N traits were highest in human-dominated habitats, where nitrogen inputs tend to be high, and tropical forests, which are generally thought to be less limited by nitrogen than

temperate ecosystems (Vitousek 1984). In contrast, N traits were lowest in cold deserts (Antarctic and Arctic), which are highly nutrient limited (Jonasson *et al.* 1999; Yergeau *et al.* 2007). However, given the low sample numbers for some habitat types, it will be important to retest these patterns as more data accumulates.

Contrary to our hypothesis, the taxa responsible for each N pathway did not vary greatly by habitat type. Within a pathway, genera that were dominant in one habitat tended to be dominant in all habitats. More generally, the environmental variables in our analyses were poor predictors of the compositional variation of the N functional groups. One possible reason for this result is that environmental preferences are conserved below the genus level and therefore would not be detected by our analysis. However, this reasoning does not explain why soil pH appears to have little influence on composition, as pH preference seems to be conserved at a broader taxonomic level (Lauber *et al.* 2009; Martiny *et al.* 2015). Perhaps N functional groups are less specialized for a particular pH environment than microorganisms with other functional roles, but distinct pH-associated lineages in ammonia-oxidizing archaea (AOA) indicate that this is not always the case (Gubry-Rangin *et al.* 2011). Alternatively, the estimates of soil pH might have been too spatially coarse to detect a pattern.

A well-recognized issue in calculating the frequencies of genes or pathways from metagenomic data is how to normalize for overall genome abundance in the library (Manor & Borenstein 2015). This normalization step is prone to uncertainties related to variation in mean genome size among communities. To address this issue, we estimated the frequencies of N pathways in two ways – using a set of conserved marker genes as well as the total number of annotated sequences within a domain. The first

approach should be sensitive to differences in genome size, whereas the second approach includes more sequence reads and is thus more statistically robust. Because the two approaches led to similar findings, we conclude that the overall patterns in N pathway frequencies are likely not an artifact of normalization.

In sum, this study provides a foundation for future trait-based investigations of soil N cycling, but also highlights two major challenges. First, we still know very little about how variability in the frequency and composition of microbial N traits will affect process rates in soil environments (Prosser 2015). Indeed, a recent review found little correlation between an individual gene's abundance and the process rates that such genes encode. However, assessment of these links using metagenomic datasets is still needed (Rocca *et al.* 2015). Second, assigning function and taxonomy from short read sequences is limited by genomic databases where annotations in some cases may be sparse and/or erroneous (Wu *et al.* 2009; Thomas *et al.* 2012). The N cycle is an archetype of this problem, as new N processes and lineages continue to be identified (Strous *et al.* 1999; Konneke *et al.* 2005; Farnelid *et al.* 2011; Sanford *et al.* 2012; van Kessel *et al.* 2015). Despite these challenges, the application of metagenomic data to a trait-based framework offers a powerful avenue for elucidating the role that microbial communities play in regulating biogeochemical processes (Barberan *et al.* 2012; Fierer *et al.* 2014).

MATERIALS AND METHODS

Dataset and curation. Metagenomic samples (sequencing type "whole genome sequencing" and environmental package "soil", n=809) in the MG-RAST database (Meyer *et al.* 2008) were classified into one of ten habitat types (Desert, Cold Desert, Grassland, Temperate Forest, Tropical Forest, Tundra, Wetland, Agriculture, Pasture, Lawn). Samples that could not be classified into these habitats (e.g., oil spill, mines, and microbial mats) were not considered further.

GPS coordinates and sample date associated with each metagenome id were downloaded from MG-RAST via the R package *matR* (Team 2011; Braithwaite & Keegan 2013). To minimize the problem of pseudoreplication, we only considered samples from one date per location (the date with the most samples). Based on the statistics provided by MG-RAST, we further removed samples if: 1) the number of uploaded sequences was equal to the number of post-QC sequences, which seemed to indicate a pre-processing step; 2) the number of identified protein features was <10,000; or 3) the total bacterial reads <10,000. The remaining metagenomic libraries (n=365) encompassed 118 unique locations. These were downloaded using the MG-RAST API version 3.2 with KEGG database annotations. Each sequence was assigned to the closest related genus in the database using an e-value of less than or equal to 10^{-5}.

Data standardization across metagenomic libraries. Because sequencing effort varied greatly among samples, we standardized the bacterial and archaeal sequences by a suite of conserved, single-copy (i.e., marker) genes to control for possible variation in average genome size among samples (Frank & Sorensen 2011)(Figure S1). The Kegg Orthology (KO) numbers for 30 Bacteria and Archaea marker genes (Frank &

Sorensen 2011) were matched to MD5 IDs using the non-redundant M5nr database. We then searched for these MD5 IDs in the samples annotated by the MG-RAST server.

The number of marker genes was also highly correlated with the total number of annotated sequences in a sample (R^2 = 0.86; Figure S1). Thus, when comparing across Archaea, Bacteria, and Fungi, we standardized the samples by total annotated sequences. Sequencing effort varied greatly among the samples, but was not overtly biased towards any particular habitat type (Figure 2.3).

Identification of N cycle pathways. In each metagenomic library, we searched for sequences from eight N pathways, defined previously in (Nelson *et al.* 2015). These pathways included nitrification (number of genes targeted: n=2), N fixation (n=20), denitrification (n=20), dissimilatory nitrate to nitrite reduction (n=9), dissimilatory nitrite to ammonia reduction (n=4), assimilatory nitrate to nitrite reduction (n=2), assimilatory nitrite to ammonia reduction (n=2), and ammonia assimilation (n=10) (Figure 2a). If a gene from a pathway was detected, we assumed the presence of the entire pathway.

Environmental Metadata. Environmental data was retrieved from a variety of publically available sources. In all cases, gridded spatial data files were downloaded, and data was extracted using the R packages *raster*, *rdgal,* and *sp* (Hijmans & van Etten 2012; Bivand *et al.* 2013). The data included average precipitation (mm) and temperature (˚C) from the month of sampling (Hijmans *et al.* 2005), soil pH (Batjes 2000), total organic carbon (kg/m^2)(Batjes 2000) , total organic nitrogen (g/m^2) (Group 2000), and nitrogen

deposition (mg N/m^2/year) (Dentener *et al.* 2006). Approximate data grid resolution for precipitation and temperature was 0.01˚, for soil pH and organic carbon was 0.5˚, for total organic nitrogen was 0.1˚, and for nitrogen deposition was 4.0˚. Environmental metadata was assigned to each sample using the associated latitude and longitude coordinates. Where data was categorized into ranges (soil pH and total organic carbon), the average value from the range was used.

Statistical Analyses. To compare the relative abundance of N pathways across samples, we calculated the frequency of each pathway in a sample for both the Bacteria and Archaea. This frequency is the estimated number of times the pathway was detected per marker gene detected, or: [# of pathway reads / # of pathway genes searched]/[# marker gene reads/30]. Thus, a pathway's frequency of detection was also standardized for the number of genes in the pathway.

To test for differences in the frequency across pathways, we used a one-way analysis of variance, using the *aov* function in R. To test for correlations between the frequencies of the individual pathways within a sample, we used Spearman's correlation coefficient. To calculate the total N pathway frequency of each sample, we summed the frequency of all eight pathways. We used *lm* in R to calculate the residuals of each N pathway against a sample's total N pathway frequency.

To tease apart the relative importance of environmental variables on the frequency of N pathways, we used a multiple regression model (*lm* function in R) including the following variables: habitat type, temperature, precipitation, soil pH, organic carbon, total nitrogen. For this analysis, we averaged data across multiple

samples from the same location at just one sampling time, yielding 118 datasets. Based on *a priori* expectations (Ramette 2007), we also included the following interaction terms: habitat-by-temperature, habitat-by-precipitation, habitat-by-soil pH, habitat-by-organic carbon, habitat-by-total nitrogen, precipitation-by-temperature, and organic carbon-by-total nitrogen. To determine the relative importance of the various significant environmental factors from our model in contributing to variation in the frequency of N pathways across samples, we used a backwards selection procedure (Mac Nally 2002; Ramette 2007). Starting with the significant terms ($p<0.01$) from our original model, we removed variables one at a time; the differences in R^2 values between each step were used to calculate the relative importance of the independent variable removed from the model. If there was no change or only a marginal change in R^2 when the term was removed, the term was assigned a relative importance of <0.01. After the initial analysis, N deposition was added to test if this parameter improved the model.

To analyze the composition within each pathway, we calculated the proportional abundance of the genera in a sample and averaged these proportions across multiple samples from the same location. We then calculated a Bray-Curtis distance matrix for all sample locations. We used a distance-based linear model (DISTLM; PRIMER v6; PERMANOVA ++, (Clarke & Warwick 2001; Anderson *et al.* 2008)) to test the significance and importance (an estimate of the proportion of R^2 explained) of the predictor variables for each pathway's composition, using a forward selection procedure.

Phylogenetic visualization. We constructed a phylogenetic tree including a representative species from all genera encoding N sequences using 16S rRNA amplicon data (chosen for their sequence quality and length of ~1400 bp) from the SILVA database (Quast *et al.* 2013). We aligned the sequences using SINA (Pruesse *et al.* 2012) and created a neighbor-joining tree with the default parameters in Geneious v9.0.5. We used the Interactive Tree of Life (iTOL) (Letunic & Bork 2007) to plot (1) the proportion of N pathways (excluding ammonia assimilation) detected within each genus and (2) the relative abundance of genera encoding each individual pathway across the unique sampling locations (n=118). For the N fixation and dissimilatory nitrate reduction pathways, we used the *ggplot2* package (Wickham 2009) in R to plot heat maps of the relative frequencies of the 15 most abundant genera by habitat.

ACKNOWLEDGMENTS

We thank John Avise, Francisco Ayala, and Brian Bowen for the invitation to participate in this Symposium. Thanks also to Alex Chase for helpful feedback on earlier drafts of the manuscript. This work was supported by a US Department of Education Graduate Assistance in Areas of National Need (GAANN) fellowship to MBN and the US Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), under Award Number DE-PS02-09ER09-25.

Tables and Figures

Table 2.1 Variation explained by the environmental variables in the regression models of the frequency of all (total) and individual N pathways.

		Individual pathways (residuals)							
Environmental variables	Total	Ammonia assimilation	Assimilatory nitrate to nitrite	Assimilatory nitrite to ammonia	N fixation	Nitrification	Dissimilatory nitrate to nitrite	Denitrification	Dissimilatory nitrite to ammonia
Bacteria									
Habitat (H)	0.14	0.02	0.23	0.07	0.29	0.11	0.06	0.09	
Precipitation (P)		<0.01							
Temperature (T)		<0.01			0.02			0.02	
pH									
Organic carbon (C)	0.12					<0.01			
Total N	0.05				0.13				
H × P	<0.01		0.07			0.08	0.32		
H × T			0.09	0.23	0.31	0.21	0.03	0.31	
H × pH	<0.01		0.09		0.05				
H × C	0.1								
H × N	0.17	0.49		0.06					
P × T			0.02			0.05		<0.01	
C × N							<0.01		
Adjusted R^2	0.58	0.51	0.5	0.36	0.8	0.45	0.41	0.41	NS
Archaea									
Habitat			0.08		0.09	0.03		0.12	
Precipitation					<0.01				
Temperature						0.02		0.03	
pH								<0.01	
Organic carbon									
Total N			0.05			0.04			
H × P			0.21		0.09				
H × T	0.09		0.18		0.33	0.13			
H × pH					0.12			0.06	
H × C									
H × N									
P × T									
C × N	0.34								
Adjusted R^2	0.43	NS	0.52	NS	0.63	0.22	NS	0.21	NA

The models for the individual pathways are based on the residual frequencies of the pathway after correcting for the Total N pathway frequency (see text). Estimates of the fraction of explained variation are only reported for significant variables (p <0.05). Samples were only included when all environmental variables could be obtained for that location (n=99).

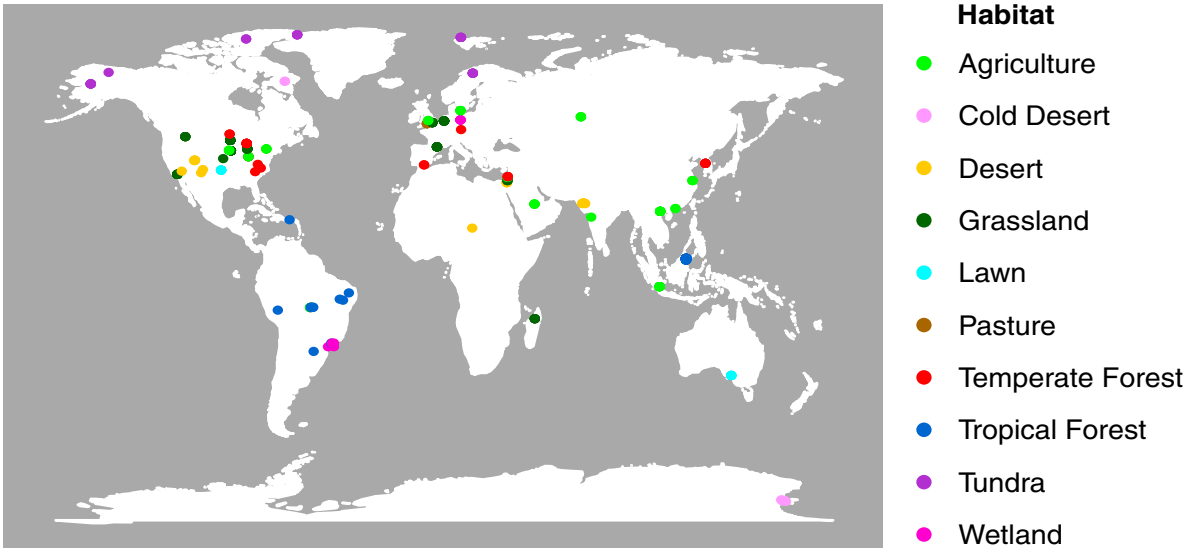


Figure 2.1 The locations (*n*=118) sampled to create the soil metagenomic libraries (*n*=365) used in this analysis. The samples represent 10 distinct habitats including Agriculture (*n*=19), Cold Desert (6), Desert (15), Grassland (14), Lawn (4), Pasture (2), Temperate Forest (12), Tropical Forest (34), Tundra (7), and Wetland (5).

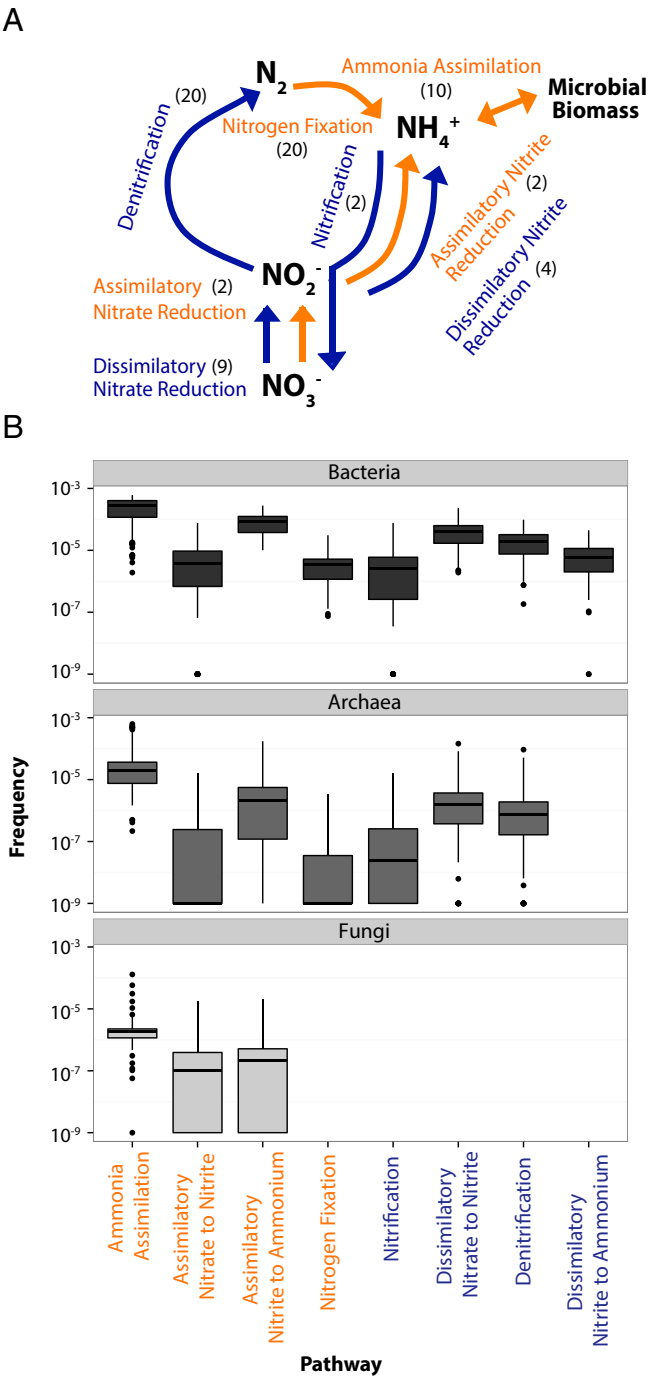


Figure 2.2 N pathways and their frequencies. A) N pathways considered in this study. The numbers in parentheses are the number of genes targeted for each pathway. Assimilatory pathways are in orange, and dissimilatory pathways in blue. B) Box plot of the frequency of each N pathway in a metagenomic library for Bacteria, Archaea, and Fungi. To compare across domains, frequencies are calculated as per annotated sequence in each domain. The upper and lower bounds of boxes correspond to the 25th and 75th percentiles, with a median line shown. Whiskers represent 1.5*IQR (interquartile range). Dots represent outliers.

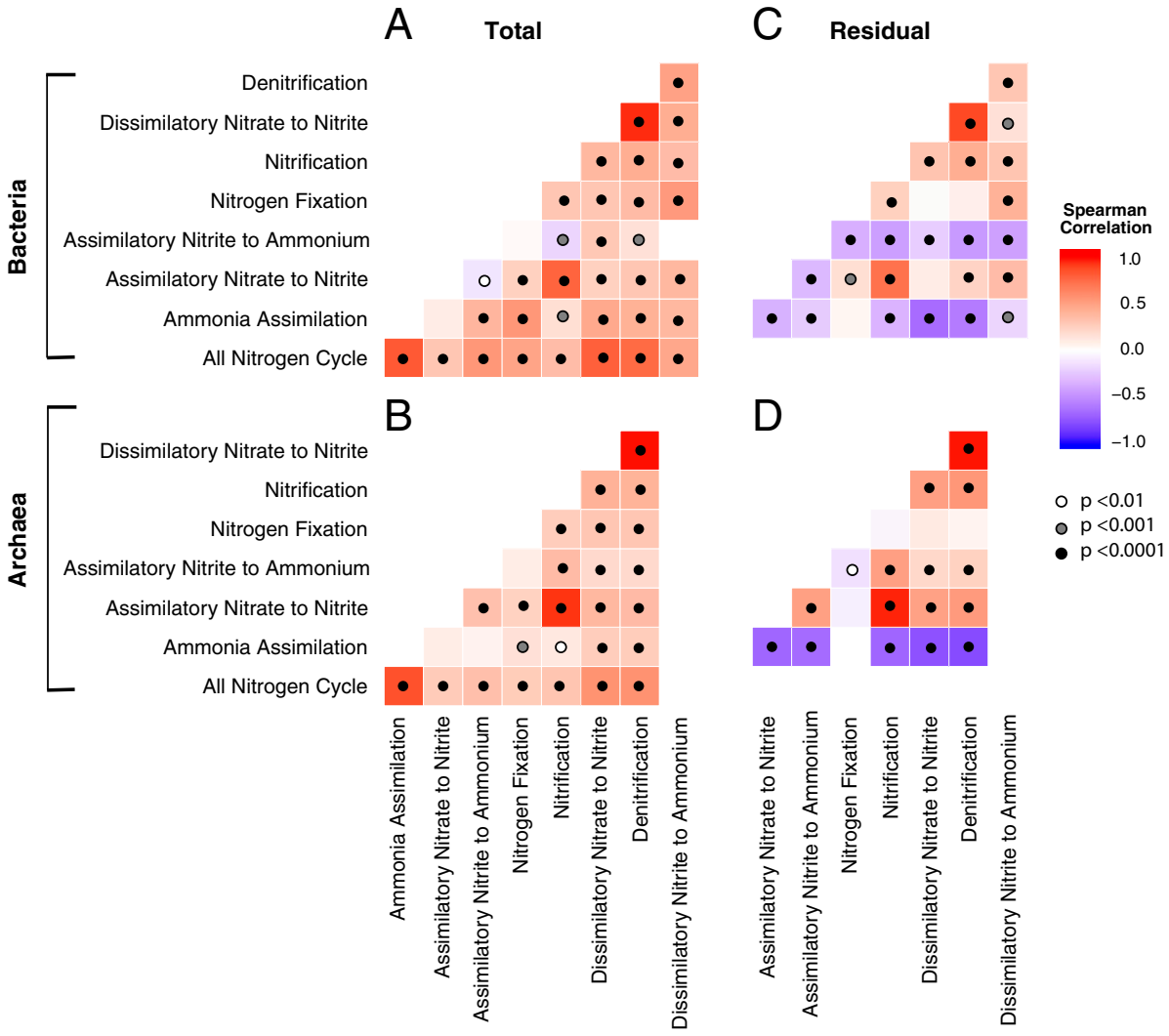


Figure 2.3 The relationships between N pathway frequencies**.** Correlations between N pathways encoded by Bacteria (A) and Archaea (B) across the samples. (C and D) Correlations between the residuals of each pathway regressed against the total frequency of all N pathways.

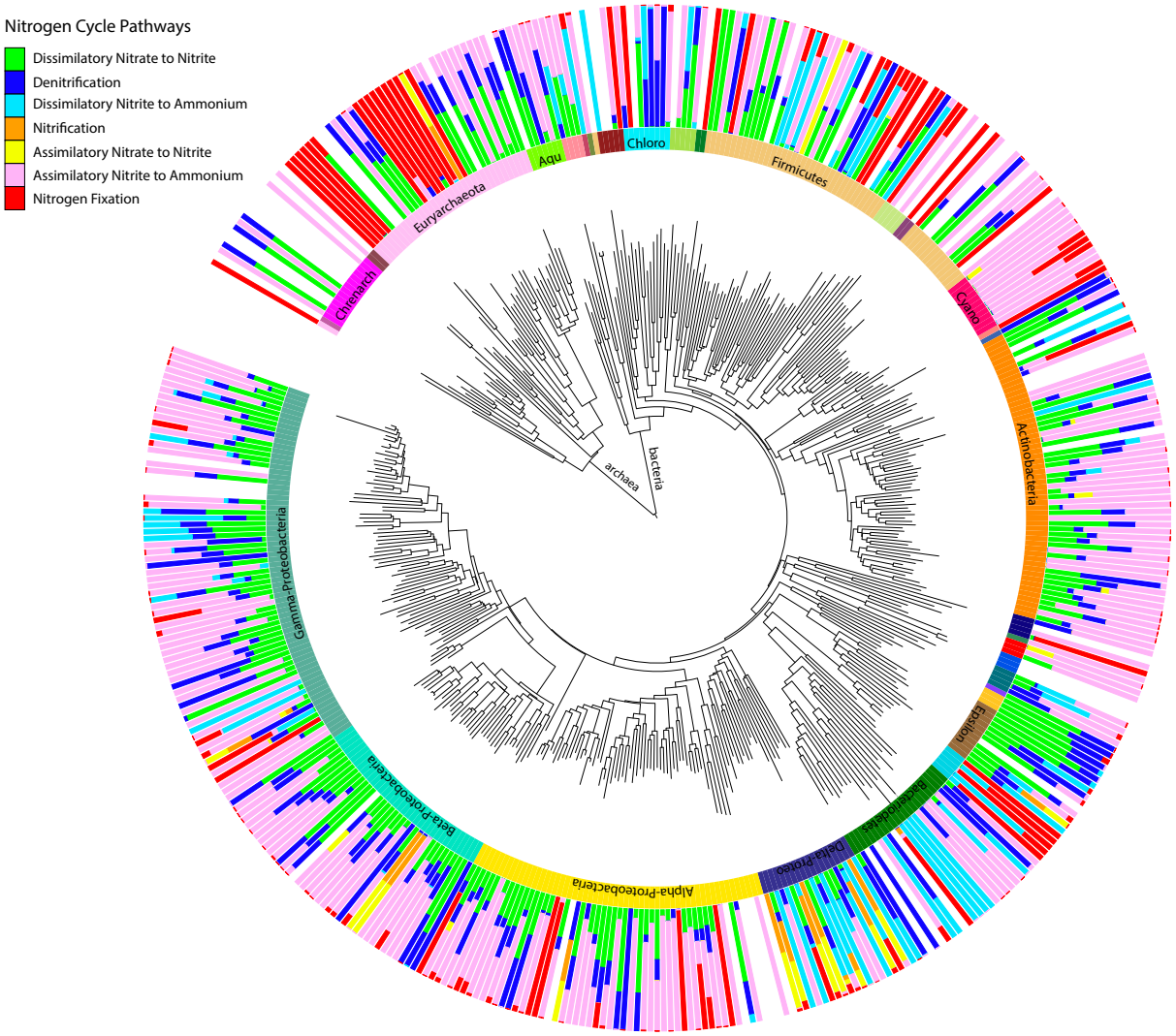


Figure 2.4 Phylogenetic distribution of N pathways in the soil metagenomes. A neighbor-joining tree was constructed using 16S rRNA data (see Methods) and includes all archaea and bacteria genera associated with N cycling sequences in the dataset. The outer-circle plots the proportion of N cycle reads assigned to each pathway within the genus. The ammonia assimilation pathway is excluded, because it was found in all genera represented by at least 10 sequences. The inner-circle indicates major classes and phyla. See Figure S5 for a high-resolution figure with genus labels.

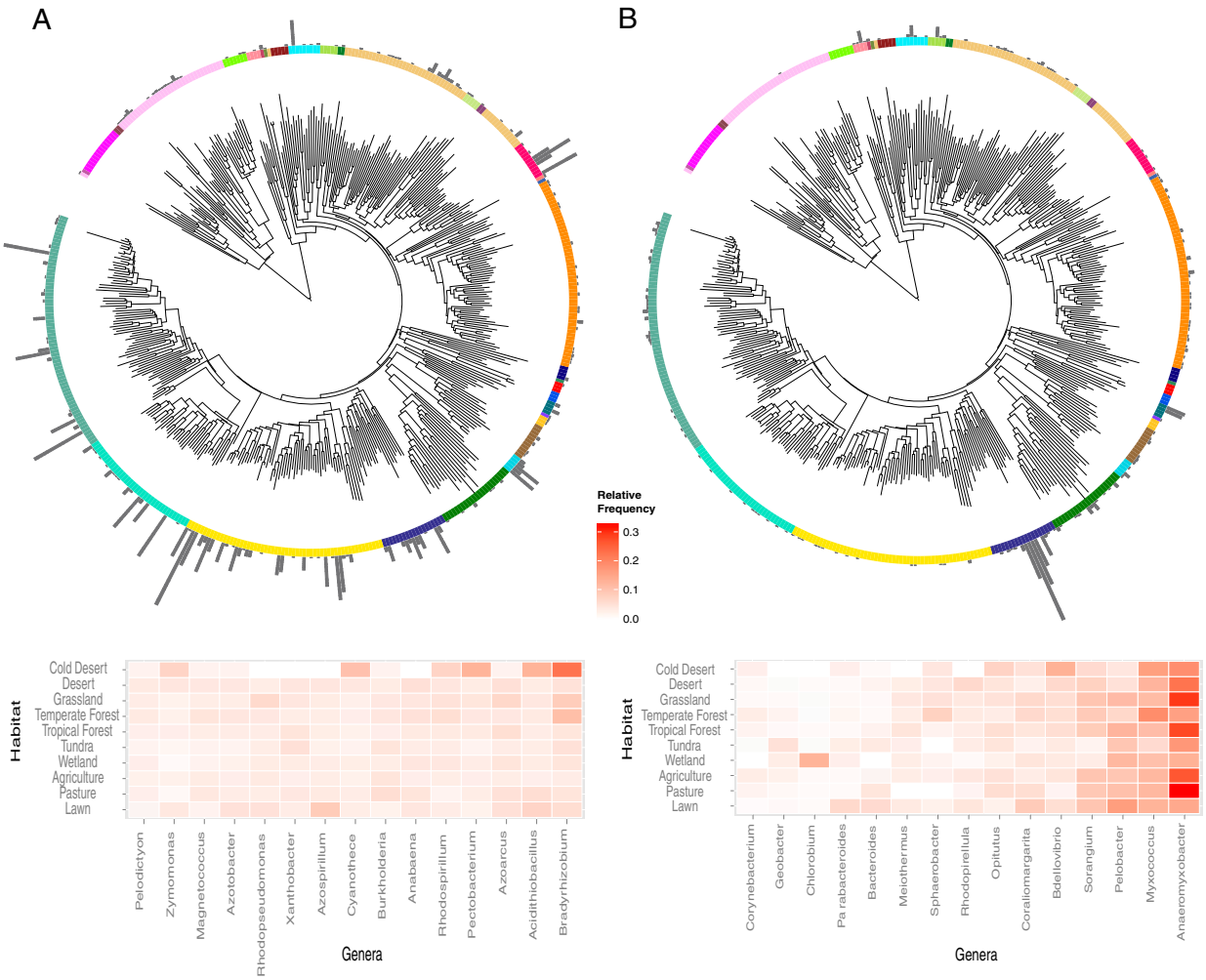


Figure 2.5 Phylogenetic distribution of genera encoding specific pathways**.** The neighbor-joining tree was constructed using 16S rRNA data (see Methods)**.** The relative abundance of genera associated with (a) nitrogen fixation and (b) dissimilatory nitrite reduction to ammonium. Within a pathway, the proportion of sequences associated with each genera was calculated for each sample. The proportions were averaged across libraries within each location and then averaged across the 10 habitat types (to provide equal weighting to the habitats). The heatmaps give the relative frequency by habitat for the 15 most abundant genera associated with each pathway. See Figure S6 for plots of the other 6 pathways.

Supporting Information

This chapter contains supporting information that can be found online at http://www.pnas.org/content/113/29/8033.long?tab=ds (doi:10.1073/pnas.1601070113).

CHAPTER 3

Dispersal alters bacterial diversity and composition in a natural community

ABSTRACT

Dispersal is central to the evolution and maintenance of microbial diversity. Quantifying microbial dispersal and its role in shaping communities remains a challenge, however. Here, we manipulated a bacterial community's dispersal rate in a grassland ecosystem and test whether this altered diversity and composition. We constructed bags of two nylon mesh sizes that were closed or open to bacterial movement and filled them with an edible or inedible substrate, irradiated plant litter or nylon sheets. We measured changes in bacterial abundance (by flow cytometry) and composition (by 16S amplicon sequencing) in the bags weekly over five months. The dispersal treatment altered bacterial colonization rates and led to differences in the abundance, richness, evenness, and composition of communities. Overall, the study demonstrates that dispersal influences the assembly of this natural bacterial community.

Introduction

Dispersal, or the movement of organisms, plays an important role in the evolution and maintenance of biodiversity (Leibold *et al.* 2004; Cottenie 2005). Recent biogeographic studies and field experiments suggest that, as with larger organisms, microorganisms are dispersal limited (Bell 2010; Lindstrom & Ostman 2011; Hanson *et al.* 2012) or in other words, that the probability of a bacterium moving away from a location varies with distance. Because of this limitation, dispersal rates can influence both diversity (richness and evenness) and composition of microbial communities

(Whitaker *et al.* 2003; Horner-Devine *et al.* 2004; Martiny *et al.* 2011; Andam *et al.* 2016). Additionally, dispersal may impact communities through mass effects (Evans *et al.* 2017). However, quantifying the dispersal of microorganisms is a challenge due to their small size and high abundance (Nemergut *et al.* 2013a), hence the relationship between dispersal rates and microbial diversity or composition remains to be tested. Indeed, these relationships are likely complex and dependent on environmental context (Chase 2007; Louca *et al.* 2016a; Evans *et al.* 2017). Thus, field experiments will be required to elucidate the role of dispersal in shaping microbial communities.

Here, we tested if varying a community's dispersal rate alters the diversity and composition of a bacterial community. We focused on bacterial communities on plant litter, the top layer of soil, in a southern California grassland (Evans *et al.* 2017). To manipulate dispersal rate (the number of cells migrating into/out of a community per time), we constructed bags from two nylon mesh sizes. The high dispersal rate treatment (18.0 μm; hereafter 'open') allowed for the migration of bacteria and larger microorganisms through the bags, whereas the low dispersal rate treatment (0.22 μm; 'closed') substantially reduced dispersal (Allison *et al.* 2013; Evans *et al.* 2017). To disentangle the influence of dispersal alone versus growth and successional dynamics, we also manipulated the substrate inside the bags. Half of the bags within each dispersal treatment contained an edible substrate (irradiated plant litter), and half contained an inedible substrate (irradiated nylon sheet), where microbes could land but not grow. In total, we deployed 192 bags into the field and collected three replicates per treatment at 16 time points over the course of 5 months. For each bag, we measured total cell abundance by flow cytometry and characterized the bacterial community using

16S amplicon sequencing. During the last two months of the experiment, we also assayed samples from surrounding plant litter (detailed methods in *Supplementary Information*).

Results & Discussion

Bacterial abundance differed by dispersal treatment (Figure 3.1a). In the litter-containing bags, abundance was 151 times higher in the open versus closed treatment after just one week (2.3×10^7 per g dry litter and 1.5×10^5 per g dry litter, respectively; t-test: $t_4=3.36$, $p=0.03$). This difference persisted for the first 7 weeks of the experiment, indicating that the closed bags successfully lowered bacterial dispersal rates (Table S1). Notably, however, abundance also increased in the closed bags such that average abundance in the open and closed treatments did not significantly differ after 7 weeks (Table S1). After 11 weeks, bacterial cell counts leveled off in both open and closed bags at 2.1×10^8 cells/ g dry litter, similar to that observed in the surrounding plant litter (Figure 1a). Thus, as with the phyllosphere (Remus-Emsermann *et al.* 2012), litter communities appear to be subject to a carrying capacity, and the capacity in our litter bags was similar to that of the surrounding, litter environment.

The open and closed nylon-containing bags also differed in their abundance; additionally, overall abundance was greatly reduced due to lack of growth. Bacterial abundance in the open bags (113 cells per cm^2 nylon) was on average 330 times higher than in the closed bags (0.34 cells per cm^2 nylon) (ANCOVA; dispersal treatment: $F_{1,14}=20.6$, $p<0.001$; inset in Figure 3.1a). We were unable to PCR amplify DNA from

the closed nylon-containing bags, confirming that dispersal was greatly reduced by the smaller mesh size.

Higher community dispersal rates led to increased bacterial diversity. Observed richness was significantly higher in the open versus closed litter-containing bags (Figure 3.1c, Table S1). Richness also appeared limited by dispersal in the litter-containing bags, as it increased significantly over time. Furthermore, higher dispersal increased evenness in the litter-containing bags (Figure 3.1d, Table S1). In general, growth on the irradiated litter led to more even communities in the later stages of the experiment, similar to the successional dynamics of a sterile substrate by marine taxa (Datta *et al.* 2016). In contrast, evenness in the open nylon-containing bags decreased over time, reflecting the pattern of the surrounding litter community (Figure 1d). We suspect that this decline in evenness is due to typical seasonal changes at this grassland site (Figure S1) (Matulich *et al.* 2015).

Dispersal rate also influenced bacterial composition (Figure 3.1b, 3.2, Table S2). The closed litter-containing bags had significantly higher within-treatment variation than either the open-litter containing bags or the environmental litter (PERMDISP; $p=0.02$ and $p=0.004$, respectively). Thus, lower bacterial dispersal rates appear to increase ecological drift and result in more divergent composition (Hanson et al. 2012)(Evans *et al.* 2017). The nylon-containing bags, which did not allow for growth, were more representative of the surrounding plant litter community than the litter-containing bags (Figure 3.1b, 3.2, Table S3). Presumably, the nylon-containing bags are a random subsample of the surrounding community whereas the communities in the litter-containing bags are in the early stages of succession and therefore also influenced by

growth and competition. Still, relative abundance of individual taxa in both the open nylon- and open litter-containing bags was positively correlated with the relative abundance of taxa in the environment (R^2= 0.77 and 0.76, respectively, both $p<0.001$; Figure S2). These strong correlations suggest that bacterial dispersal occurs primarily passively in this system, as taxa generally colonize the bags in proportion to their abundance in the surrounding plant litter. However, dispersal rates among taxa could depend on traits such as size, adhesive ability, and in the case of an edible substrate, competitive ability. Thus, it is notable that a Cytophagaceae taxon (genus *Hymenobacter*), previously characterized as a common atmospheric bacterium (Yooseph *et al.* 2013; Barberan *et al.* 2015), was more abundant on the nylon than in the environmental litter (Figure 3.2, S2). Similarly, an Oxalobacteriaceae taxon (genus *Massilia*) was more abundant in the litter bag samples than in the environment (Figure 3.2, S2), suggesting that this taxon has a growth advantage during early litter colonization. Finally, we note that although the mesh size of the open bags was small, we cannot exclude the possibility that they allowed a greater number of grazers to colonize, which may have contributed to differences in bacterial composition between the dispersal treatments.

Conclusions

This experiment reveals that bacterial dispersal, like selection by the litter substrate, contributes to the diversity and composition of this bacterial community on grassland litter. Our study also demonstrates the feasibility of manipulating bacterial dispersal in the field, offering the potential to disentangle the processes contributing to

microbial community assembly (Hanson *et al.* 2012; Nemergut *et al.* 2013a). Future work might consider the impacts of dispersal differences across ecosystems for community assembly and functioning.

Acknowledgements

We thank Claudia Weihe for assistance with field sampling and laboratory methods. Thank you to Lucía Vivanco, Alexander Chase, Kendra Walters, Sydney Glassman, and John Dunbar for comments on earlier versions of the manuscript. This work was supported by a US Department of Education Graduate Assistance in Areas of National Need (GAANN) fellowship to MA and the US Department of Energy, Office of Science, Office of Biological and Environmental Research (BER), under Award Number DE-PS02-09ER09-25 to JM.

Tables and Figures

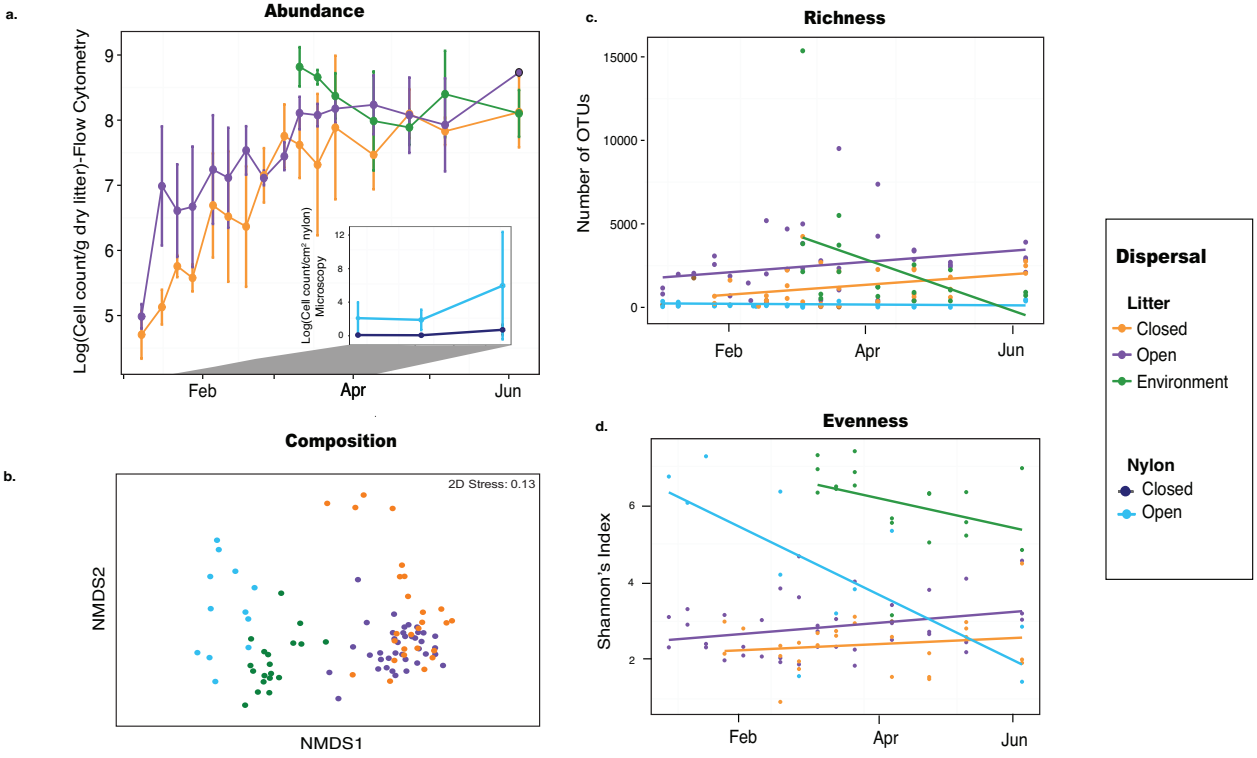


Figure 3.1 Effects of dispersal limitation on bacterial a) abundance, b) composition, as assayed by 16S sequencing, and c) richness and d) evenness, after rarefaction to standardize for sequencing effort among samples. Error bars in panel a) are 1 + 1SE. The four treatment types included litter-containing open bags (purple), litter-containing closed bags (orange), nylon-containing open bags (light blue), nylon-containing closed bags (dark blue). Plant litter samples from the surrounding environment (green) were also collected for comparison.

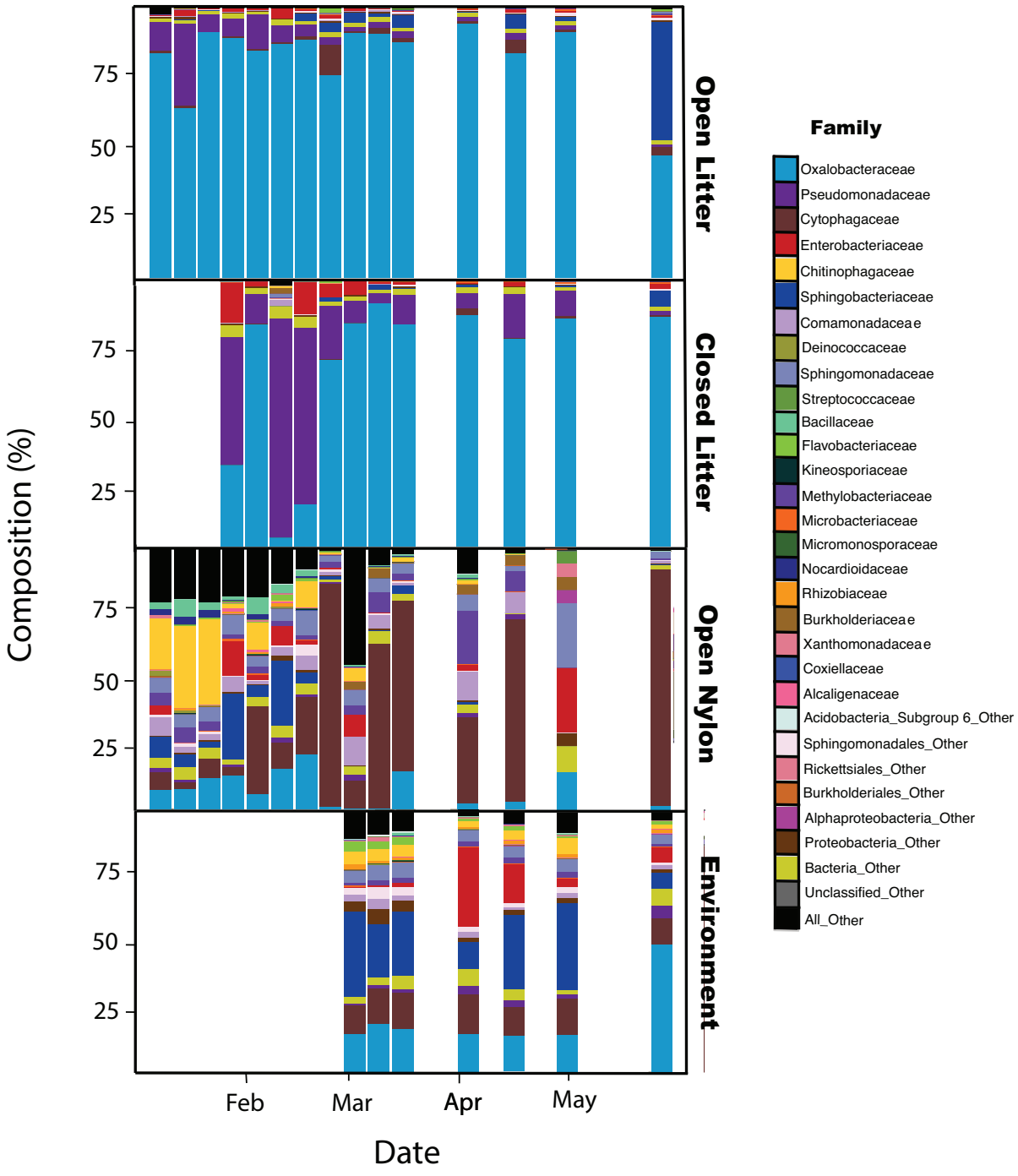


Figure 3.2 Relative abundance of bacterial families found in each treatment type (litter-containing open bags, litter-containing closed bags, nylon-containing open bags) and the surrounding plant litter (environment), averaged across each time point.

Supplementary Information

Materials and Methods

Field site

The field experiment was conducted at East Loma Ridge, Irvine CA, USA (33°44'N, 117°42'E, 365 m elevation). This southern California grassland ecosystem is dominated by exotic annual grasses and forbs (Potts *et al.* 2012; Kimball *et al.* 2014). Over the dry months, grasses die and accumulate as thatch cover. Microbes break down this material as a source of carbon and nutrients, primarily during the wet months (usually December thru March (Allison *et al.* 2013).

Dispersal manipulation

Litterbags (*n*=192) of 10 cm x 10 cm were constructed from nylon mesh. To manipulate dispersal, half of the bags were made from 18.0 µm nylon mesh (open bags; Tisch Scientific), which allows for the migration of bacteria and fungi, but not larger organisms. The other 96 litterbags were made from 0.22 µm nylon mesh (closed bags; Tisch Scientific), to prevent most microorganism dispersal. Both mesh sizes appeared to allow water to pass through similarly, as we saw no difference in the water content of litter in the closed and open bags, when collected for sampling over the course of the experiment (ANCOVA; main effect of dispersal treatment, $F_{3,89}$=0.40, p= 0.53).

Initial litter collection and sterilization

Surface litter was collected in November 2014 from the Loma Ridge site. The litter was ground using a blade coffee grinder (KitchenAid model BCG111OB) and

homogenized. The litter was autoclaved, then wetted with saline (0.9 M NaCl), left overnight and autoclaved again. Half of the open (*n*=48) and half of the closed (*n*=48) bags were filled with 6 g of sterilized litter. In the remaining open and closed bags, we placed a single sheet of 8 cm x 8 cm, 0.22 µm nylon. Nylon sheets were used as an inert substrate, on which microbes could land but not grow (or at least grow very slowly). In contrast, microbial growth could occur in litter-filled bags after initial dispersal. All bags were were subjected to 2.7 days of continuous exposure in a ^{137}Cs irradiator (UCI Radiation Oncology, School of Medicine, Irvine, CA) at a dose rate of 4Gy/min (~16000 Gy) with an additional estimated 6000 Gy picked up by samples over the exposure time due to scatter, leading to total dose of ~22 kGy.

Field experiment

Initially, 12 bags were collected as time 0 samples, and the remaining 180 bags were deployed at Loma Ridge, Irvine, CA on January 8th 2015. Bags were placed near one another (approximately 3 m x by 3 m area) to minimize environmental variation. Bags were collected weekly and then bi-weekly for 5 months, for a total of 16 samplings, including the time 0 samples. At each sampling, 3 replicates of each bag type was collected. In addition, starting in mid-March, 3 plant litter samples were collected from within the experimental area at each sampling time. Environmental data from the field site including precipitation (mm), surface soil moisture (vol/vol), and air temperature (degrees C) were monitored over the course of the field experiment (Figure S1). From each litter-containing bag, 0.1 g of plant litter was collected for bacterial counts and stored in 5-mL 1% phosphate buffered glutaraldehyde solution in the dark at

4˚ C. Similarly, 0.1 g litter was collected for DNA extraction and stored at -80˚ C. In addition, we obtained the water content of each litter bag at the time of sampling by weighing a litter subsample before and after drying in an oven at 60˚ C overnight. From each nylon-containing bags, the nylon sheet was cut in half, and the halves were stored in the same way as the litter samples.

Bacterial Cell Counts

Bacterial abundance on the litter samples was measured using flow cytometry by modifying a procedure from Allison et al. (2013). The fixed samples were extracted by adding 0.55 mL of 0.1 mol/L tetrasodium pyrophosphate and gently sonicating for 30 minutes in the dark at 4˚ C. Samples were then filtered through a 3.3-µm syringe filter to remove large particulates. 3 µL of SYBR Green (200x) was added to 600 µL of sample and incubated in the dark at room temperature for 10 min. Particle counts were performed using flow cytometry (BD Accuri C6; BD Biosciences, San Jose, California, USA). Each sample was run on the flow cytometer for 2 minutes at medium speed (45 µL/min) and 2000 threshhold (minimum fluorescence level). The gating parameters were optimized to count particle sizes in the size range of bacterial cells (tested by positive additions of single and mixed cultured isolates to sterilized plant litter). Cell counts are reported as number of stained counts.

The fixed nylon sheet samples were extracted and filtered in the same way as the litter samples. Low cell counts, combined with high background noise from nylon particulates, prevented us from using flow cytometry to estimate cell counts. Therefore, bacterial counts were assayed using fluorescence microscopy on a subset of samples

including 3 replicates of each bag type at 3 sampling dates (February 18, March 18, April 23). Microscopy slides were prepared by vacuum filtering 3 mL of the above processed samples on a filter. Filters were placed face down to incubate on a plastic petri dish containing 10 µL of SYBR green for 2 minutes in the dark and then mounted to a glass slide. Bacterial abundance was determined by counting stained cells for 12 randomly selected fields on each filter with a Axioplan2 Imaging microscope (Scientific Imaging Corporation, Campbell, CA, USA) using a fluorescein (FITC) filter at 100x objective.

DNA extraction and sequencing

From each bag, a 0.05 g litter sample was collected for microbial DNA extractions and stored in a -80˚ C freezer. DNA was extracted following with the MoBIO Soil Kit, using the low biomass DNA extraction protocol. Given the large differences in cell abundances in the samples over the time course of the experiment, either 5 µL of undiluted or a 1:10 dilutions of DNA from the extracts was used for subsequent PCR amplification. 5 µL of DNA from each extract was added to a cocktail containing: 1 Unit per reaction of Hot Start Taq DNA polymerase (BioLabs, Inc), 1 × PCR Rxn Buffer (-$MgCl_2$) (Invitrogen), 1200 µM $MgCl_2$ (Invitrogen), 200 µM dNTP, 0.2 µM Forward primer and 0.2 µM Reverse Primer, 200 mM Bovine Serum Albumin Acetylated (PROMEGA), and H_2O to a final volume of 25 µl. Dilutions were sample dependent because of low DNA quantities obtained from samples. We used the 515 forward primer (GTGYCAGCMGCCGCGGTAA) and 926 reverse primer (CCGYCAATTYMTTTRAGTTT) designed by (Caporaso *et al.* 2012) and modified by

(Apprill *et al.* 2015) to target the V4-V5 region of the 16S region. Following an initial denaturation step at 94 °C for 3 min, PCR was cycled 35 times at 94 °C for 45 s, 55 °C for 30s, 68 °C for 20s, with a final extension at 68 °C for 10 min. We amplified each subsample in duplicate from the extracted DNA.

All amplified samples were pooled based on gel pictures, with 1.0, 2.0, 3.0 µL added for strong, moderate, weak bands respectively, into a low binding tube. After pooling, PCR products were cleaned using the Agencourt AMPure XP PCR Purification Kit (Beckman Coulter Inc., Indianapolis IN, USA), following the standard manufacturer's instructions. We then performed a gel extraction on the pooled and cleaned samples to isolate the target band. Specifically, the cleaned PCR products were run on a sodium borate agarose gel at 100V for 45 minutes. The DNA was then gel extracted and cleaned using the Agencourt AMPure XP PCR Purification Kit (Beckman Coulter Inc., Indianapolis IN, USA). PCR products were assessed for quality using a High Sensitivity DNA Assay on an Agilent Bioanalyzer at the Genome High-throughput Facility at University of California, Irvine. Products were then sequenced at the University of California, Davis Genome Center at the DNA Technologies Core using multiplexed paired-end Illumina MiSeq platform. Presumably because of low bacterial abundance, we were unable to amplify DNA from the closed nylon-containing bags or from time 0 of any of the bags. Consistent with this, we only amplified samples from the closed litter treatment after 4 weeks in the field, when cell abundance reached an average of 1.1 x 10^7 per g dry litter (Figure 1a).

Illumina sequence data was processed using the QIIME (version 1.9.1) toolkit (Caporaso *et al.* 2010). Paired end files were joined and operational taxonomic units

(OTUs) were picked at 97% identity level using UCLUST (Edgar 2010) with the nearest neighbor method in QIIME. Taxonomy was assigned using SILVA v119 as the reference database (Quast *et al.* 2013) using QIIME scripts. We removed samples with <923 reads from further analysis. For the remaining samples (*n*= 104) reads varied between 923 and 163,485 with a median of 10,527. Using this data, 100 OTU-by-sample matrices were randomly subsampled at an even depth of 923 reads. This relatively low number was chosen for the rarefaction so as not to bias against samples from the early sampling points, as the low DNA yields from these samples led to these low number of reads. However, rarefaction curves suggest that the relative diversity among the treatments would not have changed with deeper sequencing (Figure S3).

We applied a square root transformation to each rarified matrix (to down-weight the influence of the most abundant OTUs). Next for each the transformed OTU-by-sample matrices, we calculated a sample-to-sample distance matrix using the Bray-Curtis beta-diversity metric in QIIME . A final median Bray-Curtis matrix was creating by selecting the median value for each cell across the 100 distance matrices. Shannon diversity (alpha-diversity metric) of all rarified OTU matrix libraries was also calculated using QIIME (Caporaso *et al.* 2010). Unprocessed sequences are available through the NCBI's Sequence Read Archive (accession number SRP102671).

Statistical Analysis

To test the trajectory of the abundance, richness, and evenness of microbial communities we used a generalized linear model to perform an ANCOVA. This model tested the effects of treatment, time, and their interaction (treatment x time) on bacterial

abundance and diversity (*lm* function, R Development Core (Team 2011). For abundance, the model compared only the dispersal treatment on the litter-containing bags (two levels of treatment), whereas for diversity it compared four levels of treatment including open litter-containing bags, closed litter-containing bags, open nylon-containing bags, and environmental samples. In addition, for bacterial abundance, we split the data into two time periods for analysis, January 5th to February 25th, when abundances were rapidly increasing, and March 5th to June 5th, when abundances of the open and closed litter bags converged and leveled off (Table S1).

To test for differences in bacterial community composition across sample groups, we performed a permutational multivariate analysis of variance (PERMANOVA) using PRIMER6 & PERMANOVA+ (Anderson 2001; Anderson *et al.* 2008), including treatment and time as fixed factors, and their interaction (treatment x time). Analyses were run using type III partial sums of squares under a reduced model with 999 permutations. Ad-hoc pairwise tests were run to compare each of the four sample groups. We then used SIMPER and PERMDISP analyses to compare the average similarity between and within treatments, respectively (PRIMER6 & PERMANOVA).

Lastly, we tested for correlations between the relative abundance of individual taxa between treatments. For this analysis, we first found the mean relative abundance of each taxon across all samples within each treatment (including the environmental samples). We then compared mean relative abundances of taxa between treatments using Pearson's correlations on cube root transformed data (*rcorr* function, Hmisc package R). For pairwise comparisons of treatments, only taxa found in both treatments were included in the analysis.

Supplementary Tables and Figures

Table S3.1. ANCOVA statistics for abundance, richness, and evenness. Abundance comparisons are for open and closed litter-containing bags. Richness and evenness comparisons are across all treatment types (open-litter containing, closed-litter containing, open-nylon containing, environmental litter).

	Abundance (Jan 5th- Feb 25th)			**Abundance** (March 5th- June 5th)		
Factor	F	P	df	F	P	df
Treatment	6.74	**0.01**	1,44	16.0	0.07	1,44
Time	7.05	**0.01**	1,44	41.4	**<0.001**	1,44
Treatment x Time	0.07	0.79	1,44	0.88	**0.04**	1,44

Richness			**Evenness**		
F	P	df	F	P	df
4.88	**0.003**	3,87	78.1	**<0.001**	3,87
0.63	0.43	1,87	2.53	0.11	1,87
5.82	**0.001**	3,87	11.4	**<0.001**	3,87

*Note: Significant p-values (<0.05) are shown in bold.

Table S3.2. PERMANOVA statistics and estimated multivariate components of variation for dispersal treatment (including environmental samples), time, and treatment-by-time interaction.

Factor	F	p	df	% Explained Variation
Treatment	8.26	**0.001**	3,53	26.2
Time	1.18	**0.019**	14,53	2.12
Treatment x Time	1.10	0.064	24,53	3.11

Table S3.3. Average similarity in composition between sample groups (SIMPER), ordered from highest to lowest similarity. All treatments were significantly different from one another (PERMANOVA: $p<0.01$).

Pairwise Comparison	Average Similarity (%)
Open Litter -- Closed Litter	70.4
Open Nylon -- Environment	50.4
Open Litter -- Environment	45.0
Closed Litter -- Environment	37.1
Open Litter -- Open Nylon	33.0
Open Nylon -- Closed Litter	28.5

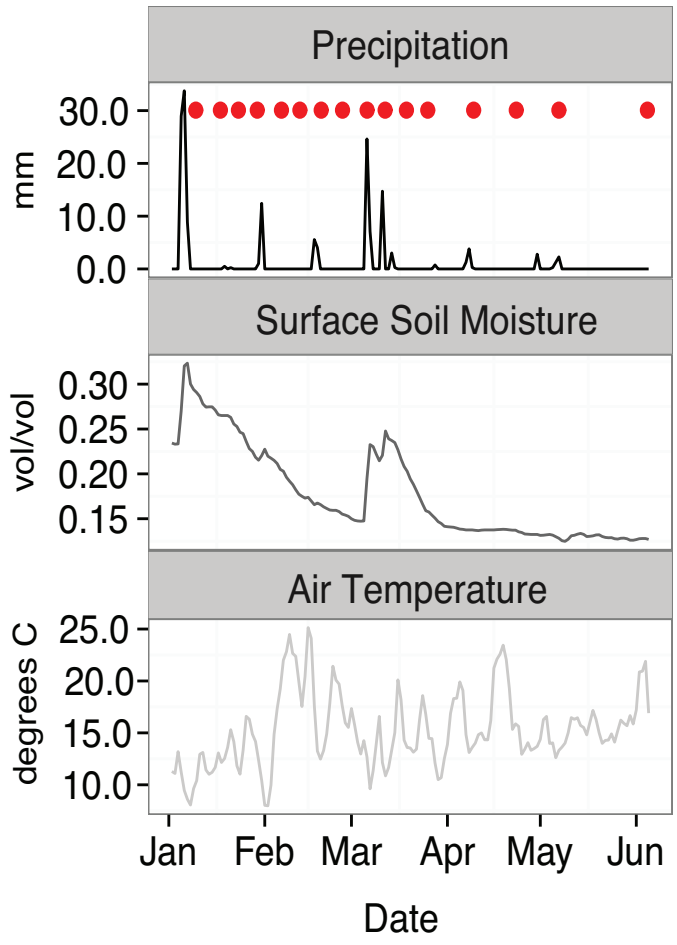


Figure S3.1. Precipitation (mm), surface soil moisture (vol/vol), and air temperature (degrees C) over the course of the field experiment, with bag collection dates denoted by red dots in the top panel. East Loma Ridge weather station data courtesy of Center for Environmental Biology, School of Biological Sciences, University of California, Irvine: http://128.200.14.200/index.html#.

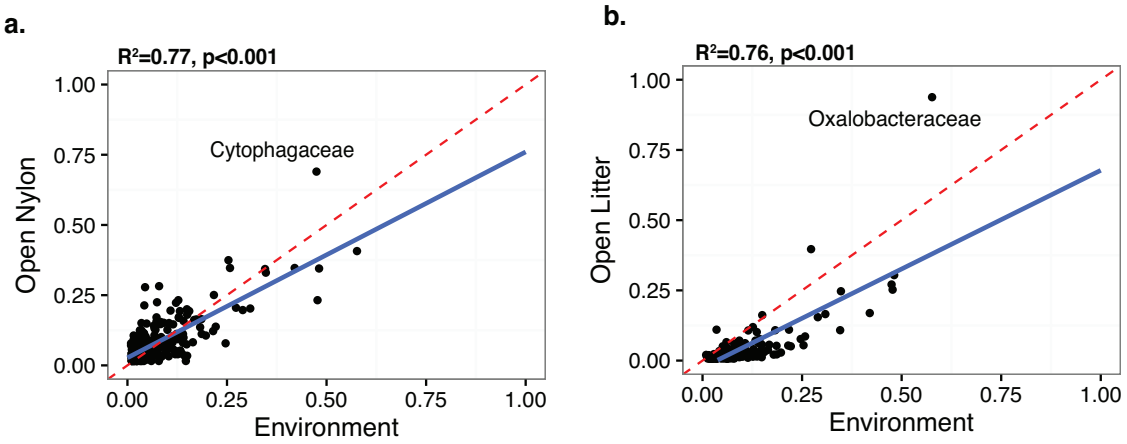


Figure S3.2. Correlation between relative abundance of individual taxa (OTU level) occurring in both sample types, for a) environmental litter versus open nylon-containing bags (*n*=301) and b) environmental litter versus open litter-containing bags (*n*=189). Relative abundances of taxa in each sample type represent the mean across sampling dates. The plots and Pearson's correlations are based on normalized data, $\sqrt[3]{(relative\ abundance)}$. Blue solid lines show a linear regression fit, while the red dotted lines illustrate a 1:1 correspondence.

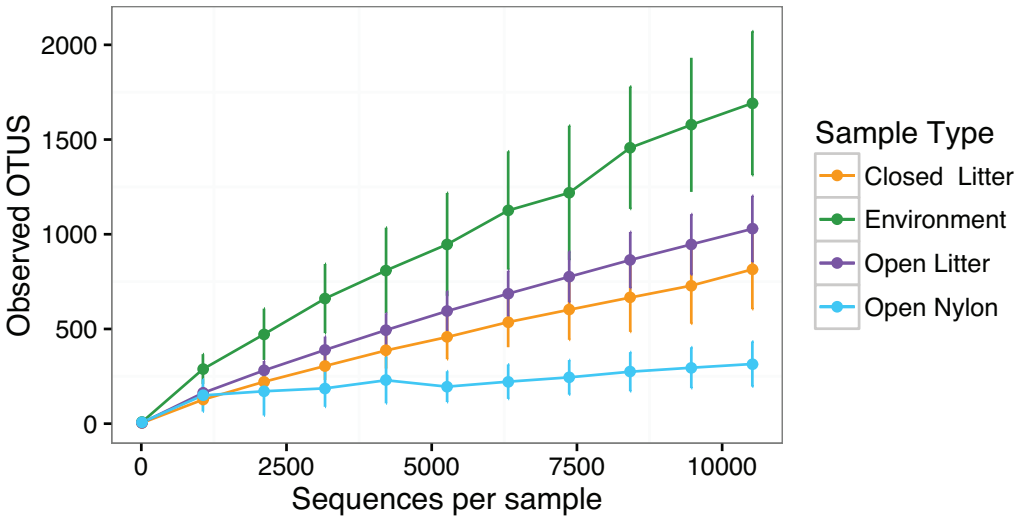


Figure S3.3. Rarefaction curves of bacterial richness (observed OTUs ± sd) for nylon-containing open bags (light blue), litter-containing open bags (purple), litter-containing closed bags (orange), and plant litter samples from the surrounding environment (green).

CHAPTER 4

Quantifying stochastic variation in taxonomy and functioning

ABSTRACT

It is increasingly recognized that both stochastic and deterministic processes drive community composition. However, quantifying the role of stochastic processes in shaping communities remains a challenge. Here we used a field experiment to address the following questions: 1) Is there stochastic variation in community composition? 2) Does stochastic variation in bacterial communities translate to stochasticity in functioning? 3)Does dispersal rate alter stochasticity. To address these questions, we performed a field experiment focusing on plant litter microbial communities. We homogenized plant litter substrate and irradiated it in replicate litterbags of the two mesh sizes (open vs. closed to microbial dispersal). We inoculated the litterbags with homogenized microbial communities and placed them in the field in close proximity to limit environmental variation. We next manipulated an environmental (deterministic) parameter, water availability (ambient vs. added water). Our sampling design allowed us to quantify how variance in measures of taxonomic composition and functioning were distributed among our treatments (dispersal and water addition), stochasticity (bag replicates), and residual variation (measurement error). We found that stochastic processes contributed to a significant amount of variation in taxonomic composition. Furthermore, effects of stochasticity in community composition translated into stochasticity in functional parameters. Ultimately, the ability to accurately quantify stochastic processes is paramount to determining the predictability of ecosystem

structure and functioning, whether focused on bacteria that degrade plant litter, microbes in the human gut, or patterns of global biodiversity.

Significance

Assessing the role that stochastic processes— processes probabilistic in terms of species identity— play in shaping communities remains a challenge. This study is the first field experiment that directly attempts to disentangle the role of stochastic processes, environmental selection, and dispersal on microbial community composition and functioning. As such, we minimized overall environmental variation and starting community composition, to create conditions where we might be most likely to observe the effects of drift. This is in contrast to previous observational studies investigating the impact of stochastic and deterministic processes on communities performed in the context of spatial and temporal gradients which emphasize the strength of selection processes and where the effects of drift will more likely go undetected. We further provide the first estimate of stochastic variation in community composition and functioning that accounts for measurement error.

Introduction

A central goal in ecology is understanding the mechanisms that underlie patterns in community diversity. Much of the variability among communities is thought to arise deterministically, from environmental factors and biotic interactions that select for predictable differences in taxonomic composition, or beta-diversity. In contrast, stochastic variability emerges from random differences in replication, death, mutation,

and dispersal among individuals in a community (Hubbell 2001; Vellend 2010; Chase & Myers 2011). The role of dispersal in ecological communities is particularly complex, as it may contribute to both stochastic or deterministic variation. On the one hand, dispersal rates moderate the role of ecological drift, defined as stochastic fluctuations in the relative abundance of organisms in a community (Hubbell 2001; Vellend 2010); high dispersal rates are thought to limit drift, as migration between locations can homogenize variation in species composition among locations. On the other hand, species may differ in their dispersal ability and thereby contribute to deterministic variation among communities.

In microbial communities, extensive evidence suggests that stochasticity influences beta-diversity in both free-living and host-associated communities (Hao *et al.* 2016; Adair & Douglas 2017; Vega & Gore 2017). This evidence is derived from both observational studies and field experiments that track changes in composition across spatial and temporal gradients, while measuring numerous deterministic, potentially selective biotic and abiotic variables (Bell 2010; Ferrenberg *et al.* 2013; Lee *et al.* 2013; Brown & Jumpponen 2014; Louca *et al.* 2016b). Statistical methods, including null-deviation models (Ferrenberg *et al.* 2013; Brown & Jumpponen 2014; Dini-Andreote *et al.* 2015) or variance partitioning (Dumbrell *et al.* 2010; Hanson *et al.* 2012; Langenheder *et al.* 2012), are then used to infer the relative importance of deterministic and stochastic processes. However, these studies likely overestimate the importance of stochasticity because of unmeasured deterministic variables and measurement error (Bell 2010; Evans *et al.* 2017). Here we define measurement error as variability due to spatial heterogeneity within a sampled community as well as technical error in

estimating community diversity. Both of these factors may be particularly important for microbial communities, which are highly under-sampled and assayed through molecular methods that introduce variability (e.g., variability added during DNA extraction, PCR amplification, and sequencing) (Hughes *et al.* 2001; Schloss *et al.* 2011; Brooks *et al.* 2015); however, such error will also be present in studies of larger organisms if diversity is assayed on a subset of the community (Gotelli & Colwell 2001).

Ultimately, however, the consequences of such stochasticity depend on whether this variability translates into variation in functioning at the ecosystem level. It is often argued that stochastic variation within communities will attenuate at the ecosystem level, because many microbial taxa perform common functional processes (e.g. (Burke *et al.* 2011; Louca *et al.* 2016a)). Alternatively, compositional stochasticity might lead to variation in ecosystem functioning, particularly for functions performed by a limited number of taxa (Schimel 1995). This has been demonstrated for microbial communities in a number of systems including including in bioreactors (Zhou *et al.* 2013), on decomposing plant material (Dickie *et al.* 2012), and in plant nectar (Vannette & Fukami 2017). Thus, how stochasticity translates from the community to ecosystem level may depend in part on the choice of functional metric.

Here, we quantified the importance of stochasticity for both composition and functioning of a natural microbial community on decomposing plant litter in a southern California grassland. The composition of microbial decomposers is known to influence the rate and quality of decomposition at this site (Allison *et al.* 2013; Martiny *et al.* 2017) and in other plant litter systems (Tiunov & Scheu 2005; Strickland *et al.* 2009). Compared to previous studies in the field, we made a concerted effort to minimize

biological and environmental heterogeneity within and between replicates, while also accounting for measurement error. To do this, we manipulated the communities within litterbags, which provided a convenient spatial scale for replication and allowed us to measure litter decomposition rates (mass loss over time). We then assayed an array of diversity and functioning metrics on three subsamples from each litterbag replicate. We hypothesized that measurement error would account for a large fraction of otherwise unexplained community variation (H1); however, any remaining variation among replicates we could then attribute to stochastic effects. We further hypothesized that stochastic variation would be higher for community composition than for functional processes, attenuating from the community to the ecosystem level (H2).

Finally, we investigated if the rate of microbial dispersal altered these stochastic effects by varying the mesh size of the nylon bag material. We hypothesized that reduced dispersal would increase the effect of stochasticity via ecological drift (H3). For comparison, we also manipulated an environmental parameter, precipitation, known to influence the microbial community in this system (Matulich *et al.* 2015). Thus, our full experimental design included two treatments: dispersal (open versus closed litterbags) and precipitation (ambient versus added water). Each litterbag was initially filled with a ground, irradiated litter substrate that was re-inoculated with a homogenized microbial community to minimize variation in substrate resources and initial community composition. We limited abiotic variation by placing the litterbags in close proximity (within $1m^2$) in the field. The four treatments (ambient-open, ambient-closed, precipitation-open, and precipitation-closed) were replicated eight times, while subsamples from within each litterbag at the end of the experiment allowed us to

quantify stochastic effects and importantly distinguish these effects from measurement error.

Results

Overall treatment effects

The composition of bacterial communities in the litterbags was similar to earlier studies of the natural plant litter at this site (Figure S4.1)(Matulich *et al.* 2015). The same phyla, including, Actinobacteria, Proteobacteria, and Bacteriodetes, were dominant, and the composition observed in the 16S rRNA and metagenomic data were similar (Figure S4.1, S4.2).

The dispersal and precipitation treatments had minimal effects on bacterial abundance and alpha diversity. On average bacterial cell densities were ~9 x 10^8 cells per gram of dry litter, and this density did not vary across treatments (ANOVA: $p>0.05$; Figure S4.3a). Fungal abundance, as assessed by hyphae counts, also did not differ (ANOVA: $p>0.05$; Figure S4.3b). As with abundance, the number of OTUs observed did not differ significantly among treatments (ANOVA: $p>0.05$; Figure S4.4a) and was similar to the richness observed (3320±2453 OTUs) in the initial inoculum. Likewise, evenness was similar among treatments (Figure S4.4b), although it did decline dramatically from the inoculum (Tukey's test; $p<0.001$).

Dispersal did alter bacterial composition as assessed by 16S rRNA amplicon sequencing (Figure 4.2a; PERMANOVA: $p=0.001$; Table S4.1a). Dispersal also interacted significantly with precipitation to affect composition ($p = 0.029$), whereas added precipitation had no effect (Table S4.1a). Finally, composition also varied

significantly among replicate litterbags within a treatment, as tested using the nested design of the experiment (p=0.001; Table S4.1a). Very similar results were observed whether composition was assessed by 16S or metagenomic sequencing, although in this latter case, the dispersal-by-precipitation interaction was not significant (Table S4.1b).

Stochasticity influences taxonomic composition

We next asked how much of the observed variation in bacterial taxonomic composition could be attributed to each of the significant treatments described above. Although the dispersal treatment was highly significant, this effect explained a relatively small proportion of total compositional variation (4.5%; Figure 4.1a). Further, the dispersal-by-precipitation interaction explained only a small additional amount of compositional variation (2.5%).

In contrast, variability among replicate litterbags within a treatment accounted for 16.5% of total variation, or more than twice as much as all treatment effects combined (Figure 4.1a, Figure S4.8). This component of variation is our estimate of stochastic variation among the communities. Still, after accounting for treatment effects and stochasticity, approximately three quarters (76%) of the estimated variation in taxonomic composition remained unexplained and thus, can be attributed to measurement error (a combination of technical error and spatial heterogeneity within the litterbags).

Similar trends were observed when taxonomic composition was assayed through metagenomic sequencing (Figure 4.1a). Stochastic variation among bacterial

composition was estimated to be 17.5% (versus 16.5% by 16S) and measurement error 74% (versus 76%). However, it is worth noting that these estimates of variation are reported as relative proportions. Thus, the average compositional dissimilarity between any two samples was twice as high as when the community was assayed by 16S sequencing than when assayed by metagenomic sequencing. Thus, total variability was higher for the 16S data, but the relative amount of that variability that could be assigned to stochasticity (or the treatments or measurement error) was similar for the metagenomic data (Figure 4.1b).

Stochasticity influences functional processes

To test whether the stochastic variation in taxonomic composition translated into stochastic variation in functional processes, we measured community functioning in three ways: 1) functional potential from metagenomics sequences to characterize differences among litterbags; 2) potential extracellular enzyme activity (EEA) as a metric of overall community functioning; and 3) mass loss of the litter to assay decomposition rate. To assess functional potential, we annotated metagenomic sequences from each sample using the protein family (Pfam) database (Finn *et al.* 2016). We then used the Pfam annotations to create a Bray-Curtis dissimilarity matrix as for the taxonomic data (see Supplementary Methods). Thus, these data provide information on the composition of the genetic functional potential of the bacterial communities.

As with taxonomic composition, dispersal and, to a lesser extent, an interaction between dispersal and precipitation significantly altered the functional potential of the litter communities (Figure 4.1a, Table S4.1c). Similarly, the distribution of the variability

assigned to treatment effects versus stochasticity (among bags) or measurement error (residual, or within bags) was almost identical to that observed for taxonomic composition. Treatment effects accounted for only 7.6% of variation in genetic functional potential, while 18.3 % was attributed to stochasticity and the remaining 74.3% to measurement error. It is again worth noting that despite the similarity in proportional assignments of variability, the average dissimilarity between samples in its functional potential was even lower still than metagenomic taxonomic composition; functional potential between two samples varied by only an average of 12% (Figure 4.1b). A similar average and a similar variability due to stochastic effects was observed even when we subset the data and only considered genes involved either in carbohydrate degradation (glycoside hydrolases and carbohydrate binding modules) or in the nitrogen cycle (data not shown).

We measured the potential activities of seven extracellular enzymes including, α-Glucosidase (AG), acid phosphatase (AP), ß-Glucosidase (BG), ß-Xylosidase (BX), cellobiohydrolase (CBH), leucine aminopeptidase (LAP), and N-acetyl-ß-D-glucosaminidase (NAG). The individual enzyme activities varied somewhat in their response to the dispersal and precipitation treatments (Table S4.2; Figure S4.5) although generally, higher activity levels were observed in the open-ambient and closed-precipitation treatments (Figure S4.5). Stochastic effects contributed to between 0 and 35% of estimated variation for the individual EEA measurements, with the lowest stochasticity observed in BG and the highest in NAG (Figure S4.8). We next created a composite metric of EEA by calculating the Euclidean distance between each pairwise sample based on normalized measurements of the seven assays. Here, we saw a

significant effect of the precipitation-by-dispersal interaction on EEA composition among the litterbags (PERMANOVA: p=0.007; Figure 4.1b; Table S4.3). Treatment effects explained 27% of the estimated variation in EEA, while stochastic effects accounted for an additional 27% (Figure 4.1a). Measurement error was lower than for community composition, accounting for 46% of the estimated variation.

Dispersal significantly affected the mass loss within each litterbag over the course of the experiment, such that open litterbags lost more mass than closed litterbags (ANOVA; dispersal treatment: $F_{1,28}$=12.356, p=0.002, Figure S4.3c). Thus, dispersal and its interaction with precipitation accounted for 47% of the estimated variation in mass loss among litterbags. The remaining 53% of variation is a combination of among bag (stochastic) variation and measurement error (Figure 4.1a). Because we could not measure mass loss for subsamples within a litterbag, these two sources of variation cannot be separated.

Variation in stochasticity between treatments

Finally, we tested whether the degree of stochastic variation differed among the dispersal or precipitation treatments. In fact, beta-diversity within a treatment group (mean distance to the centroid) of replicate communities did not vary significantly between treatments (PERMDISP; p=0.818; Figure 4.2c). However, the degree of stochastic variation in the EEA closed-precipitation treatment was higher than either the closed-ambient or open-ambient treatments (PERMDISP; p=0.004 & p=0.01, respectively). Furthermore, the pattern suggesting an interactive effect between the precipitation and dispersal treatment was qualitatively similar to that observed for both

community composition (4.2b and 4.2c) and genetic functional potential (Figure 4.2f).

Discussion

Effects of dispersal and precipitation on community diversity and functioning

In this study our goal was to disentangle the role of environmental selection, dispersal, and stochastic processes on microbial community composition and functioning. In fact, our precipitation treatment had unexpectedly little effect on community composition and functional metrics. Previous work at this field site has shown small but significant impacts of drought on bacterial community composition on plant litter (Allison *et al.* 2013). Our rain water addition may not have been frequent enough to overwhelm the effects of the natural precipitation events, and/or may have been too short-lived due to rapid evaporation in this sunny and arid climate (Figure S4.7).

In contrast, differences in dispersal rate in open versus closed bags led to differences in bacterial community composition and functioning. Dispersal rate altered community composition contributing to a small portion (4.5%) of estimated variation across samples. This altered composition under different dispersal rates may be mediated through changing biotic interactions and/or priority effects (Thompson & Gonzalez 2017; Vannette & Fukami 2017). Furthermore, higher mass loss in open versus closed bags may be a result of faster colonization or differences in the microbial taxa able to disperse into the litterbags (Bradford *et al.* 2002). Although we could not detect differences in bacterial or fungal abundance due to the dispersal treatment at the end of the experiment, it is likely that the open bags were colonized more quickly and

therefore had higher abundances early on in the experiment, as shown in a recent study measuring dispersal onto plant litter in this system (Albright & Martiny *in Revision*).

An interaction of the dispersal and precipitation treatments, in fact, created communities that differed in their composition and functioning. As mentioned above, while alone the precipitation treatment did not impact communities, we did see an effect of dispersal on composition. Thus, the precipitation treatment may have altered composition in the dispersal treatments in different and perhaps opposing ways. For example, maybe immigration into the open bags allowed for communities to respond to the added precipitation more quickly. Additionally, for the precipitation treatment we added rain water collected earlier in the season (see Supplementary Methods), this may have have contributed to dispersal of different additional bacterial taxa into the open versus closed bags, which again interacting with differences in dispersal rates, may have altered competition and successional dynamics.

Measuring stochastic effects on community diversity

Overall, our results support previous studies of macro- and micro- organisms that have highlighted the importance of stochastic processes in contributing to beta-diversity, or variation in community composition (Cottenie 2005; Hanson *et al.* 2012), while making crucial steps towards robustly quantifying these processes. Two key factors in our experimental approach allowed for this more robust quantification. First, rather than using statistical inference to account for the influence of environmental variation on beta-diversity, we manipulated the communities and environment to limit the influence of this variation. While removing all variability would be impossible, we purposely

homogenized microbial communities and limited environmental variability across replicate litterbags. This allowed us to better assume that variation not explained by treatment effects are likely due to stochasticity. Second, taking subsamples from each replicate allowed us to directly measure variability across the replicate bags and therefore tease apart the influence of measurement error (error in assessing composition in a replicate due to technical and spatial variability). Indeed, we show that without measuring this error, we would have overestimated the contribution of stochasticity by 1-4 times.

After accounting for across bag effects, the remaining residual variation, or within bag variability, then captures the measurement error. This measurement error could be due to spatial variability within each litterbag or could be introduced during subsequent sample processing for example during DNA extraction and amplification, or for metagenomes during the bioinformatics and short-read processing (Hughes *et al.* 2001; Brooks *et al.* 2015). Measurement error accounted for a large portion, up to ~75%, of the variation in community composition (H1). Without this quantification, this variation might be mistakenly attributed to stochasticity, highlighting the importance of taking into account this variability, present to some extent in all ecological studies. Given that measurement error accounted for such a large portion of the variability in composition across communities, further studies are needed to partition the source of this measurement error between spatial variability from sampling and technical error during sample processing. However, despite the high measurement error, importantly, with this sampling design we are able to distinguish between stochasticity and measurement error.

In this study stochasticity played a larger role, in shaping community composition than environmental effects (Figure 4.1a). The combined effects of the precipitation, dispersal, and precipitation-by-dispersal interactions explained a relatively small portion of estimated variation in bacterial community composition. There are a number of factors that likely contributed to this higher stochastic variation, and as other studies have pointed out, the relative influence of these factors should change under different conditions (Ferrenberg *et al.* 2013; Dini-Andreote *et al.* 2015). In particular, the precipitation treatment had an unexpectedly negligible effect on bacterial composition. A stronger environmental treatment, or perhaps running the experiment longer, would have likely changed the relative importance of stochastic effects (Chase & Myers 2011). Furthermore, the litterbags contained irradiated litter that was inoculated at the beginning of the field experiment; low initial community size relative to habitat size is predicted to enhance ecological drift (Hubbell 2001). Thus, we consider the estimates here to be an upper bound for the amount of that stochastic variation that is likely to be observed in other, natural bacterial communities. We also acknowledge that this estimate still includes effects that are feasible uncontrollable (micro-environmental variation between replicates) and not stochastic, but highly unpredictable (Huisman & Weissing 1999, 2001).

Stochasticity in composition translates to functioning

We hypothesized that stochastic effects observed at the community-level would translate to our measured functional metrics, but be relatively less important (H2). Contrary to these expectations, the relative fraction of stochasticity observed the

functional metrics was similar and possibly even higher than for composition (Figure 4.1a, Figure S4.8). The relative contribution of stochastic effects to variation in metagenome taxonomic versus functional potential among communities was similar. However, an important point here is that these stochastic effects are measured in relative terms. As found in numerous other previous studies (Burke *et al.* 2011; Louca *et al.* 2016a; Louca *et al.* 2016b), average variation across samples was lower for functioning than composition (Figure 4.1b), which is likely because metrics for functioning are often coarser than those for composition. Thus, our measurement of stochasticity for functional potential was taken across the relatively small dissimilarity in genetic functional potential among the communities.

Stochasticity in community composition also translated into stochasticity as assessed by EEA potential and was in fact even higher for the composite EEA potential than for community composition. One possible mechanism driving the functional variation could be that variation in community composition alters the biotic interactions in the litterbags leading to higher variability in functioning. Regardless of the mechanism, the fact that a significant portion of observed variation in EEA potential may be due to stochastic effects is important to consider. Extracellular enzymes often control rate limiting steps in organic matter turnover, and thus shifts in enzyme activity could have major consequences for carbon cycling (Allison *et al.* 2011). However, calculating stochasticity of EEA potential in different ways, for example by instead using an average value of stochastic effects across the seven individual enzymes, or measuring a different subset of enzymes, would have resulted in different conclusions as to the relative amount of stochasticity in community versus functional metrics. This possibility

is highlighted by the large range in the contribution of stochastic effects to individual potential EEA, where in fact, in some cases stochasticity in functioning might be lower than for community metrics (Figure S4.8). Thus, overall, we see a different perspective of the influence of stochasticity on community taxonomic composition versus functioning depending on the choice of functional metric. This highlights that traits of the functional or community metric of choice may influence the the outcome of comparisons between taxonomy and functioning. Despite this uncertainty, regardless of the method of calculating variation in functioning, we would have observed some measureable portion of this variation due to stochastic effects.

Impact of dispersal and precipitation on stochasticity

Finally, in contrast to our hypothesis (H3), dispersal rate did not consistently alter stochastic effects on community composition or functioning. If anything, we only detected a trend that dispersal interacted with precipitation to influence the degree of stochasticity. Across all assays at both the community- and ecosystem- levels, replicates in the closed precipitation were more variable than the open precipitation treatment. We speculate that water addition may have allowed for more turnover (growth and death) in the litterbags over the course of the experiment, and thus allowed for greater ecological drift of the closed communities (Nemergut *et al.* 2013b). Supporting this hypothesis is the observation that stochasticity was also generally higher in the water addition bags compared to the ambient bags, even in the litterbags with higher dispersal. Alternatively, bacteria dispersing into the bags from the rainwater addition may have increased beta-diversity due to stronger priority effects under

increased dispersal (Vannette & Fukami 2017).

Conclusions

By minimizing initial differences in the bacterial communities and the environment, and taking replicate measurements at multiple levels, we successfully quantified the contribution of stochastic effects across multiple levels of biodiversity in a natural environment. This field experiment is parallel to powerful population level laboratory experiments with bacteria investigating the impact of drift on genetic variation (Travisano *et al.* 1995). We demonstrate that previous studies likely greatly overestimate the importance of stochastic effects on microbial communities, because they do not consider measurement error among their replicates. Still, we demonstrate that stochasticity can be a major component of variation in bacterial composition and its functioning - in our case, often twice as strong as other measured deterministic treatments. Overall, this study is a step towards better assessing the predictability of ecosystem structure and functioning, which requires the ability to robustly quantify the roles of stochastic and deterministic processes.

Tables and Figures

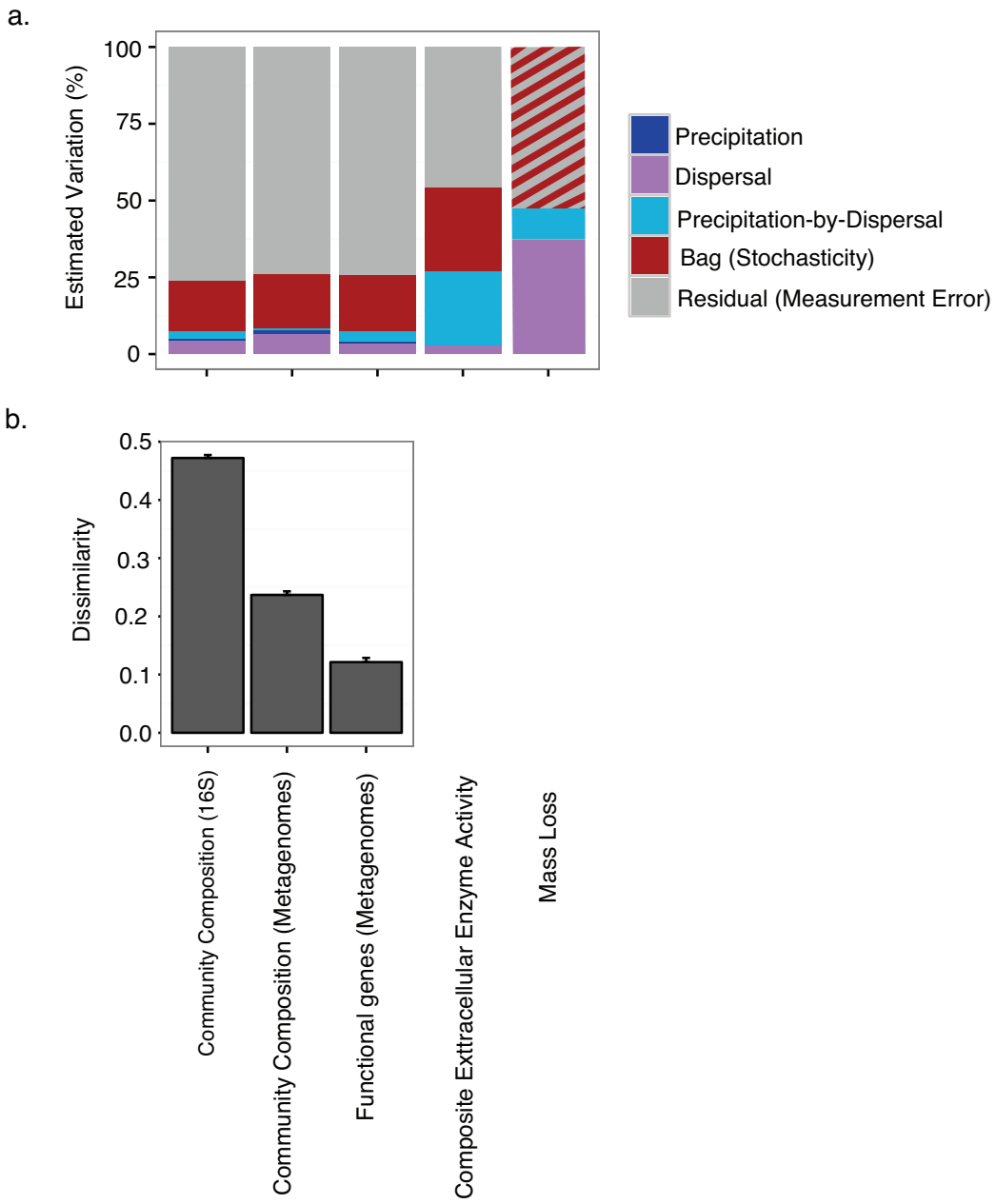


Figure 4.1. a) Percentage of estimated variation across a variety of community composition and ecosystem function metrics explained by the environment, dispersal, environment-by-dispersal, within bag (measurement error), and residual (stochastic effects). For community composition and extracellular enzyme activity estimates were derived from a permutational MANOVA model, and for mass loss a two-way ANOVA. b) Average Bray-Curtis dissimilarity between samples for for 16S and metagenomic sequence data.

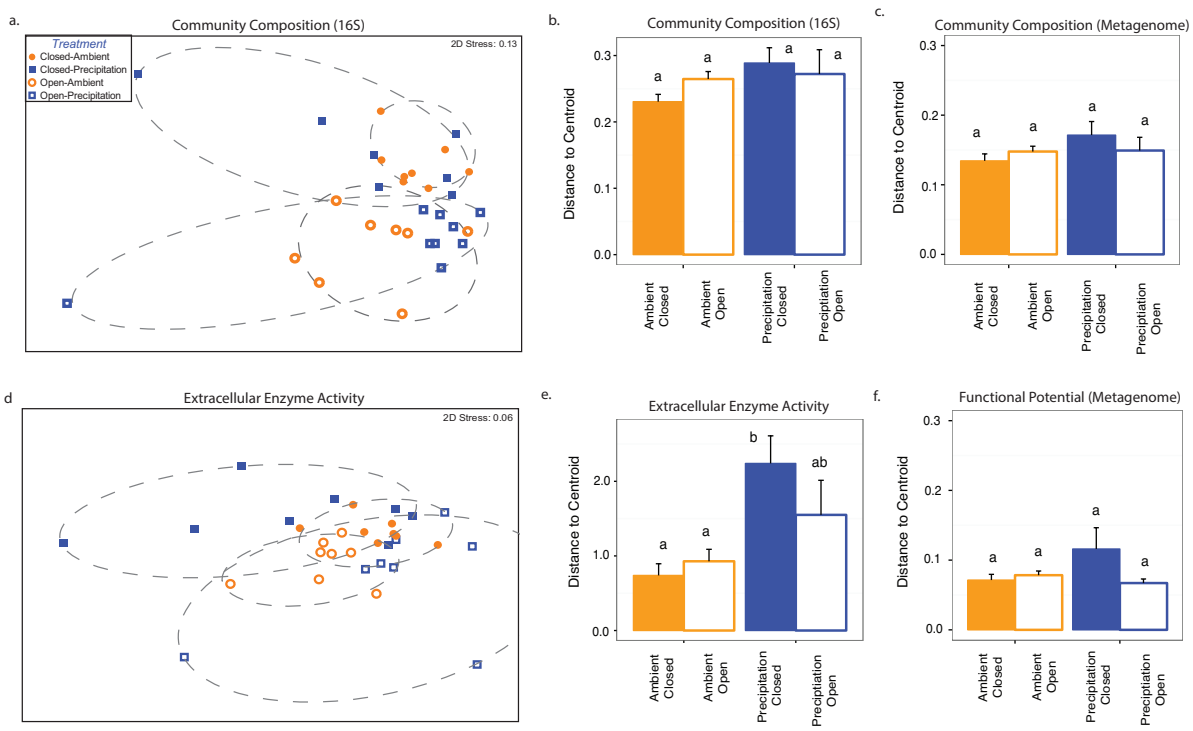


Figure 4.2. Nonmetric multidimensional scaling (NMDS) ordinations showing variability in a) 16S bacterial community composition (Bray-Curtis dissimilarity) and d) extracellular enzyme activity (Euclidean distance). Each point represents the median of the 3 replicates from the bag (n=32). Within-group distance for b) 16S bacterial community composition (Bray-Curtis dissimilarity), c) metagenome bacterial community composition e) extracellular enzyme activity (Euclidean distance), and f) functional genetic potential of replicate litterbags within treatments (closed-ambient, closed-precipitation, open-ambient, open-precipitation).

Supplementary Information

Materials and Methods

Field Experiment

The field experiment was conducted in a grassland at Loma Ridge in Irvine, CA (33°44'N, 117°42'E, 365 m elevation). The site has a semi-arid Mediterranean climate, with mean annual precipitation of 325 mm, most of which occurs between October and April and is dominated by non-native annual grasses (*Bromus diandrus, Avena fatua*)(Potts *et al.* 2012; Kimball *et al.* 2014).

We placed 32 litterbags into the field to assay the assembly of plant litter microbial communities. The litterbags (10 cm x 8 cm) were constructed from nylon mesh. Half of the litterbags were made from 18.0 um mesh (Tisch International), which allows for the dispersal of bacteria and fungi into and out of the bags. We refer to these as "open" litterbags. The other half were "closed" litterbags, made from 0.22 um nylon mesh (Tisch International) to prevent bacterial and fungal dispersal. Both litterbag types allowed water to pass through similarly (Albright & Martiny *in revision)*.

Surface leaf litter was collected in November 2014 from several spots across the field site to capture a mix of plant species. The litter was ground using a blade coffee grinder (KitchenAid model BCG111OB). To sterilize the litter, we used both autoclaving and gamma irradiation. The litter was autoclaved, then wetted down with 0.9 M NaCl, left overnight, and autoclaved again. Next, 4.7 grams of litter was added to each of the 40 litterbags, and the filled bags were gamma irradiated with > 22 kGy gamma irradiation. We reinoculated all the sterile litterbags with 0.12 grams of ground and homogenized inoculum litter (freshly collected in January 2015).

The inoculated litterbags were deployed at Loma Ridge on January 8th 2015, to coincide with most of the annual rains and litter decomposition. To minimize environmental variation, the litterbags were placed in close proximity to one another (within $1m^2$). Precipitation was manipulated by adding 120 mL of rainwater to half of the open and closed litterbags at 5 time points during the 5 months the bags were in the field. This volume was equivalent to half of the rainfall of the first storm of the season, and was slightly higher than the rainfall of each subsequent storm during the experiment (Figure S4.7). This volume completely soaked the litter inside the bags. Rainwater for the experiment was collected in December 2014 and stored at 4C until use. Litterbags were collected on June 5th 2015. Upon collection, we weighed the litter remaining in each bag and dried a subsample at 60˚ C to obtain dry mass. Mass losses are reported as the percentage loss of initial dry mass.

DNA extraction and 16S sequencing

All assays (except mass loss) were performed on eight inoculum samples and three subsamples from each litterbag to assess measurement error. From each bag, three 0.05 g litter subsamples were collected for microbial DNA extractions and stored in a -80˚ C freezer. Overall the 104 samples, included 96 final collection samples (4 treatments x 8 bags x 3 samples per bag) and 8 initial inoculum samples. DNA was extracted following the FastDNA Spin Kit for Soil (MP Biomedicals, LLC) protocol, with two modifications. 1) After the addition of sodium phosphate and MT buffer, samples were subjected to a freeze-thaw cycle three times, immersing the samples for 30 second in liquid nitrogen and 3 minutes in a 60 C water bath and 2) bead-beating was

done in a FastPrep FP120 (Bio101, Vista, CA, USA) at 5.5 m s^{-1} for 45 s.

To PCR amplify the bacterial 16S rRNA gene, 5 uL of a 1:50 dilution of DNA (average 2.95±0.88 ng DNA) from each extract was added to a cocktail containing: 1 Unit per reaction of Hot Start Taq DNA polymerase (BioLabs, Inc), 1 × PCR Rxn Buffer (-$MgCl_2$) (Invitrogen), 1200 uM $MgCl_2$ (Invitrogen), 200 uM dNTP, 0.2 uM Forward primer and 0.2 uM Reverse Primer, 200 mM Bovine Serum Albumin Acetylated (PROMEGA), and H_2O to a final volume of 25 µl. We used the 515 forward primer (GTGYCAGCMGCCGCGGTAA) and 926 reverse primer (CCGYCAATTYMTTTRAGTTT) designed by (Caporaso *et al.* 2012) and modified by (Apprill *et al.* 2015) to target the V4-V5 region of the 16S gene. Following an initial denaturation step at 94 °C for 3 min, PCR was cycled 35 times at 94 °C for 45 s, 55 °C for 30s, 68 °C for 20s, with a final extension at 68 °C for 10 min. We amplified each subsample in duplicate from the extracted DNA.

All amplified samples were pooled based on gel pictures, with 1.0, 2.0, 3.0 uL added for strong, moderate, weak bands respectively, into a low binding tube. After pooling, PCR products were cleaned using the Agencourt AMPure XP PCR Purification Kit (Beckman Coulter Inc., Indianapolis IN, USA), following the standard manufacturer's instructions. We then performed a gel extraction on the pooled and cleaned samples to isolate the target band. Specifically, the cleaned PCR products were run on a TAE agarose gel at 80V for 1 hour. The DNA was then gel extracted and purified using the standard Zymoclean™ Gel DNA recovery Kit protocol (Zymo Research Corp). PCR products were assessed for quality using a High Sensitivity DNA Assay on an Agilent Bioanalyzer and quantified (10.7 ng/uL) on a Qubit at the Genome High-throughput

Facility at University of California, Irvine. Products were then sequenced at the University of California, Davis Genome Center at the DNA Technologies Core using multiplexed paired-end Illumina MiSeq platform.

Illumina sequence data was processed using the QIIME (version 1.9.1) toolkit (Caporaso *et al.* 2010). Paired end files were joined and operational taxonomic units (OTUs) were picked at 97% identity level using UCLUST (Edgar 2010) with the nearest neighbor method in QIIME. Taxonomy was assigned using SILVA v119 as the reference database (Quast *et al.* 2013) using QIIME scripts. We removed samples with <5000 reads from further analysis. For the remaining samples (n= 99) reads varied between 5187 and 120,158 with a median of 21,883. Using this data, 100 OTU-by-sample matrices were randomly subsampled at an even depth of 5187 reads. Square root transformations were applied to matrices, subsequently pairwise sample-to-sample comparisons were generated and the median value across the 100 distance matrices was used to create a final median Bray–Curtis matrix. Shannon diversity (alpha-diversity metric) of all rarified OTU matrix libraries was also determined using QIIME (Caporaso *et al.* 2010). Unprocessed sequences are available through the NCBI's Sequence Read Archive (accession number). The relative abundance of families differed slightly in this 16S data from previous surveys of plant litter bacterial communities at the site (Matulich et al. 2015). After several tests comparing samples among the studies, we attribute this difference primarily to less degenerate 16S primers with Illumina sequencing compared to previous 454 sequencing.

Metagenomic Sequencing

Extracted DNA was quantified (Qubit dsDNA HS Assay, Invitrogen) and diluted to 0.2 ng/ul. Subsequently, metagenomic libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA) and sequenced with an Illumina HiSeq 4000 (150-bp paired ends) (DNA Technologies Core, UC Davis, CA). Sequences were uploaded and are publically available through the MG-RAST server (add #'s) (Wilke *et al.* 2016). Paired-end metagenomic library reads were joined using PEAR (Zhang *et al.* 2014). Joined sequences and the remaining unassembled forward sequences were combined and filtered using fastq-mcf in the EA-UTILS software package (Aronesty 2011). The sequences were then translated with FragGeneScan-Plus (Kim *et al.* 2015).

Taxonomic annotations of metagenomes were performed using a custom reference genomic database (https://github.com/alex-b-chase/LRGCE) constructed with phylogenetic marker genes from 3173 genomes which target soil microbes (Chase et al. unpublished). Metagenomic libraries were annotated for functional potential using hmmscan with HMMER software v3.1b2 (Eddy 1998), to search the Pfam family database (Finn *et al.* 2016), where top sequence matches with an e-value >e-04 were retained.

Metagenomic composition and functional data was processed using a similar pipeline as 16S community composition using QIIME. ID (taxonomic or pfam family)-by-sample matrices were converted to biom files and tables were filtered to remove singletons. We rarefied the ID-by-sample matrices at an even depth of 1001 reads and 107,231 reads for the taxonomic and functional annotations respectively. For rarefaction, 100 ID-by-sample matrices were randomly subsampled at an even depth.

Further processing in QIIME was performed using the same steps as with the 16S amplicon sequencing data (see above section).

Bacterial cell density

Bacterial cell densities were measured using flow cytometry, using a procedure modified from (Allison *et al.* 2013). At the time of litterbag collection, three 0.1 g subsamples of the litter were fixed with 5-mL 1% phosphate buffered glutaraldehyde solution within 6 hours of collection. Fixed samples were stored in the dark at 4˚ C for up to 1 week. To extract cells from the litter, 0.55 mL of 0.1 mol/L tetrasodium pyrophosphate was added to the sample and gently sonicated for 30 minutes in the dark at 4˚ C. The samples were then filtered through a 3.3-µm syringe filter to remove large particulates. 3 µL of SYBR Green (200x) was added to 600 µL of sample and incubated in the dark at room temperature for 10 minutes. Stained particle counts were performed using flow cytometry (BD Accuri C6; BD Biosciences, San Jose, California, USA). Each subsample was run three times for technical replication on the flow cytometer. The flow cytometer was set to run for 2 minutes on medium speed, using a threshold of 2000. Flow cytometry gating parameters were optimized to count particle sizes in the size range of typical bacterial cells (unpublished data). Cell densities are reported as number of stained counts per g dry weight litter.

Extracellular enzymes

We used a spectrophotometric assay to characterize the potential extracellular enzyme activity of the litter samples. We performed assays on 8 inoculum samples and

3 subsamples from each of the 32 final litterbags (n=104). Samples were stored for in the -80˚ C freezer for up to 3 months before processing. Sample homogenate preparation and flourimetric enzyme assays were performed using methods described in (Alster *et al.* 2013). We measured the potential activities of seven extracellular enzymes including, α-Glucosidase (AG), acid phosphatase (AP), ß-Glucosidase (BG), ß-Xylosidase (BX), cellobiohydrolase (CBH), leucine aminopeptidase (LAP), N-acetyl-ß-D-glucosaminidase (NAG). We then create a composite metric of the seven EEA assays, by calculating the Euclidean distance between each pairwise sample based on normalized measurements of the seven assays.

Statistical analyses

Our sampling design allowed us to estimate how variance in measures of community composition (16S data), abundance (bacterial cell counts), and EEA was distributed among the precipitation and dispersal treatments, stochastic effects (among bag replicates), and residuals (measurement error). For bacterial community composition and extracellular enzyme activity composition, we performed a permutational multivariate analysis of variance (PERMANOVA) (PRIMER6 & PERMANOVA+, Primer-E Ltd, Ivybridge, UK). This model included precipitation (ambient and added rainfall) and dispersal (open and closed) as main fixed factors, a precipitation-by-dispersal interaction, and bag replicates as a nested random factor. Analyses were run using type III partial sums of squares under a reduced model with 999 permutations. For bacterial cell density and individual EEA activity measurements, we used a factorial nested ANOVA design. As with community composition, the model

included precipitation and dispersal as fixed effects, a precipitation-by-dispersal interaction and bag replicates as a nested random factor. To test for mass loss differences, we performed a two-way ANOVA, because we could not measure mass loss on subsamples within a bag. The ANOVA analyses were conducted in the R software environment (Team 2011).

The above statistics provide a quantification of the among bag variation across all samples. However, they do not provide a measure of how this variation is distributed across the treatments. To measure relative variability within each treatment group (open-ambient, closed-ambient, open-precipitation, closed-precipitation), we quantified the distance to the centroid within each treatment combination (PERMDISP) (PRIMER6 & PERMANOVA, Primer-E Ltd, Ivybridge, UK). This test was performed for bacterial community composition, and the composite potential EEA. For both the EEA Euclidean distance matrix and bacterial community composition Bray Curtis dissimilarity matrix, in order to avoid pseudoreplication, we collapsed the three subsamples from each bag by finding the centroid of the subsamples using the 'Distances among centroids' feature in PERMANOVA+ (Anderson *et al.* 2008). Thus, for this analysis each bag represented one sample (total samples, n=32). Using a permutation test we ran pairwise comparisons of group mean dispersions based on the four treatment groups (PERMDISP). A two-way model was not used for this analysis, as this computation is not advised given the difficulties in testing homogeneity of dispersions across multiple main effects (Anderson 2006). Models were run under 999 permutations.

Supplementary Tables and Figures

Table S4.1. Nested Permutational MANOVA for a) bacterial community composition (16S amplicon sequencing) b) bacterial community composition from metagenomic data (marker genes) and c) functional composition from metagenomic data (Pfam protein families)

a. **Community composition (16S)**

	df	SS	MS	Pseudo-F	P(perm)
Precipitation	1	0.333	0.333	1.166	0.163
Dispersal	1	0.756	0.756	2.649	**0.001**
Precipitation X Dispersal	1	0.415	0.415	1.454	**0.029**
Bag (Precipitation x Dispersal)	29	8.418	0.293	1.605	**0.001**
Residual	59	10.67	0.181		

b. **Community composition (metagenomes)**

	df	SS	MS	Pseudo-F	P(perm)
Precipitation	1	0.11	0.11	1.38	0.11
Dispersal	1	0.27	0.27	3.48	**0.001**
Precipitation X Dispersal	1	9.17E-2	9.17E-2	1.17	0.226
Bag (Precipitation x Dispersal)	28	2.20	7.87E-2	1.69	**0.001**
Residual	61	2.84	4.66E-2		

c. **Functional composition (metagenomes)**

	df	SS	MS	Pseudo-F	P(perm)
Precipitation	1	2.89E-2	2.89E-2	1.158	0.226
Dispersal	1	5.43E-2	5.43E-2	2.177	**0.006**
Precipitation X Dispersal	1	3.93E-2	3.93E-2	1.577	**0.04**
Bag (Precipitation x Dispersal)	27	0.69	2.56E-2	1.6813	**0.01**
Residual	55	0.8374	1.53E-2		

Table S4.2. Significance of the effects of precipitation, dispersal, precipitation x dispersal and bag on extracellular enzyme activity (EEA), which was assessed using a two-way nested ANOVA. Each treatment (open-precipitation, open-ambient, closed-precipitation, closed-ambient) had 8 replicate bags and individual EEA assays were run on three subsamples of each bag.

	AG	**AP**	**BG**	**BX**	**CBH**	**LAP**	**NAG**
Precipitation	NS	NS	NS	NS	NS	NS	NS
Dispersal	NS	NS	NS	NS	NS	****	NS
Precipitation X Dispersal	**	*	*	*	•	NS	NS
Bag (Precipitation x Dispersal)	****	****	****	****	****	****	****

.

• *< 0.1, * <0.05, ** <0.01, *** <0.001, ****<0.0001*

Table S4.3. Nested Permutational MANOVA for extracellular enzyme activity (EEA) composition. To create a composite metric of the seven EEA assays, we calculated the Euclidean distance between each pairwise sample based on normalized measurements of the seven assays.

	df	SS	MS	F	P
Precipitation	1	2.7341	2.7341	0.25	0.89
Dispersal	1	23.316	23.316	2.13	0.111
Precipitation X Dispersal	1	58.728	58.728	5.37	**0.007**
Bag (Precipitation x Dispersal)	28	308.91	11.033	2.69	**0.001**
Residual	61	250.42	4.1053		

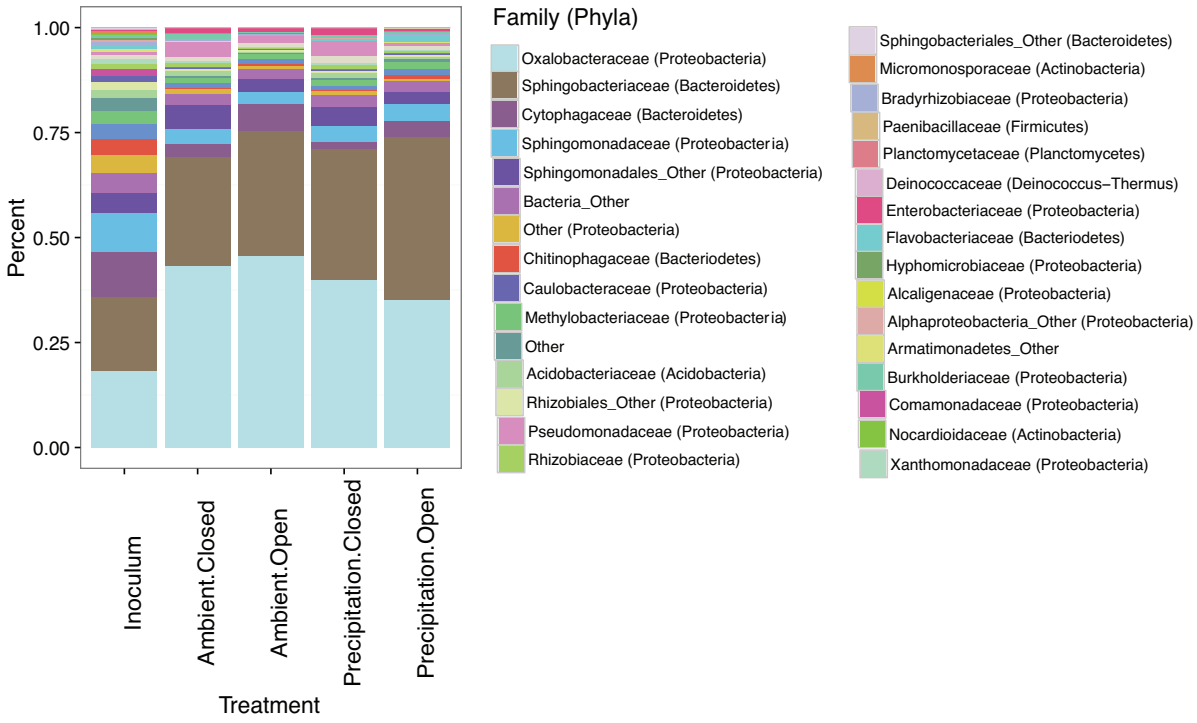


Figure S4.1. Relative abundance of bacterial families (16S amplicon sequencing) found in each treatment type and the inoculum averaged across replicates.

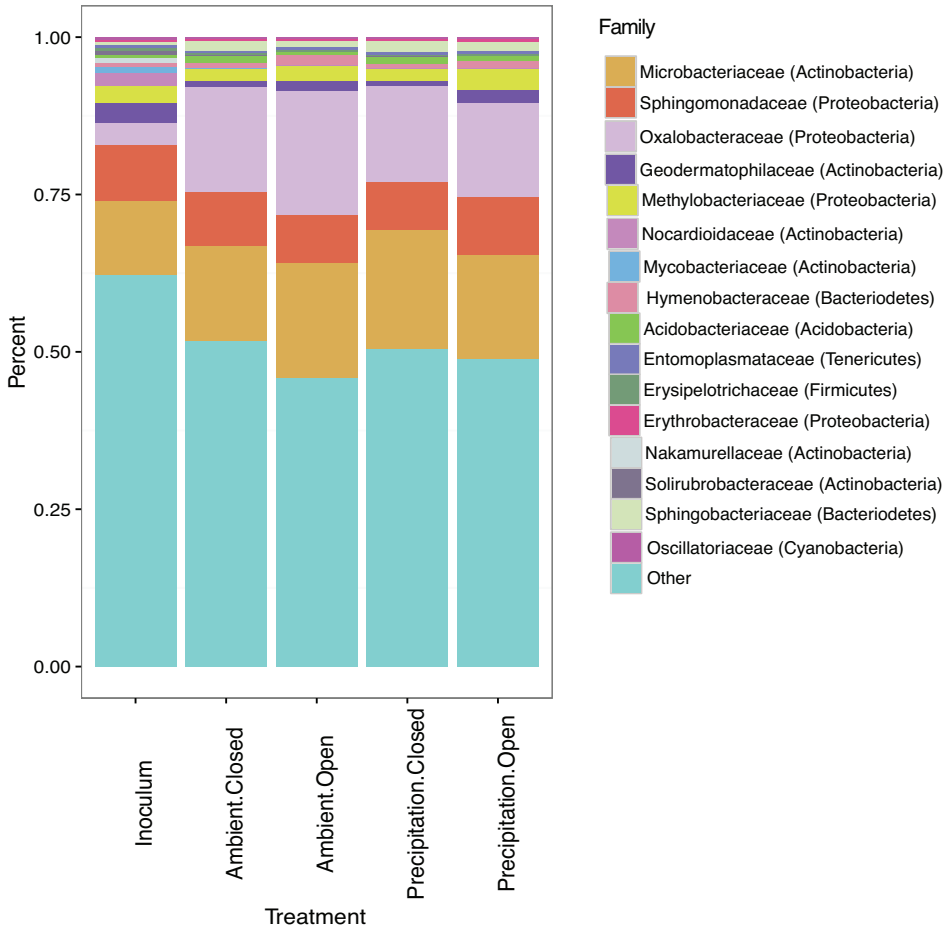


Figure S4.2. Relative abundance of bacterial families (metagenomic sequencing) found in each treatment type and the inoculum averaged across replicates.

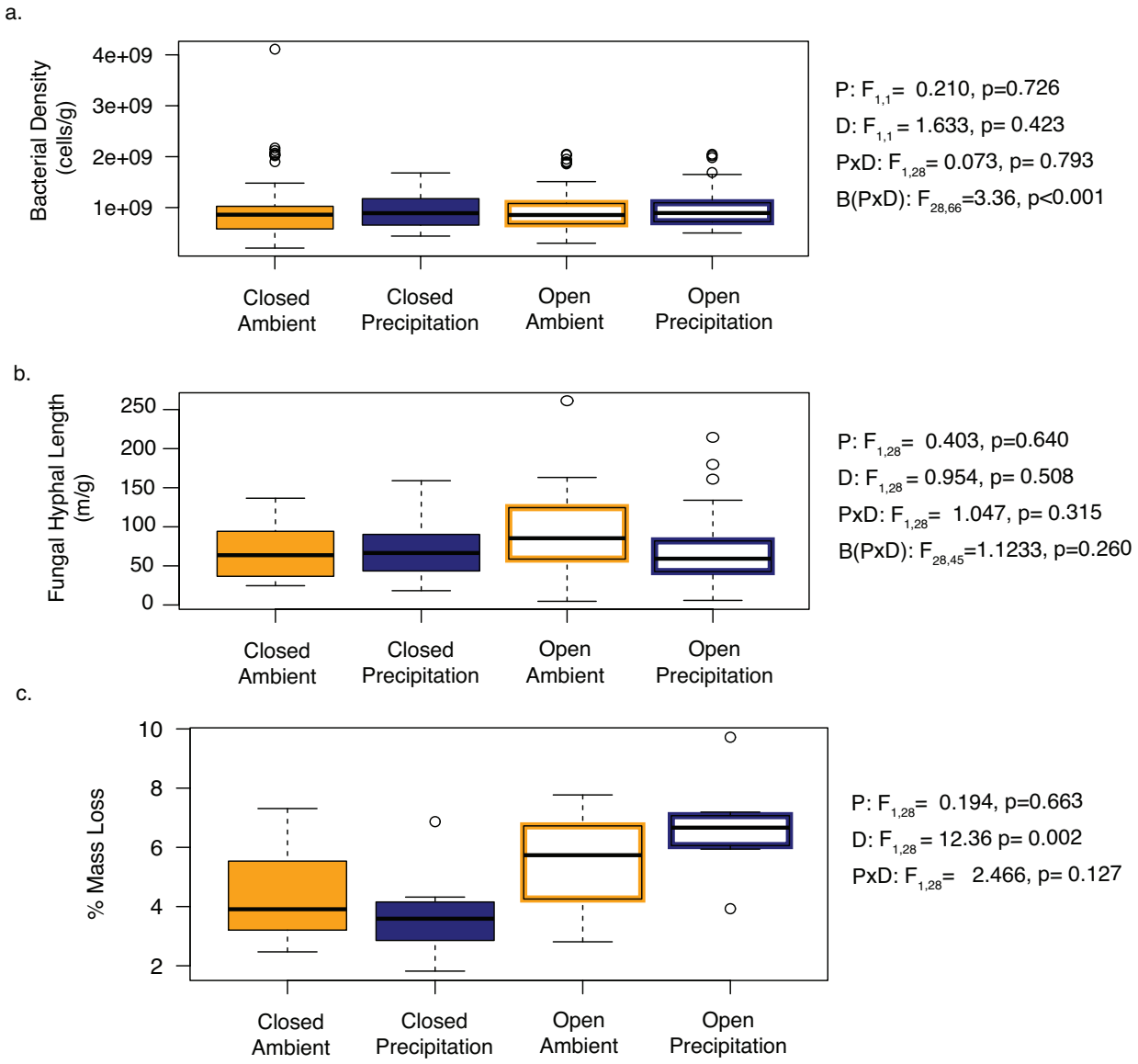
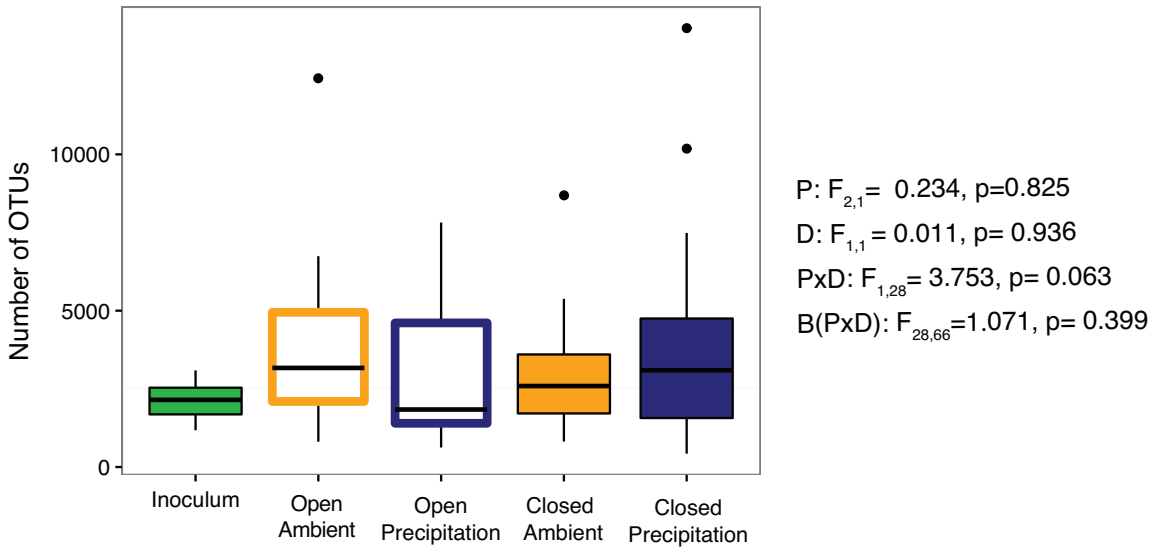


Figure S4.3. a) Bacterial cells counts per gram of dry litter, b) fungal hyphal lengths, and c) mass loss across the four treatments.

a.



b.

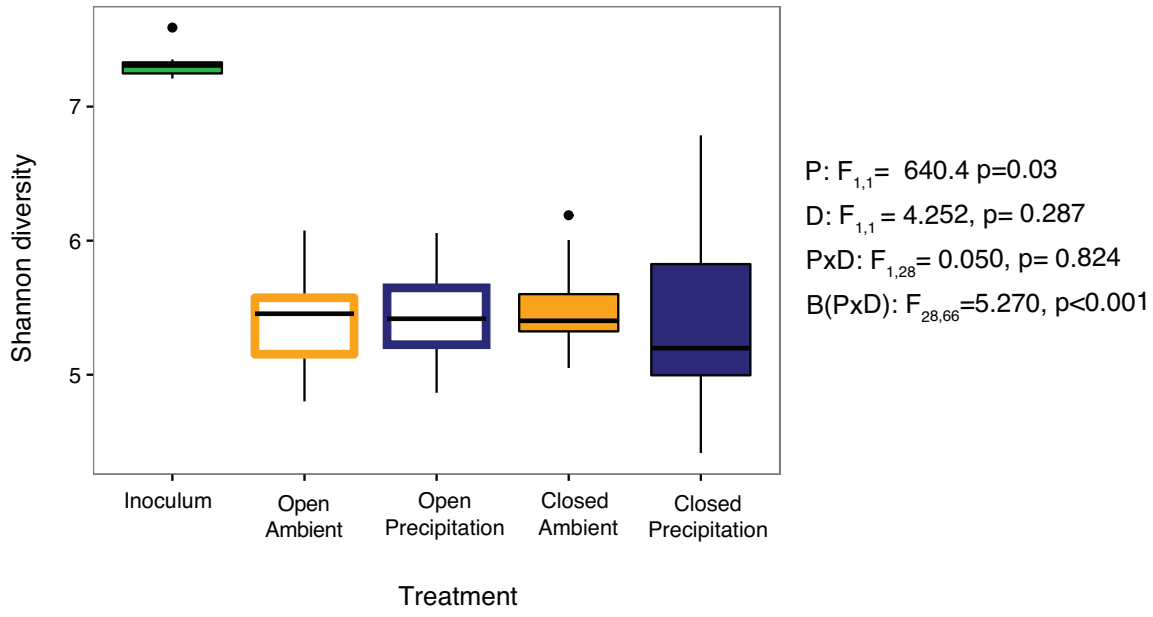


Figure S4.4. Effects of the dispersal and precipitation treatment on bacterial a) richness (number of OTUs) and b) evenness (Shannon diversity) as assayed by 16S sequencing, after rarefaction to standardize for sequencing effort among samples.

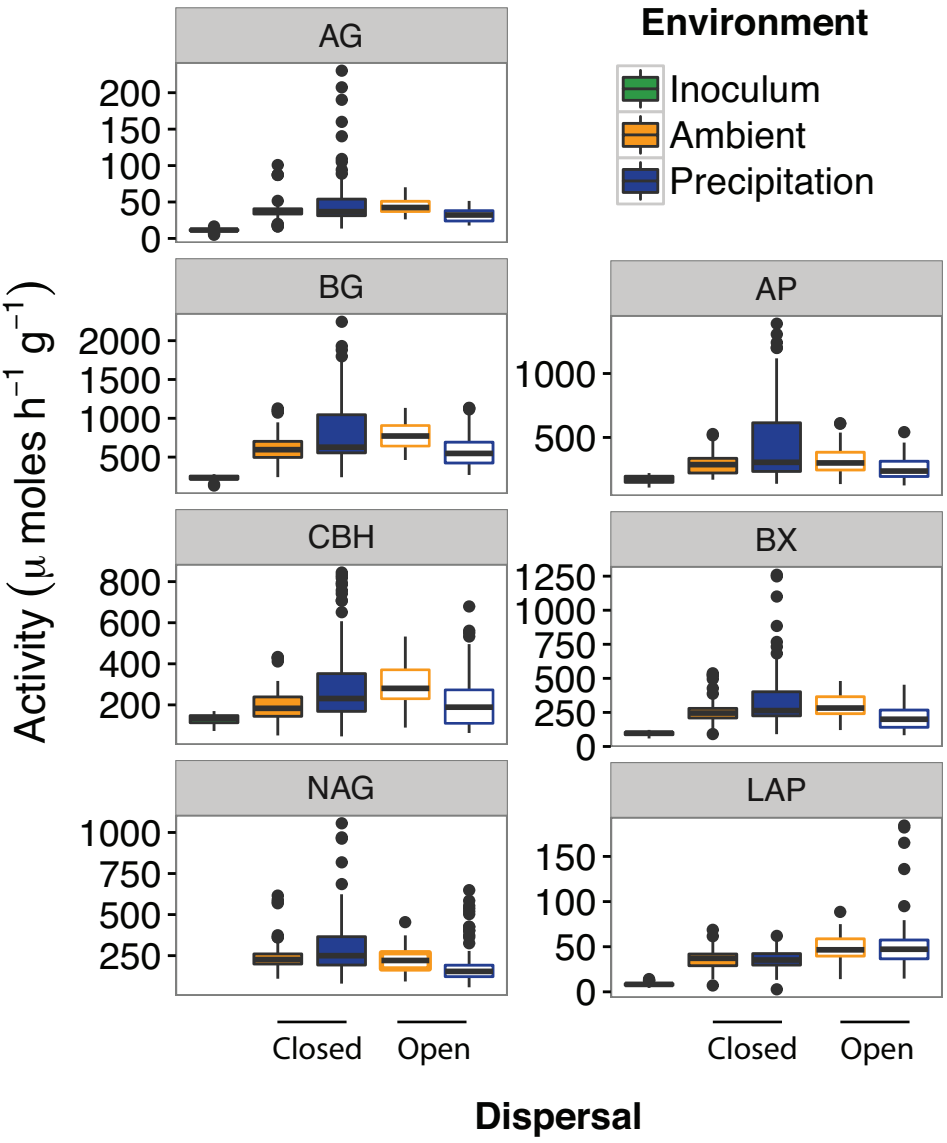


Figure S4.5. Potential activities of seven extracellular enzymes including, A-Glucosidase (AG), acid phosphatase (AP), B-Glucosidase (BG), B-Xylosidase (BX), cellobiohydrolase (CBH), leucine aminopeptidase (LAP), N-acetyl-B-D-glucosaminidase (NAG). Measurements are shown for inoculum (n=8) and 3 replicates from each final litterbag (n=96).

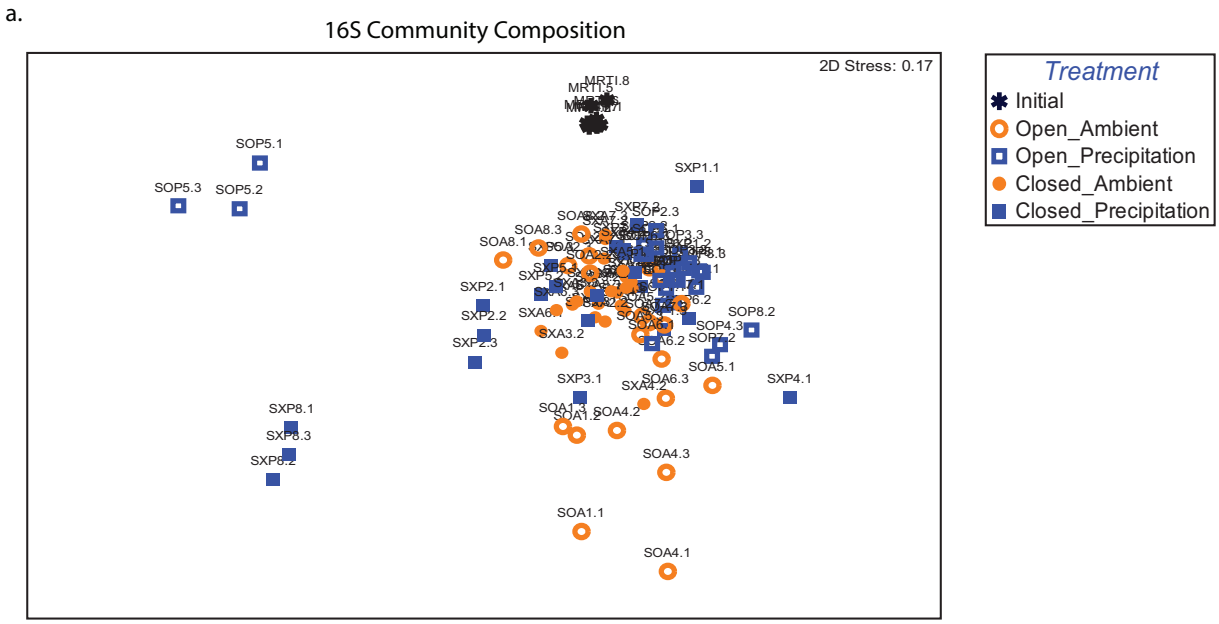


Figure S4.6. Nonmetric multidimensional scaling (NMDS) ordinations showing variability in bacterial community composition (Bray-Curtis dissimilarity) with all replicate subsamples, as well as initial inoculum subsamples plotted.

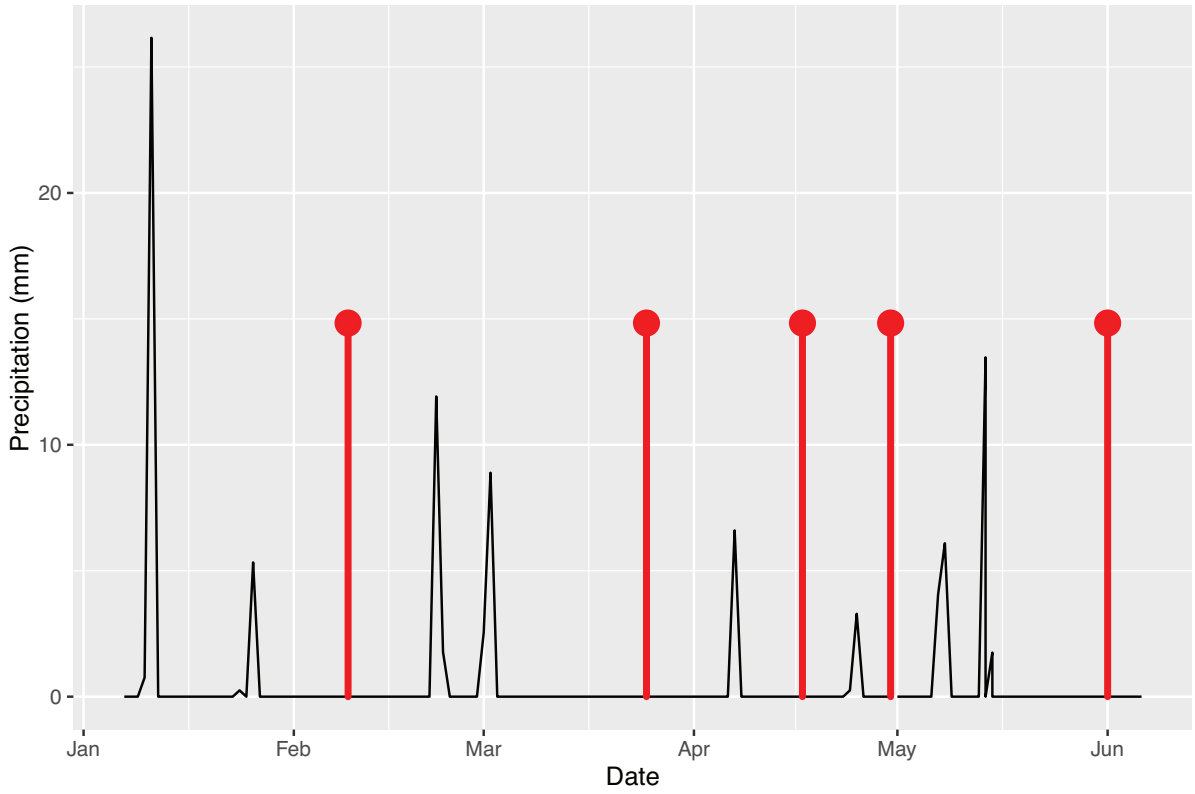


Figure S4.7. Precipitation at Loma Ridge between January 2015 and June 2015. Red lines indicate water additions.

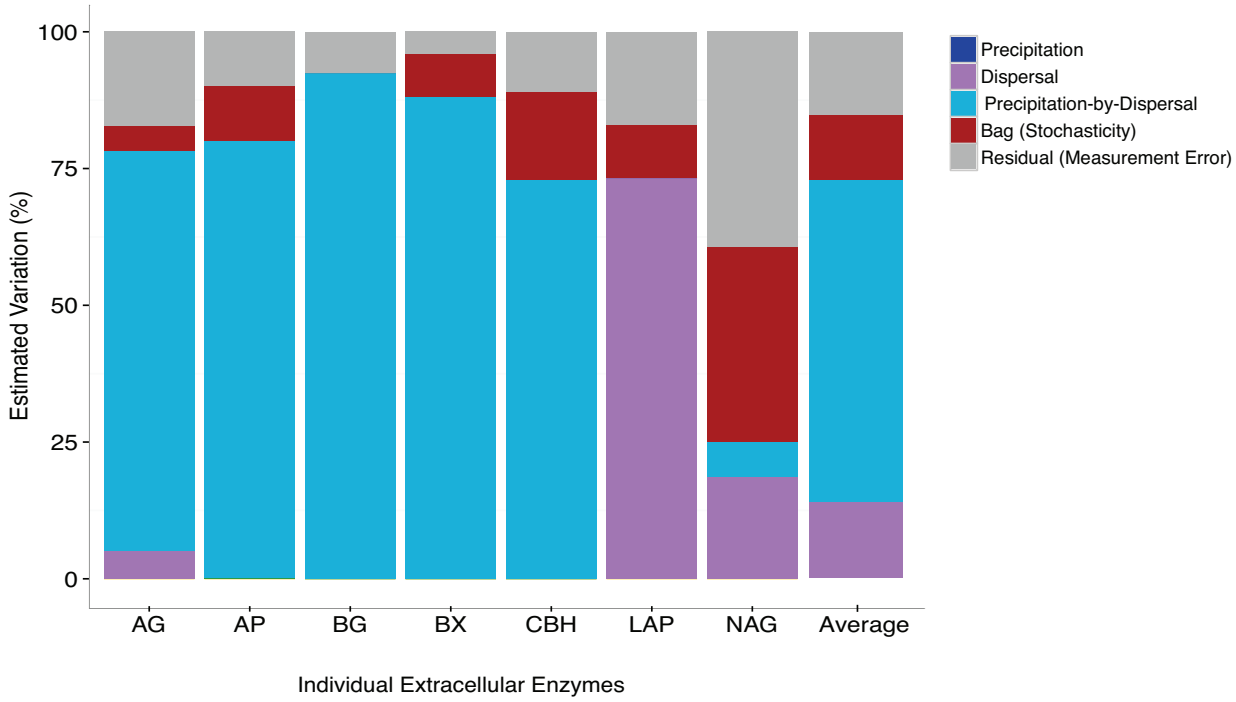


Figure S4.8. Percentage of estimated variation explained by treatment effects and measurement error for 7 individual extracellular enzymes including A-Glucosidase (AG), acid phosphatase (AP), B-Glucosidase (BG), B-Xylosidase (BX), cellobiohydrolase (CBH), leucine aminopeptidase (LAP), N-acetyl-B-D-glucosaminidase (NAG) and the average across the 7 enzymes

References

1.
Adair, K.L. & Douglas, A.E. (2017). Making a microbiome: the many determinants of host-associated microbial community composition. *Curr Opin Microbiol*, 35, 23-29.

2.
Aerts, R. (1997). Climate, leaf litter chemistry and leaf litter decomposition in terrestrial ecosystems: A triangular relationship. *Oikos*, 79, 439-449.

3.
Allen, E.B., Padgett, P.E., Bytnerowicz, A. & Minnich, R. (1998). Nitrogen deposition effects on coastal sage vegetation of southern California. In: *General Technical Report PSW-GTR-166*. USDA Forest Service Riverside, California.

4.
Allison, S.D., Lu, Y., Weihe, C., Goulden, M.L., Martiny, A.C., Treseder, K.K. *et al.* (2013). Microbial abundance and composition influence litter decomposition response to environmental change. *Ecology*, 94, 714-725.

5.
Allison, S.D. & Martiny, J.B.H. (2008). Resistance, resilience, and redundancy in microbial communities. *P Natl Acad Sci USA*, 105, 11512-11519.

6.
Allison, S.D., Weintraub, M.N., Gartner, T.B. & Waldrop, M.P. (2011). Evolutionary-Economic Principles as Regulators of Soil Enzyme Production and Ecosystem Function. *Soil Biol*, 22, 229-243.

7.
Alster, C.J., German, D.P., Lu, Y. & Allison, S.D. (2013). Microbial enzymatic responses to drought and to nitrogen addition in a southern California grassland. *Soil Biol Biochem*, 64, 68-79.

8.
Andam, C.P., Doroghazi, J.R., Campbell, A.N., Kelly, P.J., Choudoir, M.J. & Buckley, D.H. (2016). A Latitudinal Diversity Gradient in Terrestrial Bacteria of the Genus Streptomyces. *Mbio*, 7.

9.
Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecol*, 26, 32-46.

10.
Anderson, M.J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62, 245-253.

11.
Anderson, M.J., Gorley, R.N. & Clarke, K.R. (2008). PERMANOVA+ for PRIMER: guide to software and statistical methods. Plymouth, UK.

12.
Apprill, A., McNally, S., Parsons, R. & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol*, 75, 129-137.

13.

Aronesty, E. (2011). ea-utils: command-line tools for processing biological sequencing data. . http://code.google.com/p/ea-utils.
14.
Attard, E., Recous, S., Chabbi, A., De Berranger, C., Guillaumaud, N., Labreuche, J. *et al.* (2011). Soil environmental conditions rather than denitrifier abundance and diversity drive potential denitrification after changes in land uses. *Global Change Biol*, 17, 1975-1989.
15.
Barberan, A., Fernandez-Guerra, A., Bohannan, B.J.M. & Casamayor, E.O. (2012). Exploration of community traits as ecological markers in microbial metagenomes. *Molecular Ecology*, 21, 1909-1917.
16.
Barberan, A., Ladau, J., Leff, J.W., Pollard, K.S., Menninger, H.L., Dunn, R.R. *et al.* (2015). Continental-scale distributions of dust-associated bacteria and fungi. *P Natl Acad Sci USA*, 112, 5756-5761.
17.
Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R. & Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *Isme J*, 5, 908-917.
18.
Batjes, N.H. (2000). Global Data Set of Derived Soil Properties, 0.5-Degree Grid (ISRIC-WISE). Available on-line [http://www.daac.ornl.gov/] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.
19.
Batmalle, C.S., Chiang, H.I., Zhang, K., Lomas, M.W. & Martiny, A.C. (2014). Development and Bias Assessment of a Method for Targeted Metagenomic Sequencing of Marine Cyanobacteria. *Appl Environ Microb*, 80, 1116-1125.
20.
Bell, T. (2010). Experimental tests of the bacterial distance-decay relationship. *Isme J*, 4, 1357-1365.
21.
Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A. *et al.* (2011). The under-recognized dominance of Verrucomicrobia in soil bacterial communities. *Soil Biol Biochem*, 43, 1450-1455.
22.
Berlemont, R., Allison, S.D., Weihe, C., Lu, Y., Brodie, E.L., Martiny, J.B.H. *et al.* (2014). Cellulolytic potential under environmental changes in microbial communities from grassland litter. *Front Microbiol*, 5.
23.
Bivand, R.S., Pebesma, E. & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R*. Second edn, Springer, NY.
24.
Boberg, J.B., Finlay, R.D., Stenlid, J. & Lindahl, B.D. (2010). Fungal C translocation restricts N-mineralization in heterogeneous environments. *Funct Ecol*, 24, 454-459.

25.
Boonpragob, K. & Nash, T.H. (1990). Seasonal-Variation of Elemental Status in the Lichen Ramalina-Menziesii Tayl from 2 Sites in Southern California - Evidence for Dry Deposition Accumulation. *Environ Exp Bot*, 30, 415-428.
26.
Booth, M.S., Stark, J.M. & Rastetter, E. (2005). Controls on nitrogen cycling in terrestrial ecosystems: A synthetic analysis of literature data. *Ecol Monogr*, 75, 139-157.
27.
Bouwman, A.F., Beusen, A.H.W., Griffioen, J., Van Groenigen, J.W., Hefting, M.M., Oenema, O. *et al.* (2013). Global trends and uncertainties in terrestrial denitrification and N2O emissions. *Philos T R Soc B*, 368.
28.
Bradford, M.A., Tordoff, G.M., Eggers, T., Jones, T.H. & Newington, J.E. (2002). Microbiota, fauna, and mesh size interactions in litter decomposition. *Oikos*, 99, 317-323.
29.
Braithwaite, D.T. & Keegan, K.P. (2013). matR: Metagenomics Analysis Tools for R. R package version 0.9.9.
30.
Brooks, J.P., Edwards, D.J., Harwich, M.D., Rivera, M.C., Fettweis, J.M., Serrano, M.G. *et al.* (2015). The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *Bmc Microbiol*, 15.
31.
Brown, J.R., Blankinship, J.C., Niboyet, A., van Groenigen, K.J., Dijkstra, P., Le Roux, X. *et al.* (2012). Effects of multiple global change treatments on soil N2O fluxes. *Biogeochemistry*, 109, 85-100.
32.
Brown, S.P. & Jumpponen, A. (2014). Contrasting primary successional trajectories of fungi and bacteria in retreating glacier soils. *Mol Ecol*, 23, 481-497.
33.
Bru, D., Ramette, A., Saby, N.P.A., Dequiedt, S., Ranjard, L., Jolivet, C. *et al.* (2011). Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *Isme J*, 5, 532-542.
34.
Brumelle, S., Nemetz, P. & Casey, D. (1984). Estimating means and variances: The comparative efficiency of composite and grab samples. *Environmental monitoring and assessment*, 4, 81-84.
35.
Burger, M. & Jackson, L.E. (2003). Microbial immobilization of ammonium and nitrate in relation to ammonification and nitrification rates in organic and conventional cropping systems. *Soil Biol Biochem*, 35, 29-36.
36.
Burke, C., Steinberg, P., Rusch, D., Kjelleberg, S. & Thomas, T. (2011). Bacterial community assembly based on functional genes rather than species. *P Natl Acad Sci USA*, 108, 14288-14293.
37.

Butterbach-Bahl, K. & Gundersen, P. (2011). Nitrogne Processes in terrestrial ecosystems. In: *The European Nitrogen Assessment*. Cambridge University Press Cambridge, UK, pp. 99-125.
38.
Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7, 335-336.
39.
Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N. *et al.* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme J*, 6, 1621-1624.
40.
Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P. *et al.* (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59-67.
41.
Chapin, F.S., Matson, P.A. & Mooney, H.A. (2002). *Principles of Terrestrial Ecosystem Ecology*. Springer, New York.
42.
Chase, J.M. (2007). Drought mediates the importance of stochastic community assembly. *P Natl Acad Sci USA*, 104, 17430-17434.
43.
Chase, J.M. & Myers, J.A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos T R Soc B*, 366, 2351-2363.
44.
Cisneros-Dozal, L.M., Trumbore, S.E. & Hanson, P.J. (2007). Effect of moisture on leaf litter decomposition and its contribution to soil respiration in a temperate forest. *J Geophys Res-Biogeo*, 112.
45.
Clarke, K.R. & Gorley, R.N. (2006). PRIMER v6: User Manual/Tutorial. PRIMER-E Plymouth.
46.
Clarke, K.R. & Warwick, R.M. (2001). *Change in marine communities: an approach to statistical analysis and inter- pretation.* 2nd edn, Plymouth, UK.
47.
Cleveland, C.C. & Liptzin, D. (2007). C : N : P stoichiometry in soil: is there a "Redfield ratio" for the microbial biomass? *Biogeochemistry*, 85, 235-252.
48.
Cleveland, C.C., Townsend, A.R., Schimel, D.S., Fisher, H., Howarth, R.W., Hedin, L.O. *et al.* (1999). Global patterns of terrestrial biological nitrogen (N-2) fixation in natural ecosystems. *Global Biogeochem Cy*, 13, 623-645.
49.
Cobo-Díaz, J.F., Fernández-González, A.J., Villadas, P.J., Robles, A.B., Toro, N. & Fernández-López, M. (2015). Metagenomic Assessment of the Potential Microbial Nitrogen Pathways in the Rhizosphere of a Mediterranean Forest After a Wildfire. *Microbial Ecology*.
50.

Cole, J.A. (1990). Physiology, Biochemistry and Genetics of Nitrate Dissimilation to Ammonia. *Fems Symp*, 57-76.
51.
Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecol Lett*, 8, 1175-1182.
52.
Cregger, M.A., Schadt, C.W., McDowell, N.G., Pockman, W.T. & Classen, A.T. (2012). Response of the Soil Microbial Community to Changes in Precipitation in a Semiarid Ecosystem. *Appl Environ Microb*, 78, 8587-8594.
53.
Cruz-Martinez, K., Suttle, K.B., Brodie, E.L., Power, M.E., Andersen, G.L. & Banfield, J.F. (2009). Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *Isme J*, 3, 738-744.
54.
Datta, M.S., Sliwerska, E., Gore, J., Polz, M.F. & Cordero, O.X. (2016). Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nat Commun*, 7.
55.
Dehal, P.S., Joachimiak, M.P., Price, M.N., Bates, J.T., Baumohl, J.K., Chivian, D. *et al.* (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res*, 38, D396-D400.
56.
Dentener, F., Drevet, J., Lamarque, J.F., Bey, I., Eickhout, B., Fiore, A.M. *et al.* (2006). Nitrogen and sulfur deposition on regional and global scales: A multimodel evaluation. *Global Biogeochem Cy*, 20, GB4003.
57.
Diaz, S. & Cabido, M. (2001). Vive la difference: plant functional diversity matters to ecosystem processes. *Trends Ecol Evol*, 16, 646-655.
58.
Dickie, I.A., Fukami, T., Wilkie, J.P., Allen, R.B. & Buchanan, P.K. (2012). Do assembly history effects attenuate from species to ecosystem properties? A field test with wood-inhabiting fungi. *Ecol Lett*, 15, 133-141.
59.
Dini-Andreote, F., Stegen, J.C., van Elsas, J.D. & Salles, J.F. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *P Natl Acad Sci USA*, 112, E1326-E1332.
60.
Dumbrell, A.J., Nelson, M., Helgason, T., Dytham, C. & Fitter, A.H. (2010). Relative roles of niche and neutral processes in structuring a soil microbial community (vol 4, pg 337, 2010). *Isme J*, 4, 1078-1078.
61.
Eddy, S.R. (1998). Profile hidden Markov models. *Bioinformatics*, 14, 755-763.
62.
Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460-2461.

63.
Egerton-Warburton, L.M. & Allen, E.B. (2000). Shifts in arbuscular mycorrhizal communities along an anthropogenic nitrogen deposition gradient. *Ecol Appl*, 10, 484-496.
64.
Evans, S., Martiny, J.B.H. & Allison, S.D. (2017). Effects of dispersal and selection on stochastic assembly in microbial communities. *Isme J*, 11, 176-185.
65.
Falkowski, P.G., Fenchel, T. & Delong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320, 1034-1039.
66.
Farnelid, H., Andersson, A.F., Bertilsson, S., Al-Soud, W.A., Hansen, L.H., Sorensen, S. *et al.* (2011). Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *Plos One*, 6, e19223.
67.
Fazzolari, E., Nicolardot, B. & Germon, J.C. (1998). Simultaneous effects of increasing levels of glucose and oxygen partial pressures on denitrification and dissimilatory nitrate reduction to ammonium in repacked soil cores. *Eur J Soil Biol*, 34, 47-52.
68.
Fenn, M.E., Allen, E.B., Weiss, S.B., Jovan, S., Geiser, L.H., Tonnesen, G.S. *et al.* (2010). Nitrogen critical loads and management alternatives for N-impacted ecosystems in California. *J Environ Manage*, 91, 2404-2423.
69.
Fenn, M.E., Baron, J.S., Allen, E.B., Rueth, H.M., Nydick, K.R., Geiser, L. *et al.* (2003). Ecological effects of nitrogen deposition in the western United States. *Bioscience*, 53, 404-420.
70.
Fenn, M.E. & Bytnerowicz, A. (1997). Summer throughfall and winter deposition in the San Bernardino Mountains in southern California. *Atmos Environ*, 31, 673-683.
71.
Ferrenberg, S., O'Neill, S.P., Knelman, J.E., Todd, B., Duggan, S., Bradley, D. *et al.* (2013). Changes in assembly processes in soil bacterial communities following a wildfire disturbance. *Isme J*, 7, 1102-1111.
72.
Fierer, N., Barberan, A. & Laughlin, D.C. (2014). Seeing the forest for the genes: using rnetagenomics to infer the aggregated traits of microbial communities. *Front Microbiol*, 5, 614.
73.
Fierer, N. & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *P Natl Acad Sci USA*, 103, 626-631.
74.
Fierer, N. & Ladau, J. (2012). Predicting microbial distributions in space and time. *Nat Methods*, 9, 549-551.
75.

Fierer, N., Ladau, J., Clemente, J.C., Leff, J.W., Owens, S.M., Pollard, K.S. *et al.* (2013). Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. *Science*, 342, 621-624.
76.
Fierer, N., Lauber, C.L., Ramirez, K.S., Zaneveld, J., Bradford, M.A. & Knight, R. (2012a). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *Isme J*, 6, 1007-1017.
77.
Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L. *et al.* (2012b). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *P Natl Acad Sci USA*, 109, 21390-21395.
78.
Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A. & Cleveland, C.C. (2009). Global patterns in belowground communities. *Ecol Lett*, 12, 1238-1249.
79.
Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L. *et al.* (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*, 44, D279-D285.
80.
Fowler, D., Coyle, M., Skiba, U., Sutton, M.A., Cape, J.N., Reis, S. *et al.* (2013). The global nitrogen cycle in the twenty-first century. *Philos T R Soc B*, 368, 20130164.
81.
Frank, J.A. & Sorensen, S.J. (2011). Quantitative metagenomic analyses based on average genome size normalization. *Appl Environ Microb*, 77, 2513-2521.
82.
Frey, S.D., Knorr, M., Parrent, J.L. & Simpson, R.T. (2004). Chronic nitrogen enrichment affects the structure and function of the soil microbial community in temperate hardwood and pine forests. *Forest Ecol Manag*, 196, 159-171.
83.
Gans, J., Wolinsky, M. & Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, 309, 1387-1390.
84.
Geisseler, D., Horwath, W.R., Joergensen, R.G. & Ludwig, B. (2010). Pathways of nitrogen utilization by soil microorganisms – A review. *Soil Biology & Biochemistry*, 42, 9.
85.
Gholz, H.L., Wedin, D.A., Smitherman, S.M., Harmon, M.E. & Parton, W.J. (2000). Long-term dynamics of pine and hardwood litter in contrasting environments: toward a global model of decomposition. *Global Change Biol*, 6, 751-765.
86.
Gotelli, N.J. & Colwell, R.K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett*, 4, 379-391.
87.
Graham, E.B., Wieder, W.R., Leff, J.W., Weintraub, S.R., Townsend, A.R., Cleveland, C.C. *et al.* (2014). Do we need to understand microbial communities to predict

ecosystem function? A comparison of statistical models of nitrogen cycling processes. *Soil Biol Biochem*, 68, 279-282.
88.
Green, J.L., Bohannan, B.J.M. & Whitaker, R.J. (2008). Microbial biogeography: From taxonomy to traits. *Science*, 320, 1039-1043.
89.
Group, G.S.D.T. (2000). Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.
90.
Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B.C., James, P. *et al.* (2011). Niche specialization of terrestrial archaeal ammonia oxidizers. *P Natl Acad Sci USA*, 108, 21206-21211.
91.
Gubry-Rangin, C., Nicol, G.W. & Prosser, J.I. (2010). Archaea rather than bacteria control nitrification in two agricultural acidic soils. *Fems Microbiol Ecol*, 74, 566-574.
92.
Gutknecht, J.L.M., Field, C.B. & Balser, T.C. (2012). Microbial communities and their responses to simulated global change fluctuate greatly over multiple years. *Global Change Biol*, 18, 2256-2269.
93.
Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B.H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol*, 10, 497-506.
94.
Hao, Y.Q., Zhao, X.F. & Zhang, D.Y. (2016). Field experimental evidence that stochastic processes predominate in the initial assembly of bacterial communities. *Environ Microbiol*, 18, 1730-1739.
95.
Harell, J.F.E. (2001). *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag, New York.
96.
Hartley, A.E. & Schlesinger, W.H. (2000). Environmental controls on nitric oxide emission from northern Chihuahuan desert soils. *Biogeochemistry*, 50, 279-300.
97.
Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*, 25, 1965-1978.
98.
Hijmans, R.J. & van Etten, J. (2012). raster: Geographic analysis and modeling with raster data.
99.
Horner-Devine, M.C., Lage, M., Hughes, J.B. & Bohannan, B.J.M. (2004). A taxa-area relationship for bacteria. *Nature*, 432, 750-753.
100.

Hubbell, S.P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton Univeristy Press, Princeton.
101.
Hughes, J.B., Hellmann, J.J., Ricketts, T.H. & Bohannan, B.J.M. (2001). Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl Environ Microb*, 67, 4399-4406.
102.
Huisman, J. & Weissing, F.J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, 402, 407-410.
103.
Huisman, J. & Weissing, F.J. (2001). Fundamental unpredictability in multispecies competition. *Am Nat*, 157, 488-494.
104.
Jonasson, S., Michelsen, A. & Schmidt, I.K. (1999). Coupling of nutrient cycling and carbon dynamics in the Arctic, integration of soil microbial and plant processes. *Appl Soil Ecol*, 11, 135-146.
105.
Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*, 38, D355-360.
106.
Kim, D., Hahn, A.S., Wu, S.J., Hanson, N.W., Konwar, K.M. & Hallam, S.J. (2015). FragGeneScan-Plus for scalable high-throughput short-read open reading frame prediction. *2015 Ieee Conference on Computational Intelligence in Bioinformatics and Computational Biology (Cibcb)*, 88-95.
107.
Kimball, S., Goulden, M.L., Suding, K.N. & Parker, S. (2014). Altered water and nitrogen input shifts succession in a southern California coastal sage community. *Ecol Appl*, 24, 1390-1404.
108.
Knorr, M., Frey, S.D. & Curtis, P.S. (2005). Nitrogen additions and litter decomposition: A meta-analysis. *Ecology*, 86, 3252-3257.
109.
Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B. & Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, 437, 543-546.
110.
Kraft, B., Strous, M. & Tegetmeyer, H.E. (2011). Microbial nitrate respiration - Genes, enzymes and environmental distribution. *J Biotechnol*, 155, 104-117.
111.
Krause, S., Le Roux, X., Niklaus, P.A., Van Bodegom, P.M., Lennon, J.T., Bertilsson, S. *et al.* (2014). Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Front Microbiol*, 5.
112.

Langenheder, S., Berga, M., Ostman, O. & Szekely, A.J. (2012). Temporal variation of beta-diversity and assembly mechanisms in a bacterial metacommunity. *Isme J*, 6, 1107-1114.
113.
Lauber, C.L., Hamady, M., Knight, R. & Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl Environ Microb*, 75, 5111-5120.
114.
Lavorel, S. & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct Ecol*, 16, 545-556.
115.
Lee, J.E., Buckley, H.L., Etienne, R.S. & Lear, G. (2013). Both species sorting and neutral processes drive assembly of bacterial communities in aquatic microcosms. *Fems Microbiol Ecol*, 86, 288-302.
116.
Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. *et al.* (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett*, 7, 601-613.
117.
Letunic, I. & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23, 127-128.
118.
Letunic, I. & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*, 39, W475-W478.
119.
Lindstrom, E.S. & Ostman, O. (2011). The Importance of Dispersal for Bacterial Community Composition and Functioning. *Plos One*, 6.
120.
Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., D.S., S., Parfrey, L.W. *et al.* (2016a). High taxonomic variability despite stable functional struc- ture across microbial communities. *Nature Ecology & Evolution*, 1:0015.
121.
Louca, S., Parfrey, L.W. & Doebeli, M. (2016b). Decoupling function and taxonomy in the global ocean microbiome. *Science*, 353, 1272-1277.
122.
Lozupone, C.A. & Knight, R. (2007). Global patterns in bacterial diversity. *P Natl Acad Sci USA*, 104, 11436-11440.
123.
Luo, C., Rodriguez, R.L., Johnston, E.R., Wu, L., Cheng, L., Xue, K. *et al.* (2014). Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. *Appl Environ Microbiol*, 80, 1777-1786.
124.
Ma, B., Hibbing, M.E., Kim, H.S., Reedy, R.M., Yedidia, I., Breuer, J. *et al.* (2007). Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. *Phytopathology*, 97, 1150-1163.

125.
Mac Nally, R. (2002). Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodivers Conserv*, 11, 1397-1401.
126.
Mania, D., Heylen, K., van Spanning, R.J.M. & Frostegard, A. (2014). The nitrate-ammonifying and nosZ-carrying bacterium Bacillus vireti is a potent source and sink for nitric and nitrous oxide under high nitrate conditions. *Environ Microbiol*, 16, 3196-3210.
127.
Manor, O. & Borenstein, E. (2015). MUSiCC: a marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. *Genome Biol*, 16, 10.1186.
128.
Martiny, A.C., Treseder, K. & Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *Isme J*, 7, 830-838.
129.
Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L. *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol*, 4, 102-112.
130.
Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D. & Horner-Devine, M.C. (2011). Drivers of bacterial beta-diversity depend on spatial scale. *P Natl Acad Sci USA*, 108, 7850-7854.
131.
Martiny, J.B.H., Jones, S.E., Lennon, J.T. & Martiny, A.C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science*, 350, aac9323.
132.
Martiny, J.B.H., Martiny, A.C., Weihe, C., Lu, Y., Berlemont, R., Brodie, E.L. *et al.* (2017). Microbial legacies alter decomposition in response to simulated global change. *Isme J*, 11, 490-499.
133.
Mason, O.U., Scott, N.M., Gonzalez, A., Robbins-Pianka, A., Baelum, J., Kimbrel, J. *et al.* (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *Isme J*, 8, 1464-1475.
134.
Matulich, K.L. & Martiny, J.B.H. (2015). Microbial composition alters the response of litter decomposition to environmental change. *Ecology*, 96, 154-163.
135.
Matulich, K.L., Weihe, C., Allison, S.D., Amend, A.S., Berlemont, R., Goulden, M.L. *et al.* (2015). Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. *Isme J*, 9, 2477-2489.
136.
McGill, B.J., Enquist, B.J., Weiher, E. & Westoby, M. (2006). Rebuilding community ecology from functional traits. *Trends Ecol Evol*, 21, 178-185.
137.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M. *et al.* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics*, 9, 386.
138.
Miki, T., Ushio, M., Fukui, S. & Kondoh, M. (2010). Functional diversity of microbial decomposers facilitates plant coexistence in a plant-microbe-soil feedback model. *P Natl Acad Sci USA*, 107, 14251-14256.
139.
Moore, A.M. (1986). Temperature and Moisture Dependence of Decomposition Rates of Hardwood and Coniferous Leaf Litter. *Soil Biol Biochem*, 18, 427-435.
140.
Mouginot, C., Kawamura, R., Matulich, K.L., Berlemont, R., Allison, S.D., Amend, A.S. *et al.* (2014). Elemental stoichiometry of Fungi and Bacteria strains from grassland leaf litter. *Soil Biol Biochem*, 76, 278-285.
141.
Muller, C., Abbasi, M.K., Kammann, C., Clough, T.J., Sherlock, R.R., Stevens, R.J. *et al.* (2004). Soil respiratory quotient determined via barometric process separation combined with nitrogen-15 labeling. *Soil Sci Soc Am J*, 68, 1610-1615.
142.
Muller, C., Rutting, T., Kattge, J., Laughlin, R.J. & Stevens, R.J. (2007). Estimation of parameters in complex N-15 tracing models by Monte Carlo sampling. *Soil Biol Biochem*, 39, 715-726.
143.
Myrold, D.D. & Posavatz, N.R. (2007). Potential importance of bacteria and fungi in nitrate assimilation in soil. *Soil Biol Biochem*, 39, 1737-1743.
144.
Myrold, D.D., Zeglin, L.H. & Jansson, J.K. (2014). The potential of metagenomic approaches for understanding soil microbial processes. *Soil Sci Soc Am J*, 78, 3-10.
145.
Naeem, S. & Wright, J.P. (2003). Disentangling biodiversity effects on ecosystem functioning: deriving solutions to a seemingly insurmountable problem. *Ecol Lett*, 6, 567-579.
146.
Nelson, M.B., Berlemont, R., Martiny, A.C. & Martiny, J.B.H. (2015). Nitrogen cycling potential of a grassland litter microbial community. *Appl Environ Microb*, 81, 7012-7022.
147.
Nelson, M.B., Martiny, A.C. & Martiny, J.B.H. (2016). Global biogeography of microbial nitrogen-cycling traits in soil. *P Natl Acad Sci USA*, 113, 8033-8040.
148.
Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K. *et al.* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol*, 13, 135-144.
149.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F. *et al.* (2013a). Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev*, 77, 342-356.
150.
Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F. *et al.* (2013b). Patterns and Processes of Microbial Community Assembly. *Microbiol Mol Biol R*, 77, 342-356.
151.
Nielsen, U.N. & Ball, B.A. (2014). Impacts of altered precipitation regimes on soil communities and biogeochemistry in arid and semi-arid ecosystems. *Glob Chang Biol*.
152.
Nishio, T., Komada, M., Arao, T. & Kanamori, T. (2001). Simultaneous determination of transformation rates of nitrate in soil. *Jarq-Jpn Agr Res Q*, 35, 11-17.
153.
Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B. *et al.* (2013). vegan: Community Ecology Package, R package version 2.0-7
154.
Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M. *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res*, 33, 5691-5702.
155.
Parton, W. & Silver, W.L. (2007). Global-scale similarities in nitrogen release patterns during long-term decomposition (vol 315, pg 361, 2007). *Science*, 315, 940-940.
156.
Paul, E.A. & Clark, F.E. (1996). *Soil Microbiology and Biochemistry*. Academic, San Diego, California.
157.
Petersen, D.G., Blazewicz, S.J., Firestone, M., Herman, D.J., Turetsky, M. & Waldrop, M. (2012). Abundance of microbial genes associated with nitrogen cycling as indices of biogeochemical process rates across a vegetation gradient in Alaska. *Environ Microbiol*, 14, 993-1008.
158.
Pett-Ridge, J., Silver, W.L. & Firestone, M.K. (2006). Redox fluctuations frame microbial community impacts on N-cycling rates in a humid tropical forest soil. *Biogeochemistry*, 81, 95-110.
159.
Philippot, L., Bru, D., Saby, N.P.A., Cuhel, J., Arrouays, D., Simek, M. *et al.* (2009). Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environ Microbiol*, 11, 3096-3104.
160.
Potts, D.L., Suding, K.N., Winston, G.C., Rocha, A.V. & Goulden, M.L. (2012). Ecological effects of experimental drought and prescribed fire in a southern California coastal grassland. *J Arid Environ*, 81, 59-66.
161.

Prestat, E., David, M.M., Hultman, J., Tas, N., Lamendella, R., Dvornik, J. *et al.* (2014). FOAM (Functional Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res*, 42, e145.
162.
Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S. *et al.* (2015). Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecol Lett*, 18, 85-95.
163.
Prosser, J.I. (2015). Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. *Nat Rev Microbiol*, 13, 439-446.
164.
Pruesse, E., Peplies, J. & Glockner, F.O. (2012). SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, 28, 1823-1829.
165.
Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. *et al.* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, 41, D590-D596.
166.
Quinn, R.A., Lim, Y.W., Maughan, H., Conrad, D., Rohwer, F. & Whiteson, K.L. (2014). Biogeochemical Forces Shape the Composition and Physiology of Polymicrobial Communities in the Cystic Fibrosis Lung. *Mbio*, 5.
167.
Raes, J. & Bork, P. (2008). Systems microbiology - Timeline - Molecular eco-systems biology: towards an understanding of community function. *Nat Rev Microbiol*, 6, 693-699.
168.
Raes, J., Letunic, I., Yamada, T., Jensen, L.J. & Bork, P. (2011). Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol Syst Biol*, 7.
169.
Raich, J.W. & Schlesinger, W.H. (1992). The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus Series B Chemical and Physical Meteorology*, 44, 8.
170.
Ramette, A. (2007). Multivariate analyses in microbial ecology. *Fems Microbiol Ecol*, 62, 142-160.
171.
Ramirez, K.S., Craine, J.M. & Fierer, N. (2012). Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Global Change Biol*, 18, 1918-1927.
172.
Recous, S., Mary, B. & Faurie, G. (1990). Microbial Immobilization of Ammonium and Nitrate in Cultivated Soils. *Soil Biol Biochem*, 22, 913-922.
173.

Rediers, H., Vanderleyden, J. & De Mot, R. (2009). Nitrate respiration in Pseudomonas stutzeri A15 and its involvement in rice and wheat root colonization. *Microbiol Res*, 164, 461-468.
174.
Reed, H.E. & Martiny, J.B.H. (2013). Microbial composition affects the functioning of estuarine sediments. *Isme J*, 7, 868-879.
175.
Reichstein, M., Bahn, M., Mahecha, M.D., Kattge, J. & Baldocchi, D.D. (2014). Linking plant and ecosystem functional biogeography. *P Natl Acad Sci USA*, 111, 13697-13702.
176.
Reith, F., Drake, H.L. & Kusel, K. (2002). Anaerobic activities of bacteria and fungi in moderately acidic conifer and deciduous leaf litter. *Fems Microbiol Ecol*, 41, 27-35.
177.
Remus-Emsermann, M.N.P., Tecon, R., Kowalchuk, G.A. & Leveau, J.H.J. (2012). Variation in local carrying capacity and the individual fate of bacterial colonizers in the phyllosphere. *Isme J*, 6, 756-765.
178.
Resendis-Antonio, O., Hernandez, M., Salazar, E., Contreras, S., Batallar, G.M., Mora, Y. *et al.* (2011). Systems biology of bacterial nitrogen fixation: High-throughput technology and its integrative description with constraint-based modeling. *Bmc Syst Biol*, 5.
179.
Rice, C.W. & Tiedje, J.M. (1989). Regulation of Nitrate Assimilation by Ammonium in Soils and in Isolated Soil-Microorganisms. *Soil Biol Biochem*, 21, 597-602.
180.
Rocca, J.D., Hall, E.K., Lennon, J.T., Evans, S.E., Waldrop, M.P., Cotner, J.B. *et al.* (2015). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *Isme J*, 9, 1693-1699.
181.
Roesch, L.F., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D. *et al.* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *Isme J*, 1, 283-290.
182.
Rousk, J., Baath, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G. *et al.* (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *Isme J*, 4, 1340-1351.
183.
Rutting, T., Boeckx, P., Muller, C. & Klemedtsson, L. (2011). Assessment of the importance of dissimilatory nitrate reduction to ammonium for the terrestrial nitrogen cycle. *Biogeosciences*, 8, 1779-1791.
184.

Sanford, R.A., Wagner, D.D., Wu, Q.Z., Chee-Sanford, J.C., Thomas, S.H., Cruz-Garcia, C. *et al.* (2012). Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *P Natl Acad Sci USA*, 109, 19709-19714.
185.
Schimel, J. (1995). *Arctic and Alpine Biodiversity: Patterns, Causes, and Ecosystem Consequences*. Springer Verlag, New York.
186.
Schimel, J., Balser, T.C. & Wallenstein, M. (2007). Microbial stress-response physiology and its implications for ecosystem function. *Ecology*, 88, 1386-1394.
187.
Schimel, J.P. & Bennett, J. (2004). Nitrogen mineralization: Challenges of a changing paradigm. *Ecology*, 85, 591-602.
188.
Schimel, J.P. & Hattenschwiler, S. (2007). Nitrogen transfer between decomposing leaves of different N status. *Soil Biol Biochem*, 39, 1428-1436.
189.
Schimel, J.P. & Schaeffer, S.M. (2012). Microbial control over carbon cycling in soil. *Front Microbiol*, 3, 348.
190.
Schloss, P.D., Gevers, D. & Westcott, S.L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *Plos One*, 6.
191.
Schloss, P.D. & Handelsman, J. (2006). Toward a census of bacteria in soil. *Plos Comput Biol*, 2, 786-793.
192.
Seager, R., Ting, M.F., Held, I., Kushnir, Y., Lu, J., Vecchi, G. *et al.* (2007). Model projections of an imminent transition to a more arid climate in southwestern North America. *Science*, 316, 1181-1184.
193.
Silver, W.L., Herman, D.J. & Firestone, M.K. (2001). Dissimilatory nitrate reduction to ammonium in upland tropical forest soils. *Ecology*, 82, 2410-2416.
194.
Sinsabaugh, R.L., Hill, B.H. & Follstad Shah, J.J. (2009). Ecoenzymatic stoichiometry of microbial organic nutrient acquisition in soil and sediment. *Nature*, 462, 795-798.
195.
Smil, V. (1999). Nitrogen in crop production: An account of global flows. *Global Biogeochem Cy*, 13, 647-662.
196.
Sokal, R.R. & Rohlf, F.J. (2012). *Biometry: the principles and practice of statistics in biological research*. 4th edition edn. W. H. Freeman and Co., New York.
197.
Souza, R.C., Hungria, M., Cantao, M.E., Vasconcelos, A.T.R., Nogueira, M.A. & Vicente, V.A. (2015). Metagenomic analysis reveals microbial functional redundancies and specificities in a soil under different tillage and crop-management regimes. *Appl Soil Ecol*, 86, 106-112.

198.
Stone, M.M.D., J.L. ; Plante, A.F. (2014). Changes in extracellular enzyme activity and microbial community structure with soil depth at the Luquillo Critical Zone Observatory. *Soil Biol Biochem*, 75, 237-247.
199.
Strickland, M.S., Lauber, C., Fierer, N. & Bradford, M.A. (2009). Testing the functional significance of microbial community composition. *Ecology*, 90, 441-451.
200.
Strous, M., Fuerst, J.A., Kramer, E.H.M., Logemann, S., Muyzer, G., van de Pas-Schoonen, K.T. *et al.* (1999). Missing lithotroph identified as new planctomycete. *Nature*, 400, 446-449.
201.
Sun, W.M. & Cupples, A.M. (2012). Diversity of Five Anaerobic Toluene-Degrading Microbial Communities Investigated Using Stable Isotope Probing. *Appl Environ Microb*, 78, 972-980.
202.
Tas, N., Prestat, E., McFarland, J.W., Wickland, K.P., Knight, R., Berhe, A.A. *et al.* (2014). Impact of fire on active layer and permafrost microbial communities and metagenomes in an upland Alaskan boreal forest. *Isme J*, 8, 1904-1919.
203.
Team, R.D.C. (2011). R: a language and environment for statistical computing. (ed. Computing, RFfS) Vienna, Austria.
204.
Tedersoo, L., Bahram, M., Polme, S., Koljalg, U., Yorou, N.S., Wijesundera, R. *et al.* (2014). Global diversity and geography of soil fungi. *Science*, 346, 1256688.
205.
Thomas, T., Gilbert, J. & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial informatics and experimentation*, 2, 3.
206.
Thompson, P.L. & Gonzalez, A. (2017). Dispersal governs the reorganization of ecological networks under environmental change. *Nat Ecol Evol*, 1, 162.
207.
Thurber, R.V., Willner-Hall, D., Rodriguez-Mueller, B., Desnues, C., Edwards, R.A., Angly, F. *et al.* (2009). Metagenomic analysis of stressed coral holobionts. *Environ Microbiol*, 11, 2148-2163.
208.
Tiunov, A.V. & Scheu, S. (2005). Facilitative interactions rather than resource partitioning drive diversity-functioning relationships in laboratory fungal communities. *Ecol Lett*, 8, 618-625.
209.
Toth, I., Humphris, S., Campbell, E. & Pritchard, L. (2015). Why genomics research on *Pectobacterium* and *Dickeya* makes a difference. *Am J Potato Res*, 92, 218-222.
210.
Travisano, M., Mongold, J.A., Bennett, A.F. & Lenski, R.E. (1995). Experimental tests of the roles of adaptation, chance, and history in evolution. *Science*, 267, 87-90.
211.

Tringe, S.G. & Rubin, E.M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet*, 6, 805-814.
212.
Tsiknia, M., Paranychianakis, N.V., Varouchakis, E.A. & Nikolaidis, N.P. (2015). Environmental drivers of the distribution of nitrogen functional genes at a watershed scale. *Fems Microbiol Ecol*, 91, fiv052.
213.
van der Heijden, M.G.A., Bardgett, R.D. & van Straalen, N.M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol Lett*, 11, 296-310.
214.
van der Lee, G.E.M., de Winder, B., Bouten, W. & Tietema, A. (1999). Anoxic microsites in Douglas fir litter. *Soil Biol Biochem*, 31, 1295-1301.
215.
van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op den Camp, H.J.M., Kartal, B. *et al.* (2015). Complete nitrification by a single microorganism. *Nature*, 528, 555-559.
216.
Vannette, R. & Fukami, T. (2017). Dispersal enhances beta diversity in nectar microbes. *Ecol Lett*, 20, 901-910.
217.
Varin, T., Lovejoy, C., Jungblut, A.D., Vincent, W.F. & Corbeil, J. (2010). Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems. *Limnol Oceanogr*, 55, 1901-1911.
218.
Vega, N.M. & Gore, J. (2017). Stochastic assembly produces heterogeneous communities in the Caenorhabditis elegans intestine. *PLoS Biol*, 15, e2000633.
219.
Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *Q Rev Biol*, 85, 183-206.
220.
Violle, C., Reich, P.B., Pacala, S.W., Enquist, B.J. & Kattge, J. (2014). The emergence and promise of functional biogeography. *P Natl Acad Sci USA*, 111, 13690-13696.
221.
Vitousek, P.M. (1984). Litterfall, nutrient cycling, and nutrient limitation in tropical forests. *Ecology*, 65, 285-298.
222.
Vitousek, P.M., Aber, J.D., Howarth, R.W., Likens, G.E., Matson, P.A., Schindler, D.W. *et al.* (1997). Human alteration of the global nitrogen cycle: Sources and consequences. *Ecol Appl*, 7, 737-750.
223.
Wang, Q., Quensen, J.F., Fish, J.A., Lee, T.K., Sun, Y.N., Tiedje, J.M. *et al.* (2013). Ecological Patterns of nifH Genes in Four Terrestrial Climatic Zones Explored with Targeted Metagenomics Using FrameBot, a New Informatics Tool. *Mbio*, 4.
224.

Welsh, A., Chee-Sanford, J.C., Connor, L.M., Loffler, F.E. & Sanford, R.A. (2014). Refined NrfA phylogeny improves PCR-based nrfA gene detection. *Appl Environ Microb*, 80, 2110-2119.
225.
Whitaker, R.J., Grogan, D.W. & Taylor, J.W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*, 301, 976-978.
226.
Whitman, W.B., Coleman, D.C. & Wiebe, W.J. (1998). Prokaryotes: The unseen majority. *P Natl Acad Sci USA*, 95, 6578-6583.
227.
Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York.
228.
Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P. *et al.* (2016). The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Res*, 44, D590-D594.
229.
Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E.M., Kyrpides, N. *et al.* (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *Bmc Bioinformatics*, 13.
230.
Woods, D.D. (1938). The reduction of nitrate to ammonia by Clostridium welchii. *Biochem J*, 32, 2000-2012.
231.
Wu, D.Y., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N. *et al.* (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, 462, 1056-1060.
232.
Yergeau, E., Kang, S., He, Z., Zhou, J. & Kowalchuk, G.A. (2007). Functional microarray analysis of nitrogen and carbon cycling genes across an Antarctic latitudinal transect. *Isme J*, 1, 163-179.
233.
Yoon, S., Cruz-Garcia, C., Sanford, R., Ritalahti, K.M. & Loffler, F.E. (2015). Denitrification versus respiratory ammonification: environmental controls of two competing dissimilatory NO3-/NO2- reduction pathways in Shewanella loihica strain PV-4. *Isme J*, 9, 1093-1104.
234.
Yooseph, S., Andrews-Pfannkoch, C., Tenney, A., McQuaid, J., Williamson, S., Thiagarajan, M. *et al.* (2013). A Metagenomic Framework for the Study of Airborne Microbial Communities. *Plos One*, 8.
235.
Yu, K. & Zhang, T. (2012). Metagenomic and Metatranscriptomic Analysis of Microbial Community Structure and Gene Expression of Activated Sludge. *Plos One*, 7.
236.

Yuste, J.C., Penuelas, J., Estiarte, M., Garcia-Mas, J., Mattana, S., Ogaya, R. *et al.* (2011). Drought-resistant fungi control soil organic matter decomposition and its response to temperature. *Global Change Biol*, 17, 1475-1486.
237.
Zhang, J.J., Kobert, K., Flouri, T. & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30, 614-620.
238.
Zhou, J.Z., Liu, W.Z., Deng, Y., Jiang, Y.H., Xue, K., He, Z.L. *et al.* (2013). Stochastic Assembly Leads to Alternative Communities with Distinct Functions in a Bioreactor Microbial Community. *Mbio*, 4.