LINKING SOIL MICROBIOLOGY AND ENVIRONMENTAL CONDITIONS TO VARIABILITY IN NITROUS OXIDE PRODUCTION IN BIOENERGY CROPPING SYSTEMS

by

David Sean Duncan

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

(Agronomy)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 7/14/2016

The dissertation is approved by the following members of the Final Oral Committee:

Randall D. Jackson, Professor, Agronomy

Garret Suen, Assistant Professor, Bacteriology

David E. Stoltenberg, Professor, Agronomy

Paul J. Weimer, Associate Professor, Bacteriology

Thomas M. Schmidt, Professor, Microbiology and Immunology, University of Michigan

ACKNOWLEDGEMENTS

A dissertation this long and resource-intensive was only possible because of my advisor, Randy Jackson, who throughout the process displayed the patience and generosity of an early Amazon investor. The structure of the dissertation reflects an attempt to match the standards he set for narrative-based scientific inquiry. The Great Lakes Bioenergy Research Center provided the financial and logistical resources to conduct this work but also defined its overarching conceptual framework. I am particularly indebted to Steve Slater, possibly the first and last person this decade who will ever say "You need 200 soil metagenomes? Sounds good, let's do it." I received a humbling degree of support from the Department of Agronomy and College of Agriculture and Life Sciences, who honored me with enough fellowships to melt down two, possibly three master rings, affording an incredible degree of latitude to experiment and explore. The administrative staff in both Agronomy and GLBRC, particularly Sandy, Joanna, Amy, Lauren, Pam, and Mary, were positively aquiline in their capacity to mitigate failures in my planning.

The members of my committee were amazing throughout this process, neither losing patience with me or forgetting my name during long periods of radio silence. My interactions with them and with their labs shaped this project. Tom Schmidt and Tracy Teal inspired the functional gene approach at the heart of this project, showing me that addressing ecological questions with metagenomic data did not need to be a torturous process. Dave Stoltenberg helped me maintain focus on cropping systems and their agronomics, and while I rarely interacted with his lab in a professional capacity, he consistently attracted the Agronomy students with whom I got along best. Garret Suen continually drove me to contextualize my work in the broader field whenever I became too inward-looking; my work in his lab with Kelsea Jewell and Tony Neumann was the first time this project really started gaining traction. Paul Weimer's Environmental Microbiology course kicked off this entire interest in microbial ecology, and Dave Stevenson and Chris Odt from his lab ultimately helped me make it through the morass of soil DNA extraction.

I thought myself self-sufficient until I ran out of fingers while counting people who had helped me. Matt Raboin was there for the first soil sampling, with its protean protocols. Michaela Miller was a guided

missile of a minion, helping me set up and run field experiments, attempt manual bioinformatics, and prep samples; she allowed me to get away with a style of lazy supervising that I have since struggled to shed. Nicole Tautges walked with me into the abyss of trace gas flux estimation methodology. I was nothing but bad news for Zoe Andersen, who got stuck collecting samples for me from an overgrown field site during a heat wave and running quality control on trace gas data. I inflicted my soil incubation prototypes on Carly Vinkavich, then dragged her away from her beloved archiving work to process my soils. Alex Henkel turned out to be a thinly-disguised self-cleaning machine who generated half a thesis's worth of data over the 12 months I worked with him. David Williams and Harry Read generated the PLFA data from the first two chapters. I profited tremendously from the unbroken lineage of the guardians of order in the field and the lab. Andrew Dean founded the tradition, leading the field crews during the establishment of the cropping systems trials that generated most of my data. His successor, James Tesmer, kept everything working while sustaining tolerable casualties; he also built the computer that ran the analyses that resulted in the figures that made Chapter 5 happen. Jimmy Sustachek is the latest line of defense against encroaching chaos, as well as the source of our soil nitrogen data. Casey Menick became our collective scientific conscience, defusing a terrifying array of errors lurking in our datasets. Numerous other technicians, student workers, and volunteers contributed many hours under all manner of conditions to contribute to the data that are being presented. Some of them didn't give me much in the way of data, but made working in the lab more enjoyable. Drea Miller generated both a fun atmosphere and strong déjà vu. Haley MeLampy provided a low-key sense that things were not about to erupt into flame in the near future. Jess Mayry helped increase the overall nerd quotient of any conversation.

While I rarely exploited their data-generating efforts, the graduate students and other researchers in the Jackson lab contributed tremendously to my understanding of ecology and academia. Julie Doll helped create the welcoming atmosphere that first drew me to the lab, while also setting a ridiculously high bar for graduate student performance. Herika Kummel blazed the trail for both the agroecology degree and the subsequent transition to a PhD. Sue Chamberlain was a wonderful enabler of my pyromania and coached our ragtag softball team all the way to, well, at least one win, I think. Craig Maier was a

continual source of novel mechanisms for characterizing community data. Andy (now Dr. Mr. Jakubowski, sir) was a useless undergrad assistant but a great peer who shaped my understanding and approach toward statistics, literature management, and the general conduct of science. Jessica Meisel helped nudge me down the slippery path to soil ecology and was one of my first windows into ecological research traditions outside of UW-Madison. Maddy Raudenbush really propelled my adoption of ggplot and introduced me to viewing the conduct of science as inextricably bound with social dynamics. Erica Diehl brought landscape design into the lab conversation and was a wonderful person with whom to engage in general nerdery. Anna Herzberger was the best REU student ever and showed me the joy of watching a mentee develop into a capable scientist. Brianna Laube Duran kept the tradition of lab-based social activity alive, but she had her good points too and her research has been extremely valuable for shaping our narrative of nitrogen cycling in perennial systems. My collaborations with Chao Liang have been highly enjoyable and contribute disproportionately to the length of my CV. Even though it didn't quite work out in the end, getting jazzed about the duckfuels project with Carol Williams was both fun and educational. I don't know if Stacey Marion was ever officially part of the lab, but it was fun to have her attend our lab meetings and challenge some of our thought habits. Mike Cruse may as well have been a member of the lab; I am so sorry for having egged him into Bayesian statistics before I knew what that entailed. People still in the lab are pretty okay too. I've enjoyed going full action movie buddy with Gary Oates on manuscripts and data analysis projects, and have always benefited from his willingness to allow a lack of preparation on my part to constitute an emergency on his. Gregg Sanford has been a source of wide-ranging impromptu conversations and is allegedly involved in some way with the systems I studied. Laura Smith has been an endless font of interesting data analysis and interpretation puzzles and the victim of an unhealthy number of my end of the workday kvetches. I'm glad that Anna Cates will be keeping microbial ecology research going in the lab. Adam von Haden asks the methodological questions we fear, but need, to be asked. I'm hopeful that Greta Landis will help the lab find its way back to its agroecological and grazing roots. Karen Stahlheber-Nikolakakis has provided a wonderfully fresh perspective in our lab meetings, and showed me the fun to be had with geek material culture.

I've benefited from interactions with multiple people outside of our lab as well. The people at JGI generated all of my metagenomic data and tolerated my continual delays with good grace. I'm also grateful they allowed me to serve on their CSP review panels, which were responsible for two of the happiest moments in my academic career. I never found a way to collaborate formally with Julie Sinistore, but it was good to make common cause with a fellow crunchy agroecologist. Hannah Gaines-Day, Mary Bulan, and Alex Lyon provided great dissertation writing support, or would have if they hadn't graduated before I started writing. We had some great collaborators at Michigan State and KBS. Mis platicas con Sarynna Lopez Mesa siempre fueron de las partes mas divertidas de las reuniones del Centro. It was great to collaborate with Phil Robertson along with Sarah, Sven, Kevin, Neville, Ilya, and Suzanne as we worked to understand how to properly work with our environmental dataset. Jim Tiedje's lab, particularly Aaron Garoutte, Jiaronng Guo, and Adina Howe helped shape my understanding of metagenomics and microbial community analysis, and were welcome familiar faces at both internal and external conferences.

Support networks are for wimps. That said, Amy and Martin's stream of graduations, relocations, weddings, good weekends to visit and general stuff kept my world from becoming too small and enclosed. My two best non-consanguineous, non-work friends, Ryan and Naya, independently provided considerable roundabout encouragement for my completion of this degree and more direct encouragement of a general tolerance for the reality I inhabit. The rest of my family were real troopers throughout this process, tolerating exceptionally poor performances as a son/brother/nephew/cousin/etc. without cutting me from the team. James, Gunnar, Amanda, and Andrew (and by extension, Reed, Kriv, Drusilia, and Tsab respectively) gamely sallied forth into a world filled with arbitrarily shifting rules, poor exposition, and underwhelming loot drops. That was cool of them.

There are giants in this field. Some are so large and have been here so long that you can't stand anywhere without stepping on them. Others might pass for normal, if not for the speed at which they are growing. I spent insufficient time on their shoulders, but what I saw from there was exhilarating. I am very excited to explore the world they are discovering.

TABLE OF CONTENTS

CHAPTER 3

Bioenergy feedstock cropping systems exert significant but inconsistent effects on soil microbial functional gene abundances

CHAPTER 4

Effects of environmental constraints on soil N_2O fluxes differ among potential bioenergy feedstock cropping systems

CHAPTER 5

Aggregate annual nitrous oxide emissions correlate to soil microbial functional gene abundance profiles at the plot level and may reflect inherent nitrous oxide production capacities in bioenergy feedstock cropping systems

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

ABSTRACT

Agroecosystems may differ in multiple ecosystem properties, among them nitrous oxide (N_2O) production and soil microbial community composition. We hypothesized that perenniality, plant species richness, and exogenous nitrogen inputs all influence N_2O production directly through regulation of substrate concentrations and other environmental conditions and indirectly through changes to soil microbial functional characteristics. We studied the interplay among cropping systems, microbial communities, and N_2O production in the context of an agronomic trial of potential bioenergy feedstock cropping systems.

We measured N_2O production from 2009-2014 and collected accompanying data on soil temperature, water-filled pore space, and inorganic nitrogen concentrations. Individual N_2O fluxes and aggregate annual N_2O emissions were lower in perennial systems relative to annual ones, but were not consistently influenced by plant species richness in perennial systems. Environmental variables defined upper limits for N_2O fluxes, but did little to explain cropping system effects or their lack.

We explored microbial community differences between continuous corn and prairie systems using membrane lipid profiling, amplicon sequencing, and functional gene annotations from shotgun metagenomic sequencing. The strength of cropping system effects differed among methods, with the strongest effects observed in lipid profiles. We used elastic net modeling to correlate community profiles to aggregate N_2O emissions. Only the corn system could be effectively modeled, with the best models created from 16S rRNA amplicons and functional gene abundances.

We used bacterial functional gene abundance profiles to characterize microbial communities across a broader range of cropping systems. The strength of cropping system effects varied among site years. Ecological factors such as perenniality and species diversity did not determine abundance patterns for either the full set of genes explored or for groups of genes with similar functions. Similarly, individual denitrification pathway genes did not systematically differ among cropping systems.

Cropping system effects on N_2O production and functional gene abundances were weaker than anticipated. Despite this, elastic net modeling linked gene abundance patterns to variation in N_2O emissions with considerable accuracy. This indicates that within-cropping system variability in N_2O production and functional genes are in some way connected.

PROLOGUE

Project background and rationale

This dissertation presents an attempt to link the inherent nitrous oxide (N_2O) production capacities of bioenergy feedstock cropping systems to the functional genetic composition of their soil microbial communities. This project formed part of the US Department of Energy-funded Great Lakes Bioenergy Research Center (GLBRC) portfolio of sustainability research. GLBRC is one of three Bioenergy Research Centers funded through the surge of public investment in energy research that resulted from the oil crisis in the mid-2000s (Slater et al., 2010). The GLBRC research portfolio emphasized the ecological impacts of bioenergy feedstock production from the outset, considering both risks to be avoided and potential for beneficial ecological changes (Robertson et al., 2008). The Jackson Grassland Ecology Lab's expertise with studying biogeochemical cycling in agroecosystems (Bleier and Jackson, 2007; Craine and Jackson, 2010; Jackson et al., 2006) situated it naturally to take a leading role in the study of greenhouse gas emissions and nitrogen dynamics in potential bioenergy feedstock cropping systems (Duran et al., 2016; Oates et al., 2016). As GLBRC research projects were being set up, I found myself in possession of a genetics-based hammer (Duncan et al., 2015) and a growing interest in microbe-shaped nails.

To paraphrase one frequently excited soil ecologist, microbiology has always been exiting, but the period in which I conducted this research was special. The ongoing development and increasing accessibility of culture-free microbial community characterization methods revealed extensive underestimates of microbial diversity, while also providing new and exciting tools to explore that diversity (Hirsch et al., 2010; Singh et al., 2009). Genetic methods led to major changes in the understanding of microbial taxa that involved in major biogeochemical transformations such as ammonia oxidation (Leininger et al., 2006) and nitrous oxide reduction (Sanford et al., 2012). At the same time, a growing body of evidence illustrated the critical importance of soil microbial activity and communities on ecosystem functioning, notably in biogeochemical cycling (van der Heijden et al., 2008). Despite this, the connection of microbial community composition and structure to ecosystem functionality remained difficult to demonstrate (Choudoir et al., 2012; Nannipieri et al., 2003) and as such provided a tempting

target for an ambitious ecologist. In particular, the microbial ecology of bioenergy feedstock cropping systems intersected with my research interests as well as those of the GLBRC and the Jackson Lab.

Soil microbial ecology formed an integral part of the GLBRC sustainability research portfolio from the outset. Early work characterized microbial communities in extant agricultural and established grassland systems (Jesus et al., 2016; Liang et al., 2012, 2016), and in short-rotation woody crop systems (Xue et al., 2016). In the realm of ecological theory, the relationship between microbial diversity and function were investigated (Levine et al., 2011; Werling et al., 2014). The Bioenergy Cropping Systems Experiments (BCSEs) at the University of Wisconsin-Madison Arlington Agricultural Research Station (ARL) and the Michigan State University W.K. Kellogg Biological Station (KBS) provided a platform where a broad array of cropping systems could be compared in an agronomic experimental framework (Sanford et al., 2016). Work on these systems emphasized community fingerprinting or Characterization of these systems' microbial communities largely entailed taxonomic profiling or emphasis on particular functional genes (Herzberger et al., 2014; Jesus et al., 2016; Liang et al., 2016). Needing a niche, I focused my work on functional gene profiling.

While most ecology focuses on organisms or communities as the basic unit of inquiry, complex microbial communities lend themselves to a gene-centric approach (Tringe et al., 2005; Wilmes et al., 2008). The concept of the gene as a meaningful unit for ecological selection reflects the potential disconnect between microbial taxonomy and function enabled by horizontal/lateral gene transfer (Chia and Goldenfeld, 2011; Lawrence, 2002). In many cases, communities of organisms performing a function can be defined and detected through indicator genes involved in that function, e.g. nitrate reductase (*nirK/S*) for denitrifiers (e.g. Yoshida et al., 2010) or nigrogenase (*nifH*) for diazotrophs (e.g. Wakelin et al., 2009). At its most direct, this approach examines the relative abundance of a functional gene within a population as a means of evaluating the ecological importance of that function (Petersen et al., 2012). Simultaneous abundance profiling of large numbers of genes was initially done through microarrays (He et al., 2010), a technology whose familiarity no doubt contributed to my enthusiasm for gene-centric science (Ma et al., 2007). While the confluence of increased availability of high-throughput sequencing

and profit-driven international business decisions ultimately resulted in my use of shotgun metagenomic sequencing rather than microarrays, functional gene abundance profiles remained the primary mechanism by which I characterized soil microbial communities in this project.

With the means of investigating microbial communities thus settled, I needed to specify which of their effects to study. My initial interests involved soil and/or agroecosystem "health" (Kibblewhite et al., 2008). My readings at the time suggested that microbial communities could be highly sensitive integrators of complex functions and behaviors (Ritz et al., 2009; Winding et al., 2005). As the nigh-impossibility of operationally defining such concepts set in, I focused on the production of trace gasses with global warming potential, notably CO_2, N_2O, and CH_4. Inability to isolate heterotrophic respiration ruled out CO_2 as an informative response, while our estimates of CH_4 fluxes were too noisy to instill much confidence, leaving what was fortunately the most interesting of the three. In many senses, N_2O production in soil provides an ideal framework for studying the effects of microbial community composition on an ecologically relevant process. The details of this are presented extensively, and somewhat repetitively, throughout this dissertation, but briefly stated, N_2O is produced by relatively simple biochemical pathways whose key genetic components are well-characterized and found broadly, if variably, throughout microbial populations (Philippot and Hallin, 2005). The gas itself is one of the chief mechanisms by which agricultural sectors in developed countries contribute to global warming (Hu et al., 2015). Thus, this system provided a compelling ecological question, a sizeable body of knowledge to study it, and clear practical motivations for studying it.

Structural overview

My dissertation project was conducted on the BCSEs at ARL and KBS. This framework enabled me to largely isolate cropping system effects from the effects of climate and physical soil properties and provided a statistically robust experimental design. Nitrous oxide and other environmental data were collected from these experiments as one of the core data-generating activities of the GLBRC, creating a far more extensive dataset than I could have generated myself. The dissertation consists of five data and analysis chapters:

Chapter 1 compares the soil microbial communities of continuous corn and restored prairie at ARL using four characterization methods: membrane lipid profiling, 16S rRNA and nitrous oxide reductase (*nosZ*) amplicon sequencing, and shotgun sequencing. This study served multiple purposes. It allowed comparison of the shallow shotgun metagenomic sequencing approach employed in this dissertation to more established methods for community characterization. It gave me an opportunity to go through the process of generating and analyzing sequencing data using a dataset of tractable size. Finally, by comparing the two most ecologically dissimilar cropping systems in the BCSE, it was a test of the feasibility of my underlying approach. The method comparison and repeated annual measurements were relatively novel in the literature, resulting in this chapter's publication.

Chapter 2 correlates the microbial community data from Chapter 1 to N_2O emissions data generated in a prior publication (Oates et al., 2016). This is the chapter that almost wasn't, as it was initially intended as a proof of concept exercise rather than a planned publication. However, posters and presentations of this exercise generated sufficient interest to motivate packaging it into what I hope to make a standalone publication. This chapter forced an early confrontation with the challenge of bringing potentially thousands of predictors from sequencing data to bear on the prediction of, at most, a few dozen flux observations.

Chapter 3 explores the microbial ecology of BCSE cropping systems, focusing on the extent to which systems shape characteristic functional gene profiles in their associated microbial communities. It was necessary to determine the extent to which microbial communities differed among and within cropping systems in order to evaluate whether they contained sufficient information to capture variance in an ecosystem property like N_2O production potential.

Chapter 4 focuses on how environmental conditions constrain individual N_2O flux events. The initial intent with this chapter was to focus on generating estimators of inherent N_2O production, but several factors changed that. The most exciting estimation method required considerable structural overhead to describe, and ultimately did not work well at all. At the same time, initially minor analysis of ecological constraints proved unexpectedly compelling. In the end, it was cleaner to present fluxes and

their environmental constraints separately from estimators of N_2O production and their prediction by microbial data.

Chapter 5 is an odd duck. It generates four estimators of inherent N_2O production capacity, which it then models from functional gene profiles using the elastic net approach explored in Chapter 2. This chapter makes sense within the context of the dissertation, but would make for an unwieldy publication. This chapter contains elements that contribute to the stories in Chapters 3 and 4, as well as some analyses that failed to provide results worth publishing. Its tone and structure differ from those of the preceding chapter, with the aim of communicating the rather dense set of analyses and methods as clearly as possible.

A brief general conclusion outlines a proposed framework for publishing the contents of this dissertation. This primarily entails moving analyses from Chapter 5 to Chapters 3 and 4 to broaden the story. In this section, I also present some themes that occurred across the chapters, but were not necessarily their focus.

References

Bleier, J.S., Jackson, R.D., 2007. Manipulating the quantity, quality, and manner of C addition to reduce soil inorganic N and increase C4:C3 grass biomass. Restoration Ecology 15, 688–695.

Chia, N., Goldenfeld, N., 2011. Statistical mechanics of horizontal gene transfer in evolutionary ecology. Journal of Statistical Physics 142, 1287–1301.

Choudoir, M.J., Campbell, A.N., Buckley, D.H., 2012. Grappling with Proteus: Population level approaches to understanding microbial diversity. Frontiers in Microbiology 3, 1–5.

Craine, J.M., Jackson, R.D., 2010. Plant nitrogen and phosphorus limitation in 98 North American grassland soils. Plant and Soil 334, 73–84.

Duncan, D.S., Krohn, A.L., Jackson, R.D., Casler, M.D., 2015. Conservation implications of the introduction history of meadow fescue (*Festuca pratensis* Huds.) to the Driftless Area of the Upper Mississippi Valley, USA. Plant Ecology & Diversity 8, 91–99.

Duran, B.E.L., Duncan, D.S., Oates, L.G., Kucharik, C.J., Jackson, R.D., 2016. Nitrogen fertilization

effects on productivity and nitrogen loss in three grass-based perennial bioenergy cropping systems. PLOS ONE 11, e0151919.

He, Z., Deng, Y., van Nostrand, J.D., et al., 2010. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. The ISME Journal 4, 1167–1179.

Herzberger, A.J., Duncan, D.S., Jackson, R.D., 2014. Bouncing back: Plant-associated soil microbes respond rapidly to prairie establishment. PLoS ONE 9, e115775.

Hirsch, P.R., Mauchline, T.H., Clark, I.M., 2010. Culture-independent molecular techniques for soil microbial ecology. Soil Biology & Biochemistry 42, 878–887.

Hu, H.-W., Chen, D., He, J.-Z., 2015. Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates. FEMS Microbiology Reviews 39, 729–749.

Jackson, R.D., Allen-Diaz, B., Oates, L.G., Tate, K.W., 2006. Spring-water nitrate increased with removal of livestock grazing in a California oak savanna. Ecosystems 9, 254–267.

Jesus, E. da C., Liang, C., Quensen, J.F., Susilawati, E., Jackson, R.D., Balser, T.C., Tiedje, J.M., 2016. Influence of corn, switchgrass, and prairie cropping systems on soil microbial communities in the upper Midwest of the United States. GCB Bioenergy 8, 481–494.

Kibblewhite, M.G., Ritz, K., Swift, M.J., 2008. Soil health in agricultural systems. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 1492, 685–701.

Lawrence, J.G., 2002. Gene transfer in bacteria: speciation without species? Theoretical Population Biology 61, 449–460.

Leininger, S., Urich, T., Schloter, M., et al., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature 442, 806–809.

Levine, U.Y., Teal, T.K., Robertson, G.P., Schmidt, T.M., 2011. Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. The ISME Journal 5, 1683–1691.

Liang, C., Jesus, E. da C., Duncan, D.S., Jackson, R.D., Tiedje, J.M., Balser, T.C., 2012. Soil microbial

communities under model biofuel cropping systems in southern Wisconsin, USA: Impact of crop species and soil properties. Applied Soil Ecology 54, 24–31.

Liang, C., Jesus, E. da C., Duncan, D.S., Quensen, J.F., Jackson, R.D., Balser, T.C., Tiedje, J.M., 2016. Switchgrass rhizospheres stimulate microbial biomass but deplete microbial necromass in agricultural soils of the upper Midwest, USA. Soil Biology & Biochemistry 94, 173–180.

Ma, J., Duncan, D., Morrow, D.J., Fernandes, J., Walbot, V., 2007. Transcriptome profiling of maize anthers using genetic ablation to analyze pre-meiotic and tapetal cell types. The Plant Journal 50, 637–48.

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G., 2003. Microbial diversity and soil functions. European Journal of Soil Science 54, 655–670.

Oates, L.G., Duncan, D.S., Gelfand, I., Millar, N., Robertson, G.P., Jackson, R.D., 2016. Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. GCB Bioenergy 8, 539–549.

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. Environmental Microbiology 14, 993–1008.

Philippot, L., Hallin, S., 2005. Finding the missing link between diversity and activity using denitrifying bacteria as a model functional community. Current Opinion in Microbiology 8, 234–239.

Ritz, K., Black, H.I.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. Ecological Indicators 9, 1212–1221.

Robertson, G.P., Dale, V.H., Doering, O.C., et al., 2008. Sustainable biofuels redux. Science 322, 49–50.

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems and Environment 216, 344–355.

Sanford, R.A., Wagner, D.D., Wu, Q., et al., 2012. Unexpected nondenitrifier nitrous oxide reductase

gene diversity and abundance in soils. Proceedings of the National Academy of Sciences of the United States of America 109, 19709–19714.

Singh, J., Behal, A., Singla, N., et al., 2009. Metagenomics: Concept, methodology, ecological inference and recent advances. Biotechnology Journal 4, 480–494.

Slater, S., Keegstra, K., Donohue, T.J., 2010. The US Department of Energy Great Lakes Bioenergy Research Center: Midwestern biomass as a resource for renewable fuels. BioEnergy Research 3, 3–5.

Tringe, S.G., von Mering, C., Kobayashi, A., et al., 2005. Comparative metagenomics of microbial communities. Science 308, 554–557.

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. Ecology Letters 11, 296–310.

Wakelin, S.A., Gregg, A.L., Simpson, R.J., Li, G.D., Riley, I.T., McKay, A.C., 2009. Pasture management clearly affects soil microbial community structure and N-cycling bacteria. Pedobiologia 52, 237–251.

Werling, B.P., Dickson, T.L., Isaacs, R., et al., 2014. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. Proceedings of the National Academy of Sciences of the United States of America 111, 1652–1657.

Wilmes, P., Simmons, S.L., Denef, V.J., Banfield, J.F., 2008. The dynamic genetic repertoire of microbial communities. FEMS Microbiology Reviews 33, 109–32.

Winding, A., Hund-Rinke, K., Rutgers, M., 2005. The use of microorganisms in ecological soil classification and assessment concepts. Ecotoxicology and Environmental Safety 62, 230–48.

Xue, C., Penton, C.R., Zhang, B., et al., 2016. Soil fungal and bacterial responses to conversion of open land to short-rotation woody biomass crops. GCB Bioenergy 1–14.

Yoshida, M., Ishii, S., Otsuka, S., Senoo, K., 2010. *nirK*-harboring denitrifiers are more responsive to denitrification-inducing conditions in rice paddy soil than *nirS*-harboring bacteria. Microbes and Environments 25, 45–48.

CHAPTER 1

Detection of short-term cropping system-induced changes to soil bacterial communities differs among four molecular characterization methods

David S Duncan [a,b,*], Kelsea A Jewell [c†], Garret Suen [c], Randall D Jackson [a,b]

[a] Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA

[b] DOE-Great Lakes Bioenergy Research Center, Department of Bacteriology, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

[c] Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706, USA

[†] Current address: Biological Sciences Program, Centralia College, 600 Centralia College Blvd, Centralia, WA 98531, USA

(This chapter contains minor differences from the form published in Soil Biology & Biochemistry (2016) **96**:160-168, reflecting recommendations from the dissertation committee.)

Abstract

Perennial grass-based agroecosystems are under consideration as sustainable sources of bioenergy feedstocks. Establishing these systems on land previously used for conventional agricultural production is expected to dramatically alter the composition and functional capacity of their associated soil bacterial communities, but the rate at which these changes will occur is unclear. Methods for characterizing bacterial communities are both varied and useful for documenting different aspects of the soil microbiota and their dynamics during this transition. Here, we studied the soil-associated bacterial communities of continuous corn and restored prairies systems within a cropping systems experiment 2 to 4 years after establishment using 1) phospholipid fatty acid (PLFA) profiling, 2) shotgun metagenomic sequencing, 3) amplicon sequencing of the 16S rRNA gene and 4) sequencing of nitrous oxide reductase (*nosZ*). All characterization methods discriminated the bacterial communities between the two cropping systems, but the largest differences were observed with PLFA profiling. Differences between the two cropping systems did not significantly increase during the study period. The community compositions described by sequence-based methods were mutually correlated, but were only weakly correlated to the composition described by PLFA profiling. Shotgun metagenomics detected a much higher abundance of Actinobacteria than amplicon sequencing and revealed more consistent changes between cropping systems over time. Cropping system and interannual effects on the ratios of biomarkers associated with Gram-negative and Gram-positive bacteria were entirely different for PLFAs, rRNA amplicons, and shotgun-sequenced 16S rRNA. Because these characterization methods reflect different aspects of the bacterial community, none are clearly superior or optimal, but instead all provide insight into how communities respond to cropping system effects.

1.1 Introduction

The interaction between soil microbial communities and their environment is a major driver of terrestrial agroecosystem dynamics. Microbial communities directly influence agroecosystem productivity and plant community composition (Callaway et al., 2004; Kardol et al., 2007), soil carbon dynamics (Bardgett et al., 2008), and major nutrient cycles (Hawkes et al., 2005; Nannipieri et al., 2003). Agroecosystems in turn influence the composition and function of soil microbial communities (Oehl et al., 2010; Sayer et al., 2013). The move to develop cropping systems that can produce biomass feedstocks for bioenergy and bioproducts in a sustainable manner has motivated increased interest in the ecological effects of introducing diverse perennial systems onto land previously used for agriculture (Robertson et al., 2008), including the effect on microbial communities (Li et al., 2015; Mao et al., 2013). There is substantial evidence that the composition and functional capabilities of a soil microbial community change when transitioning between conventional agricultural and less intensively managed perennial systems (Allison et al., 2005; Liang et al., 2012; Xue et al., 2013). However, the timing of microbial community change following such a transition remains unclear.

Soil microbial communities shift on a decadal timescale following establishment of perennial systems on soils with a previous history of annual cropping (Allison et al., 2005). Long-term shifts are not monotonic, with systems passing through transitional states as they continue to mature (Jangid et al., 2010); short-term dynamics are less well characterized. Soil microbial communities exhibit less temporal variation than those in other environments (Shade et al., 2013) possibly because of the prevalence of dormancy in soil microbiota (Lennon and Jones, 2011), and microbial community compositions may reflect land uses from decades in the past (Jangid et al., 2011). This apparent resistance to change suggests transitions following land use change should occur in consistent, progressive steps toward major transitional or terminal states. At the same time, there is evidence that soil microbial communities are susceptible to external perturbation (Allison and Martiny, 2008) and that community succession can occur over a single growing season (Schmidt et al., 2007). From this perspective, microbial community transitions could be uneven in the short-term, exhibiting expected patterns only over longer periods of

time. Given the linkage between microbial community composition and function (Frey et al., 2004; Reed and Martiny, 2007; Wakelin et al., 2008), the consistency with which soil microbial communities change following land use conversion could influence the variability of microbially-mediated functions during the establishment phase of perennial cropping systems.

Modern efforts to track changes in soil microbial communities typically employ culture-free characterization, particularly sequencing-based methods that interrogate the metagenome (Hirsch et al., 2010). These metagenomes are dominated by bacterial sequences (Fierer et al., 2012), effectively causing most recent work in soil microbiology to primarily reflect bacterial communities and dynamics. Many characterization methods focus on distinct aspects of a community, such as the taxonomic makeup of individuals or the relative abundance of functional genes, which could potentially differ in their responsiveness to land use change.

We previously used phospholipid fatty acid (PLFA) profiling to examine the soil microbial community of continuous corn and sown tallgrass prairie cropping systems during the two to four years following their establishment on historically agricultural soil (Herzberger et al., 2014). In the present study, we reexamine these soils, focusing on their bacterial component. To better resolve the phylogenetic and functional changes undergone by this community, we supplemented the bacterial PLFA data with amplicon sequencing of 16S rRNA and *nosZ*, a gene that identifies denitrifiers, as well as functional genes and 16S rRNA sequences derived from shotgun metagenomics sequencing (SMG). Our goals were to compare how these four distinct methods characterized the divergence between the soil bacterial communities of recently-established corn (*Zea mays* L.) and sown tallgrass prairie, and evaluate whether yearly increases in this divergence were visible despite interannual variability in community composition.

1.2 Methods

1.2.1 Site description and soil sampling

This experiment was conducted on the Bioenergy Cropping Systems Experiment (BCSE) at Arlington Agricultural Research Station (Arlington, Wisconsin, USA, 43°18′10 N, 89°20′40 W). The BCSE consisted of eight cropping systems, including continuous corn and sown tallgrass prairie, in a

randomized complete block design with five replicates (full details are given in Sanford et al., 2016). Prior to 2005, the site was used for agronomic trials and production, with corn, soybean, and alfalfa as the primary crops. From 2005 to 2007, the site was under a hayfield mix of alfalfa (*Medicago sativa* L.) and orchardgrass (*Dactylis glomerata* L.) (blocks 1 to 3) or corn (blocks 4 and 5). All treatments were established in spring 2008. The experiment is situated on a highly productive Plano silt loam (Fine-silty, Mixed, Superactive, Mesic Typic Arguidolls). These are deep (>1 m), well-drained soils with little relief, formed under tallgrass prairie vegetation in loess deposits over calcareous glacial till. Mean annual minimum and maximum air temperatures are -14.6 and 27.6 °C, respectively, with 869 mm mean annual precipitation.

Soils were sampled in late August in 2010, 2011, and 2012. For each plot, 5 cores (3.7 cm diameter, 15 cm depth) were collected at variable, arbitrary distances (<5 m) from a transect running lengthwise through the plot. Cores were immediately placed on wet ice, sieved to 2 mm within 24 h of sampling, and frozen at -20 °C within 48 h of sampling. Frozen soils were lyophilized for storage prior to lipid and DNA extraction. All molecular characterization methods were performed on each individual sample.

1.2.2 PLFA extraction and quantification

We employed a combined phospholipid fatty acid (PLFA) and fatty acid methyl ester (FAME) extraction method, as described in Herzberger et al. (2014). Soils were extracted with 1:2 $CHCl_3$:CH_3OH with the extracted lipids saponified with NaOH and methylated with HCl. Lipids were analyzed with a Hewlett-Packard Agilent 6890A gas chromatograph (Agilent Tech. Co., Santa Clara, CA) equipped with a 25-m×0.2-mm×0.33-µm Agilent Ultra-2 (5% phenyl)-methylpolysiloxane capillary column (Hewlett Packard, Palo Alto, CA) and flame ionization detector. Fatty acids were identified using MIDI's EUKARY method database. We restricted our analysis to PLFAs frequently associated with bacteria, as described in Table S1.1 (Balser and Firestone, 2005; Frostegård et al., 1993; Hill et al., 2000; Wilkinson, 1988; Zelles et al., 1992).

1.2.3 DNA extraction

DNA was extracted from the soil using an adaptation of a method developed by Stevenson and Weimer (2007). Lyophilized soil was ground with liquid nitrogen in a ceramic mortar to disrupt aggregates. We prepared two tubes per sample, with each tube receiving 0.5 g ground soil (1 g total), 0.5 g 0.1-mm silica-zirconia beads (BioSpec Products, Bartlesville, OK), 1 ml extraction buffer (EB; 100 mM Tris-HCl, 10 mM EDTA, 0.15 M NaCl, pH 8.0), 50 µl 20% SDS, and 500 µl cold phenol (buffer equilibrated to pH 7.9). Tubes were bead beaten for 10 min on a converted paint shaker, incubated 10 min at 60 °C, then beaten an additional 10 min. Samples were separated by centrifugation (16,000 × *g*, 10 min) and the aqueous layer was washed successively with 500 µl phenol, 500 µl 1:1 phenol:chloroform, and 500 µl chloroform, with centrifugation (16,000 × *g*, 10 min) at each step and the aqueous layer brought to 1 ml volume with EB. To precipitate, 900 µl of aqueous layer was combined with 100 µl 3 M Na acetate and 600 µl isopropanol, incubated for 30 min at 4 °C and centrifuged (16,000 × *g*, 20 min). The DNA pellet was washed once with 70% ethanol, then both tubes from a sample were recombined in 150 µl 1/10 TE (10 mM Tris pH8, 0.1 mM EDTA). These samples were subsequently cleaned with a Power Soil Cleanup Kit (Mo Bio Laboratories, Carlsbad, CA) following manufacturer instructions.

1.2.4 DNA sequencing and classification

We characterized the composition of the overall and denitrifying bacterial communities using 16S rRNA and *nosZ* amplicons. We amplified the V6-V8 variable region of the bacterial 16S rRNA gene using primers described by de Oliveira et al. (2013) and a region of the catalytic subunit of *nosZ* using nosZF (Kloos et al., 2001) and nosZ1622R (Throbäck et al., 2004) primers. Primer sequences are given in Table S1.2. Segments were amplified using 5-ng template DNA and 0.5 µM of each primer in 25 µl volume using Platinum Blue PCR SuperMix (Life Technologies, Carlsbad, CA). Polymerase chain reaction (PCR) was performed following de Oliveria et al (2013): denaturation for 2 min at 94 °C; 30 cycles of 30 s at 94 °C, 45 s at 50 °C, and 1.75 min at 68 °C; and a final extension for 10 min at 68 °C. We limited our amplification to 30 cycles to allow for detection of low-copy template and to ensure the overwhelming majority of amplicons would be created in the 15-25 cycle range (Lee et al., 1996). PCR products were

size-selected by electrophoresis on a 1% Aqua*Pōr* LM low-melt agarose (National Diagnostics, Atlanta, GA) TAE gel and recovered with a Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA). Sequencing was carried out on a Roche 454 GS Junior Titanium sequencer following manufacturer protocols and using Lib-L emulsion PCR kits (Roche Life Sciences, Indianapolis, IN) with ~0.8 DNA molecules per bead.

The program mothur v. 1.33.3 (Schloss et al., 2009) was used to process 16S rRNA amplicon sequences. Reads were denoised using *shhh.flows*. Quality filtration removed sequences under 200 nucleotides or with >6 nucleotide homopolymers, >2 primer mismatches, or any barcode mismatches. Alignment was done against the SILVA reference database (Pruesse et al., 2007), removing sequences of eukaryotic origin. Chimeric sequences were removed using *chimera.uchime*. Operational Taxonomic Units (OTUs) were defined at 95% sequence identity and a representative sequence for each OTU was taxonomically classified with the Ribosomal Database Project Classifier (Wang et al., 2007) using an assignment cutoff of 0.8.

Processing of *nosZ* amplicons was similarly carried out in mothur through the quality filtration step. Further processing was conducted through the Functional Gene Pipeline (Fish et al., 2013). Unique sequences were translated with Framebot (Wang et al., 2013) with a minimum length of 66 amino acids and an identity cutoff of 0.3. These translations were aligned with Aligner (Cole et al., 2014), then clustered using mcClust (Fish et al., 2013). *NosZ* OTUs were classified at 90% amino acid sequence similarity, following Mao et al. (2013).

Shotgun sequencing was carried out at the Department of Energy Joint Genome Institute using the Illumina HiSeq platform (Illumina, San Diego, CA), multiplexing 12 samples per lane. Trimmed and screened reads were assembled using SOAPdenovo (v1.05). Gene calling was done with FragGenScan (v1.16), prokaryotic GeneMark.hmm (v2.8), Metagenome Annotator (v1.0), and Prodigal (v2.50). Reads identified as belonging to the 16S rRNA genes were taxonomically classified in the same fashion as the 16S rRNA amplicons. Reads were annotated using the updated clusters of orthologous groups (COGs) database (Galperin et al., 2015; Tatusov et al., 2000).

1.2.5 Statistical analysis

Data were analyzed in the R statistical package and environment (v3.1.1, R Core Team, 2014). Multivariate analyses used the 'vegan' package (Oksanen et al., 2013). Prior to multivariate analysis, individual PLFAs, functional genes, or OTUs were removed from the data if not present in at least three out of five blocks and two out of three years. We did not relativize PLFA values by sample, as PLFA extractions are quantitative (Liang et al., 2012). COG abundances were relativized following He et al. (2015). COG read counts were first relativized by the length of their consensus sequence to neutralize the overrepresentation of longer sequences. We then used the mean abundances of 37 COGs present in a single copy in nearly all prokaryotic genomes (Table S1.3) to relativize across samples. Amplicon read counts were relativized to the total number of reads in the sample. Bray-Curtis distances were calculated for all relativized measures. This metric was not appropriate for PLFA, which used absolute distances; Euclidian distances were used here instead.

Standard statistical comparisons used linear mixed effects models from the 'lme4' package (Bates et al., 2015) using block as a random term. Means testing was done with the 'lsmeans' package (Lenth, 2013).

1.2.6 Accession numbers

PLFA data are available on the Dryad database (http://datadryad.org) under doi: 10.5061/dryad.rk384. Amplicon sequences were deposited in the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) as BioProject 279094. Shotgun metagenomes are in the Joint Genome Institute Genomes Online Database (https://gold.jgi.doe.gov) under Study ID Gs0095510. Accession numbers for individual samples are given in Table S1.4.

1.3 Results

1.3.1 Sequencing overview

Shotgun metagenomic (SMG) sequencing generated 1.3 to 2.0 Gbp per sample (48.3 Gbp total), with 2.4 to 4.0×10^6 functional gene reads of which 2,151 to 8,135 reads per sample mapped to the 16S rRNA.

Diversity and coverage statistics for individual samples are given in Table S1.4. With amplicon sequencing, we generated 2,382 to 6,206 reads per sample for 16S rRNA and 1,116 to 6,681 reads per sample for *nosZ*. We found 3,717 operational taxonomic units (OTUs) using 16S rRNA amplicon sequencing (490 to 750 OTUs per sample), and 1,072 *nosZ* OTUs (150 to 290 per sample). Good's coverage for 16S rRNA amplicons (89 to 95%, 98.5% globally) was lower than for *nosZ* amplicons (93 to 99%, 99.6% globally). Shannon's diversity indices for 16S rRNA OTUs were lower for 2010 corn than for all other crop-year combinations, which were not different from each other, while Shannon's diversity indices for *nosZ* OTUs were the same for all crop-year combinations (Table S1.5). Read number was correlated to unique OTUs for 16S rRNA amplicons ($r = 0.79$, $P < 0.0001$), but not for *nosZ* amplicons ($r = 0.14$, $P = 0.22$). Within samples, unique OTUs for both amplicon types were correlated ($r = 0.77$, $P < 0.0001$), but total read numbers were not ($r = 0.21$, $P = 0.13$).

1.3.2 Factors determining microbial community composition

We used *adonis*, a permutational analogue to multivariate analysis of variance, to quantify treatment effects on overall community composition (Table 1.1). All characterization methods found statistically significant cropping system and interannual effects, although interannual effects were only weakly significant in the PLFA and *nosZ* data. The amount of variation these factors explained differed among methods. Much of the variation in the PLFA data existed between cropping systems, while interannual effects were less important. Conversely, more of the variance in the SMG COG and rRNA amplicon data existed among years than between systems. These were the only methods to have significant interactions between years and cropping systems, likely driven by the 2010 continuous corn samples (Fig. 1.1). Soil carbon and nitrogen concentrations did not change between 2009 and 2013 and were not different between the two cropping systems (Table S1.5). Soil bulk density did not differ between cropping systems in 2009, but was lower overall in 2013, more so in the sown tallgrass prairie than in the continuous corn system (Table S1.5).

1.3.3 Ordination of community compositions

Two-dimensional NMDS ordination adequately represented the community compositions found by all four characterization methods (Fig. 1.1), although two-dimensional stress was greater for the two amplicon sequencing datasets (Table 1.2). PLFA profiles of the continuous corn and restored prairie bacterial communities were clearly distinct, with greater variability within the prairie communities (Fig. 1.1A). By contrast, in both rRNA amplicon and SMG COG profiles there was more variability in the corn system than the prairie system, largely stemming from the differentiation of 2010 corn-associated soil from all other systems (Fig. 1.1 B, C). Despite overall separation of the bacterial communities from both cropping systems, there was no indication that this separation changed over the study period. In the *nosZ* amplicon data, both corn and prairie communities shifted in the same direction in 2011 before reversing that transition in 2012 (Fig. 1.1 D).

The three DNA-based characterization methods were correlated both in the Bray-Curtis distances between pairs of samples and in the distances among samples following NMDS ordination (Table 1.2). The SMG COG and 16S rRNA amplicon data were particularly well correlated. Inter-sample distance correlations between PLFA and the DNA-based data were significant, but generally weaker (for exact values, see Table 1.2).

1.3.4 Taxonomic profiles from amplicon and short read sequencing

We obtained 16S rRNA sequences from amplicons and SMG data, with roughly equal numbers of reads generated from both methods (Table S1.4). At the phylum level, both datasets were dominated by Actinobacteria, Acidobacteria, and Proteobacteria (Fig. 1.2, Table S1.6). In both datasets, Verrucomicrobia abundance increased over time and was higher in sown tallgrass prairie than corn cultivation. Similarly, both datasets observed high γ-Proteobacteria abundance in the corn system in 2010.

Beyond these similarities, the taxonomic profiles generated by amplicon and SMG sequencing contained substantial differences. The proportions of reads assigned to a given phylum differed significantly between the two sequencing approaches for all common phyla (Table S1.6, common phyla defined as accounting for >5% of all reads from at least one of the systems). Actinobacteria were twice as

abundant in the SMG sequences as in the amplicon sequences. The SMG sequence data also contained cropping system and temporal trends that were not present in the amplicon sequence data. In the SMG data, Firmicutes and Bacteroidetes were significantly more abundant in sown tallgrass prairie than continuous corn cultivation. SMG reads from major Gram-positive phyla (Actinobacteria and Firmicutes) became rarer over time, while reads from rare or unidentified taxa became more abundant.

1.3.5 Functional group biomass and gene abundance

We compared ratios of PLFAs that are frequently associated with either Gram-positive or Gram-negative bacteria (Table S1.1) to ratios of SMG and amplicon 16S rRNA reads grouped by phylum-level cell wall type (Gram-positive or Gram-negative, Table S1.6). Interannual and cropping system effects on these ratios differed among the three community characterization methods (Fig. 1.3). PLFA biomass associated with Gram-negative taxa increased in the prairie system, but did not change over time. In contrast, SMG 16S rRNA reads belonging to Gram-negative phyla were not different between cropping systems but became more prevalent over time in the prairie system. There were no interannual or cropping system effects in the 16S rRNA amplicons.

1.4. Discussion

1.4.1 Characterization methods differed in their detection of cropping system divergence

In this study, we used four methods to characterize change over time in the bacterial communities of continuous corn and sown tallgrass restored prairie cropping systems. All methods revealed significant differences between systems, but varied in the extent of the differences they showed. The greatest cropping system effects were observed using bacterial PLFA profiles, which produced similar results to those previously observed using PLFAs from the entire microbial community (Herzberger et al., 2014). PLFAs associated with Gram-negative bacteria were more abundant in the prairie, matching the difference observed between corn and prairie sites throughout south-central Wisconsin (Liang et al., 2012). Differences between cropping systems were considerably smaller, but still significant, for the DNA-based methods.

Biomass-based microbial community measures, such as PLFA, may be more responsive to environmental change than those based on DNA. Prior studies have observed responses to seasonal variability (Jangid et al., 2010, 2011) and vegetation composition (Ritz et al., 2004) in lipid profiles, but not in 16S rRNA data. The apparently limited responsiveness of DNA-based characterization methods may in part stem from the prevalence of dormant organisms (Lennon and Jones, 2011) as well as the potential persistence of DNA from nonviable organisms (Córdova-Kreylos et al., 2006), both of which could delay detection of changes in the microbial community. Our results highlight the continued value of lipid profiling as a valuable tool for assessing microbial community change and a useful complement to metagenomic methods.

We expected that the sown tallgrass restored prairie system would gradually and progressively shift its associated bacterial community toward compositions typically found under restored prairies in this region (Allison et al., 2005; Liang et al., 2012). Progressive divergence from the corn system bordered on statistical significance in the PLFA data, similar to increasing trends in AMF, Gram-negative, and total microbial biomass we had previously reported (Herzberger et al., 2014), and suggested that continued measurement of the systems would have captured increasing divergence. For DNA-based methods, however, we saw no indication of increasing differentiation during the study period. Increases in PLFA biomass are detectable in the top 5 cm of soil shortly after prairie restoration, but are only visible at greater depths once the restoration has been in place for considerably longer (Allison et al., 2007). This likely reflects the increased availability of oxygen and resources near the soil surface, due in part to the greater concentration of plant roots. Sampling from the top 15 cm of soil may have diluted the more rapidly shifting near-surface microbial community. This dilution may not have been as impactful for PLFA profiles, which should reflect the concentration of microbial biomass in the top 5 cm of soil (Fierer et al., 2003), as it would be for DNA-based methods, potentially explaining the apparent difference in their responsiveness. At the same time, we observed considerable interannual variability in bacterial communities, suggesting our observations were not driven solely by a slower response to land use change by bacteria deeper in the soil profile. The dramatic shift in the composition of corn associated bacteria

between 2010 and 2011, observed in both the 16S rRNA and SMG data, best exemplifies this. Soil microbial communities may exhibit considerable seasonal variability (Schmidt et al., 2007). Some of the interannual variability we observed may have been due to slight differences in the seasonal timing of our sampling. Nonetheless, all methods detected differences between the two cropping systems only two years after their establishment. The change in disturbance regimes may have spurred the initial rapid differentiation (Allison et al., 2005; Jangid et al., 2010), with broader cropping-system differences driving more gradual shifts in the community.

A variable not directly addressed in our work is the impact of storage on sample quality. There are varying views on the appropriate time a soil may be stored unfrozen prior to microbial analyses (Bloem et al., 2005). Although it is generally accepted that freezing and extraction should happen as soon after sampling as possible, some guidelines suggest soils may be stored under refrigeration for several weeks prior to PLFA extraction (Palojärvi, 2005) and in well-drained soils no changes in PLFA biomass were observed over 30 days of refrigeration (Wang et al., 2014), although other studies report changes within shorter time periods (Wu et al., 2009). Soils from our two cropping system might have responded differently to storage. Agricultural soils may be less perturbed by storage than grassland or forest soils (Cui et al., 2014), possibly because microbial communities in the latter systems typically receive more labile carbon (Gonzalez-Quiñones et al., 2009). Although no differences in total soil carbon developed during our study period (section 3.2), it is possible that a difference in the pool of available carbon could have gone unnoticed. If so, the sown tallgrass prairie, with its more active rhizosphere, could have lost more biomass during storage than the corn-associated soils, causing us to underestimate divergence between the two systems.

1.4.2 DNA-based methods identified similar community compositions

Soil bacterial community composition was well correlated among DNA-based methods. Fierer et al. (2012) reported high correlation between taxonomic composition and functional gene abundance over an ecological gradient that included arctic tundra, desert, and tropical forests. Our findings suggest this relationship may hold within a gradient of ecologically similar systems. Although many functional genes

in bacteria are not restricted to a narrow set of taxa (Ragan and Beiko, 2009), there are likely linkages between taxonomic and functional composition at the community level (Langille et al., 2013).

We observed limited *nosZ* diversity, dominated by a small number of taxa, similar to previously published results (Ruiz-Rueda et al., 2009). It is also worth noting that the *nosZ* primers we used do not capture 'atypical' *nosZ* genes (Orellana et al., 2014; Sanford et al., 2012), indicating our analyses exclude a sizeable component of the denitrifying community. It is, however, interesting that the number of unique *nosZ* types we observed in a sample correlated strongly with the number of unique 16S rRNA OTUs. This suggests that the subset of the *nosZ*-containing community we sampled largely responded to broad drivers of bacterial diversity, even though *nosZ* community composition did not display the same patterns as compositions based on taxonomic or functional gene abundances.

1.4.3 Characterization methods detected different abundances of specific groups

The abundance of specific bacterial groups was more influenced by interannual and cropping system effects in the SMG sequences than the 16S rRNA amplicon sequences. There were phylum-level abundance differences between the cropping systems in the SMG data that were not present in the amplicon data, but the inverse was not true. SMG-derived 16S rRNA sequences have several advantages over amplicon sequences, notably in the avoidance of primer biases (Logares et al., 2013). Primer biases were likely behind the underrepresentation of Actinobacteria in our 16S rRNA amplicon data. Fierer et al. (2012) found a similar underrepresentation of Actinobacteria in amplicon data, albeit using a different 16S rRNA region. Although abundance of Bacteroidetes and Firmicutes did not differ markedly between methods, only SMG data detected cropping system differences in their abundance. The increased abundance of γ-Proteobacteria in the 2010 continuous corn soils stands as an interesting counterpoint, as this dynamic was present in both datasets. It seems reasonable that γ-Proteobacteria were similarly well detected by both methodologies, and further implies that many responses to cropping system and interannual effects occur in bacterial groups that were less well detected by our amplicon sequencing.

The interpretation of certain PLFAs as indicative of Gram-positive or Gram-negative bacteria is problematic, both because PLFAs are not strict phylogenetic markers (Frostegård et al., 2011), and

because Gram staining is of decreasing utility as a framework for classifying bacteria. Nonetheless, this terminology remains common in soil microbiology, possibly because these sets of lipids frequently exhibit coherent and interesting ecological behaviors. Rhizosphere-extracted PLFAs associated with Gram-negative bacteria are enriched in carbon from plant exudates (Butler et al., 2003), while both sets of PLFAs can correlate to distinct nitrogen cycling processes in the soil (Balser and Firestone, 2005). In our earlier study, we found that PLFAs associated with Gram-negative organisms increased in the sown tallgrass prairie while PLFAs associated with Gram-positive organisms were similar for both systems (Herzberger et al., 2014). In this study, we tested whether these PLFA-based patterns behaved similarly to DNA from phyla that could be classified as Gram-positive or Gram-negative, finding that the two methods had entirely different dynamics. This may indicate a mismatch between classification methods, microbial activity vs abundance, or be another instance of the differences in how PLFAs and DNA respond to environmental influences (section 4.1). Our findings provide further evidence that comparison of this set of fatty acids can be an ecologically informative aspect of PLFA profiles. We hope this further highlights the need to revisit the use of the Gram-negative/-positive framework for interpreting these fatty acids and the potential utility of identifying the taxa that these biomarkers actually represent.

1.5 Conclusions

The detected soil bacterial community composition of continuous corn and restored tallgrass restored prairie cropping systems were found to be different by all four characterization methods we employed. The systems differed more in the composition of their bacterial community biomass, as detected by PLFA profiling, than in their functional gene or taxonomic compositions as detected by DNA amplicon sequencing. Differences between the two cropping systems did not increase over the three years that we sampled the soils. All characterization methods identified some effect of prairie establishment, but with DNA-based methods interannual variability appeared to have an equivalent impact. Overall community compositions detected by the alternative methods were fairly well correlated, although abundance estimates for specific taxa could vary substantially among methods. We found that soil microbial

communities could change considerably from year to year, but that these changes did not contribute toward a progressive divergence or convergence between two very different land uses.

Acknowledgments

We thank DA Williams for PLFA extraction, HW Read for PLFA quantification, A Neumann for assistance with amplicon sequencing, and K Stahlheber and two anonymous reviewers for helpful comments. Funding was provided by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830). K Jewell was supported by a USDA NIFA Fellowship Grant 2012-01193. The sequencing work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Funding sources were not directly involved in the study design, data analysis, manuscript preparation, or decision to publish.

References

Allison, S.D., Martiny, J.B.H., 2008. Resistance, resilience, and redundancy in microbial communities. Proceedings of the National Academy of Sciences of the United States of America 105, 11512–11519.

Allison, V.J., Miller, R.M., Jastrow, J.D., Matamala, R., Zak, D.R., 2005. Changes in soil microbial community structure in a tallgrass prairie chronosequence. Soil Science Society of America Journal 69, 1412–1421.

Allison, V.J., Yermakov, Z., Miller, R.M., Jastrow, J.D., Matamala, R., 2007. Using landscape and depth gradients to decouple the impact of correlated environmental variables on soil microbial community composition. Soil Biology and Biochemistry 39, 505–516.

Balser, T.C., Firestone, M.K., 2005. Linking microbial community composition and soil processes in a California annual grassland and mixed-conifer forest. Biogeochemistry 73, 395–415.

Bardgett, R.D., Freeman, C., Ostle, N.J., 2008. Microbial contributions to climate change through carbon cycle feedbacks. The ISME Journal 2, 805–814.

Bates, D., Maechler, M., Bolker, B.M., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1 – 48.

Bloem, J., Schouten, A.J., Sørensen, S.J., Rutgers, M., Werf, A. van der, Breure, A.M., 2005. Monitoring and evaluating soil quality, in: Bloem, J., Hopkins, D.W., Benedetti, A. (Eds.), Microbiological Methods for Assessing Soil Quality. CABI, Wallingford, pp. 23–49.

Butler, J.L., Williams, M.A., Bottomley, P.J., Myrold, D.D., 2003. Microbial community dynamics associated with rhizosphere carbon flow. Applied and Environmental Microbiology 69, 6793–6800.

Callaway, R.M., Thelen, G.C., Barth, S., Ramsey, P.W., Gannon, J.E., 2004. Soil fungi alter interactions between the invader Centaurea maculosa and North American natives. Ecology 85, 1062–1071.

Cole, J.R., Wang, Q., Fish, J.A., et al., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research 42, D633–D642.

Córdova-Kreylos, A.L., Cao, Y., Green, P.G., et al., 2006. Diversity, composition, and geographical distribution of microbial communities in California salt marsh sediments. Applied and Environmental Microbiology 72, 3357–66.

Cui, H., Wang, C., Gu, Z., Zhu, H., Fu, S., Yao, Q., 2014. Evaluation of soil storage methods for soil microbial community using genetic and metabolic fingerprintings. European Journal of Soil Biology 63, 55–63.

de Oliveira, M.N.V., Jewell, K.A., Freitas, F.S., et al., 2013. Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology 164, 307–314.

Fierer, N., Leff, J.W., Adams, B.J., et al., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences of the United States of America 109, 21390–21395.

Fierer, N., Schimel, J.P., Holden, P.A., 2003. Variations in microbial community composition through two soil depth profiles. Soil Biology and Biochemistry 35, 167–176.

Fish, J.A., Chai, B., Wang, Q., Sun, Y., Brown, C.T., Tiedje, J.M., Cole, J.R., 2013. FunGene: the functional gene pipeline and repository. Frontiers in Microbiology 4, 291.

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 Ecology and Management 196, 159–171.

Frostegård, Å., Bååth, E., Tunlid, A., Amebrant, K., 1993. Shifts in the structure of soil microbial communities in limed forests as revealed by phospholipid fatty acid analysis. Soil Biology & Biochemistry 25, 723–730.

Frostegård, Å., Tunlid, A., Bååth, E., 2011. Use and misuse of PLFA measurements in soils. Soil Biology & Biochemistry 43, 1621–1625.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E. V, 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research 43, D261–9.

Gonzalez-Quiñones, V., Banning, N.C., Ballesta, R.J., Murphy, D. V, 2009. Influence of cold storage on soil microbial community level physiological profiles and implications for soil quality monitoring. Soil Biology and Biochemistry 41, 1574–1576.

Hawkes, C. V., Wren, I.F., Herman, D.J., Firestone, M.K., 2005. Plant invasion alters nitrogen cycling by modifying the soil nitrifying community. Ecology Letters 8, 976–985.

He, S., Malfatti, S.A., McFarland, J.W., et al., 2015. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. mBio 6, e00066–15.

Herzberger, A.J., Duncan, D.S., Jackson, R.D., 2014. Bouncing back: plant-associated soil microbes respond rapidly to prairie establishment. PLoS ONE 9, e115775.

Hill, G.T., Mitkowski, N.A., Aldrich-Wolfe, L., et al., 2000. Methods for assessing the composition and diversity of soil microbial communities. Applied Soil Ecology 15, 25–36.

Hirsch, P.R., Mauchline, T.H., Clark, I.M., 2010. Culture-independent molecular techniques for soil microbial ecology. Soil Biology & Biochemistry 42, 878–887.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Blair, J.M., Coleman, D.C., Whitman, W.B., 2010. Development of soil microbial communities during tallgrass prairie restoration. Soil Biology and Biochemistry 42, 302–312.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Schmidt, T.M., Coleman, D.C., Whitman, W.B., 2011. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. Soil Biology & Biochemistry 43, 2184–2193.

Kardol, P., Cornips, N.J., van Kempen, M.M.L., Bakx-Schotman, J.M.T., van der Putten, W.H., 2007. Microbe-mediated plant-soil feedback causes historical contingency effects in plant community assembly. Ecological Monographs 77, 147–162.

Kloos, K., Mergel, A., Rösch, C., Bothe, H., 2001. Denitrification within the genus *Azospirillum* and other associative bacteria. Functional Plant Biology 28, 991–998.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., et al., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31, 814–821.

Lee, S.-Y., Bollinger, J., Bezdicek, D., Ogram, A., 1996. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. Applied and Environmental Microbiology 62, 3787–3793.

Lennon, J.T., Jones, S.E., 2011. Microbial seed banks: The ecological and evolutionary implications of dormancy. Nature Reviews. Microbiology 9, 119–130.

Lenth, R. V, 2013. lsmeans: Least-squares means [WWW Document]. R Package Version 1.06-05. URL http://cran.r-project.org/package=lsmeans

Li, D., Voigt, T.B., Kent, A.D., 2015. Plant and soil effects on bacterial communities associated with *Miscanthus* × *giganteus* rhizosphere and rhizomes. GCB Bioenergy 8, 183-193.

Liang, C., Jesus, E. da C., Duncan, D.S., Jackson, R.D., Tiedje, J.M., Balser, T.C., 2012. Soil microbial communities under model biofuel cropping systems in southern Wisconsin, USA: Impact of crop species and soil properties. Applied Soil Ecology 54, 24–31.

Logares, R., Sunagawa, S., Salazar, G., et al., 2013. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environmental Microbiology 16, 2659–2671.

Mao, Y., Yannarell, A.C., Davis, S.C., Mackie, R.I., 2013. Impact of different bioenergy crops on N-

cycling bacterial and archaeal communities in soil. Environmental Microbiology 15, 928–942.

Nannipieri, P., Ascher, J., Ceccherini, M.T., Landi, L., Pietramellara, G., Renella, G., 2003. Microbial diversity and soil functions. European Journal of Soil Science 54, 655–670.

Oehl, F., Laczko, E., Bogenrieder, A., Stahr, K., Bösch, R., van der Heijden, M.G.A., Sieverding, E., 2010. Soil type and land use intensity determine the composition of arbuscular mycorrhizal fungal communities. Soil Biology & Biochemistry 42, 724–738.

Oksanen, J., Blanchet, F.G., Kindt, R., et al., 2013. vegan: Community ecology package [WWW Document]. R Package Version 2.0-10. URL http://cran.r-project.org/package=vegan

Orellana, L.H., Rodriguez-R, L.M., Higgins, S., et al., 2014. Detecting nitrous oxide reductase (*nosZ*) genes in soil metagenomes: method development and implications for the nitrogen cycle. mBio 5, e01193–14.

Palojärvi, A., 2005. Phospholipid fatty acid (PLFA) analyses, in: Bloem, J., Hopkins, D.W., Benedetti, A. (Eds.), Microbiological Methods for Assessing Soil Quality. CABI, Wallingford, pp. 204–211.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glöckner, F.O., 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35, 7188–96.

R Core Team, 2014. R: A language and environment for statistical computing [WWW Document]. Version 3.1.1. URL http://www.r-project.org

Ragan, M.A., Beiko, R.G., 2009. Lateral genetic transfer: Open issues. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 364, 2241–2251.

Reed, H.E., Martiny, J.B.H., 2007. Testing the functional significance of microbial composition in natural communities. FEMS Microbiology Ecology 62, 161–170.

Ritz, K., McNicol, J.W., Nunan, N., et al., 2004. Spatial structure in soil chemical and microbiological properties in an upland grassland. FEMS Microbiology Ecology 49, 191–205.

Robertson, G.P., Dale, V.H., Doering, O.C., et al., 2008. Sustainable biofuels redux. Science 322, 49–50.

Ruiz-Rueda, O., Hallin, S., Bañeras, L., 2009. Structure and function of denitrifying and nitrifying

bacterial communities in relation to the plant species in a constructed wetland. FEMS Microbiology Ecology 67, 308–319.

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems & Environment 216, 344–355.

Sanford, R.A., Wagner, D.D., Wu, Q., et al., 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proceedings of the National Academy of Sciences of the United States of America 109, 19709–19714.

Sayer, E.J., Wagner, M., Oliver, A.E., Pywell, R.F., James, P., Whiteley, A.S., Heard, M.S., 2013. Grassland management influences spatial patterns of soil microbial communities. Soil Biology & Biochemistry 61, 61–68.

Schloss, P.D., Westcott, S.L., Ryabin, T., et al., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75, 7537–7541.

Schmidt, S.K., Costello, E.K., Nemergut, D.R., et al., 2007. Biogeochemical consequences of rapid microbial turnover and seasonal succesion in soil. Ecology 88, 1379–1385.

Shade, A., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., 2013. A meta-analysis of changes in bacterial and archaeal communities with time. The ISME Journal 7, 1493–1506.

Stevenson, D.M., Weimer, P.J., 2007. Dominance of Prevotella and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75, 165–174.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E. V, 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research 28, 33–6.

Throbäck, I.N., Enwall, K., Jarvis, A., Hallin, S., 2004. Reassessing PCR primers targeting *nirS*, *nirK* and *nosZ* genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiology Ecology 49, 401–417.

Wakelin, S.A., Macdonald, L.M., Rogers, S.L., Gregg, A.L., Bolger, T.P., Baldock, J.A., 2008. Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. Soil Biology & Biochemistry 40, 803–813.

Wang, J., Chapman, S.J., Yao, H., 2014. The effect of storage on microbial activity and bacterial community structure of drained and flooded paddy soil. Journal of Soils and Sediments 15, 880–889.

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

Wang, Q., Quensen, J.F., Fish, J.A., Lee, T.K., Sun, Y., Tiedje, J.M., Cole, J.R., 2013. Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. mBio 4, e00592–13.

Wilkinson, S.G., 1988. Gram negative bacteria, in: Ratledge, C., Wilkinson, S.G. (Eds.), Microbial Lipids. Academic Press, London, pp. 299–428.

Wu, Y., Ding, N., Wang, G., Xu, J., Wu, J., Brookes, P.C., 2009. Effects of different soil weights, storage times and extraction methods on soil phospholipid fatty acid analyses. Geoderma 150, 171–178.

Xue, K., Wu, L., Deng, Y., et al., 2013. Functional gene differences in soil microbial communities from conventional, low-input, and organic farmlands. Applied and Environmental Microbiology 79, 1284–1292.

Zelles, L., Bai, Q.Y., Beck, T., Beese, F., 1992. Signature fatty acids in phospholipids and lipopolysaccharides as indicators of microbial biomass and community structure in agricultural soils. Soil Biology & Biochemistry 24, 317–323.

Table 1.1 Factors accounting for multivariate variance from four methods assessing microbial community change over time between corn and prairie cropping systems

		PLFA			**SMG functional genes**			**16S rRNA amplicon OTUs**			***nosZ* amplicon OTUs**		
Factor	***df***	SS	R^2	*P*	SS	R^2	*P*	SS	R^2	*P*	SS	R^2	*P*
Cropping system	1	0.071	0.62	<0.01	0.001	0.13	<0.01	0.173	0.09	<0.01	0.352	0.15	<0.01
Year	2	0.008	0.07	0.04	0.003	0.23	<0.01	0.361	0.20	<0.01	0.225	0.10	0.05
System × Year	2	0.006	0.06	0.07	0.001	0.08	0.04	0.181	0.10	<0.01	0.069	0.03	0.99
Residuals	24	0.029	0.25		0.007	0.56		1.121	0.61		1.682	0.72	
Total	29	0.114			0.012	1.00		1.836			2.327		

PLFA: bacterial phospholipid fatty acids; SMG: shotgun metagenomic sequencing; OTU: operational taxonomic unit; SS: sum of squares. Values estimated using *adonis*. *P*-values based on 9999 data permutations.

Table 1.2 Correlations of community distance matrices and nonmetric multidimensional scaling (NMDS) ordinations among four methods of assessing microbial community composition

	Bray-Curtis distance (Mantel statistic)				**NMDS distance (Procrustes correlation)**				
Measurement type	PLFA	COGs	16S rRNA	*nosZ*	PLFA	COGs	16S rRNA	*nosZ*	**NMDS linear R^2**
PLFA	----	0.180**	0.168**	0.313***	----	0.352*	0.417*	452**	0.999
COGs	0.180**	----	0.750***	0.461***	0.352*	----	0.747***	0.542***	0.964
16S rRNA OTU	0.168*	0.750***	----	0.557***	0.417*	0.747***	----	0.576***	0.915
nosZ cluster	0.313***	0.461***	0.557***	----	0.452**	0.542***	0.576***	----	0.890

PLFA: bacterial phospholipid fatty acids; COGs: clusters of orthologous groups of proteins, based on shotgun metagenomic sequence. NMDS linear R^2 is the coefficient of determination between Bray-Curtis and NMDS ordination distances for all pairs of samples. Significance of correlation-like statistics was based on 9999 random permutations: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Table S1.1 Interpretations of PLFAs

Functional group	**PLFAs**
Gram-negative bacteria	a15:0
	i15:0
	a17:0
	i17:0
Gram-positive bacteria	16:1ω7c
	17:0cy
	18:1ω5c
	18:1ω7c
	19:0cy
Common to bacteria	15:0
	16:0
	16:1ω9c
	17:0
	a19:0
	i19:0

Table S1.2 Primers used for multiplex 16S and *nosZ* amplicon sequencing

Gene	Primer	Sequencing adapter	Barcode	Primer	Reference
16S subunit	1392R-GS01	ccatctcatccctgcgtgtctccgactcag	CATCG	acgggcggtgtgtRc	1
	1392R-GS02	ccatctcatccctgcgtgtctccgactcag	CATAT	acgggcggtgtgtRc	1
	1392R-GS03	ccatctcatccctgcgtgtctccgactcag	CAGCT	acgggcggtgtgtRc	1
	1392R-GS04	ccatctcatccctgcgtgtctccgactcag	CAGTG	acgggcggtgtgtRc	1
	1392R-GS05	ccatctcatccctgcgtgtctccgactcag	ACTCT	acgggcggtgtgtRc	1
	1392R-GS06	ccatctcatccctgcgtgtctccgactcag	ACTAC	acgggcggtgtgtRc	1
	1392R-GS07	ccatctcatccctgcgtgtctccgactcag	ACGCG	acgggcggtgtgtRc	1
	1392R-GS08	ccatctcatccctgcgtgtctccgactcag	ACGAT	acgggcggtgtgtRc	1
	1392R-GS09	ccatctcatccctgcgtgtctccgactcag	ACATG	acgggcggtgtgtRc	1
	1392R-GS10	ccatctcatccctgcgtgtctccgactcag	ACAGC	acgggcggtgtgtRc	1
	1392R-GS11	ccatctcatccctgcgtgtctccgactcag	ATCTC	acgggcggtgtgtRc	1
	1392R-GS12	ccatctcatccctgcgtgtctccgactcag	ATCGT	acgggcggtgtgtRc	1
	1392R-GS13	ccatctcatccctgcgtgtctccgactcag	ATCAG	acgggcggtgtgtRc	1
	1392R-GS14	ccatctcatccctgcgtgtctccgactcag	ATGTG	acgggcggtgtgtRc	1
	1392R-GS15	ccatctcatccctgcgtgtctccgactcag	ATGAC	acgggcggtgtgtRc	1
		cctatcccctgtgtgccttggcagtctcag		aaactYaaaKgaattgacgg	1
nosZ	nosZR-01	ccatctcatccctgcgtgtctccgactcag	TCGCTAG	cgSaccttSttgccstYgcg	2
	nosZR-02	ccatctcatccctgcgtgtctccgactcag	TCGCTAG	cgSaccttSttgccstYgcg	2
	nosZR-03	ccatctcatccctgcgtgtctccgactcag	TCGCTAG	cgSaccttSttgccstYgcg	2
	nosZR-04	ccatctcatccctgcgtgtctccgactcag	TCGCTAG	cgSaccttSttgccstYgcg	2
	nosZR-05	ccatctcatccctgcgtgtctccgactcag	TCGCTAG	cgSaccttSttgccstYgcg	2
	nosZR-06	ccatctcatccctgcgtgtctccgactcag	TGCATAG	cgSaccttSttgccstYgcg	2
	nosZR-07	ccatctcatccctgcgtgtctccgactcag	TGTGTAC	cgSaccttSttgccstYgcg	2

Table S1.2 continued.

Gene	Primer	Sequencing adapter	Barcode	Primer	Reference
nosZ	nosZR-08	ccatctcatccctgcgtgtctccgactcag	TGACTGA	cgSaccttSttgccstYgcg	2
	nosZR-09	ccatctcatccctgcgtgtctccgactcag	TACTCGA	cgSaccttSttgccstYgcg	2
	nosZR-10	ccatctcatccctgcgtgtctccgactcag	TATCTCG	cgSaccttSttgccstYgcg	2
	nosZR-11	ccatctcatccctgcgtgtctccgactcag	TATACTG	cgSaccttSttgccstYgcg	2
	nosZR-12	ccatctcatccctgcgtgtctccgactcag	CTCTGAG	cgSaccttSttgccstYgcg	2
	nosZR-13	ccatctcatccctgcgtgtctccgactcag	CTCATGA	cgSaccttSttgccstYgcg	2
	nosZR-14	ccatctcatccctgcgtgtctccgactcag	CTGCGAT	cgSaccttSttgccstYgcg	2
	nosZR-15	ccatctcatccctgcgtgtctccgactcag	CTGTCGA	cgSaccttSttgccstYgcg	2
	nosZF	cctatcccctgtgtgccttggcagtctcag		cgYtgttcMtcgacagccag	3

1) de Oliveria et al., 2013; 2) Kloos et al., 2001; 3) Throback et al., 2004.

Table S1.3 Single-copy housekeeping COGs used to relativize within samples

COG	Function	Sequence length (bp)
COG0016	Phenylalanyl-tRNA synthetase alpha subunit	1005
COG0048	Ribosomal protein S12	387
COG0049	Ribosomal protein S7	444
COG0051	Ribosomal protein S10	312
COG0052	Ribosomal protein S2	756
COG0072	Phenylalanyl-tRNA synthetase beta subunit	1950
COG0080	Ribosomal protein L11	423
COG0081	Ribosomal protein L1	684
COG0087	Ribosomal protein L3	654
COG0088	Ribosomal protein L4	642
COG0090	Ribosomal protein L2	825
COG0091	Ribosomal protein L22	360
COG0092	Ribosomal protein S3	699
COG0093	Ribosomal protein L14	366
COG0094	Ribosomal protein L5	540
COG0096	Ribosomal protein S8	396
COG0097	Ribosomal protein L6P/L9E	534
COG0098	Ribosomal protein S5	543
COG0099	Ribosomal protein S13	363
COG0100	Ribosomal protein S11	387
COG0103	Ribosomal protein S9	390
COG0127	Xanthosine triphosphate pyrophosphatase	582
COG0149	Triosephosphate isomerase	753
COG0164	Ribonuclease HII	597
COG0184	Ribosomal protein S15P/S13E	267
COG0185	Ribosomal protein S19	279
COG0186	Ribosomal protein S17	261
COG0197	Ribosomal protein L16/L10E	438
COG0200	Ribosomal protein L15	456
COG0244	Ribosomal protein L10	525
COG0256	Ribosomal protein L18	375
COG0343	Queuine/archaeosine tRNA-ribosyltransferase	1116
COG0481	Membrane GTPase LepA	1809
COG0504	CTP synthase (UTP-ammonia lyase)	1599
COG0532	Translation initiation factor 2 (IF-2; GTPase)	1527
COG0533	Metal-dependent proteases with possible chaperone activity	1026
COG0541	Signal recognition particle GTPase	1353

Table S1.4 Sequence coverage and diversity statistics by sample

			16S amplicon sequencing						
System	**Year**	**Block**	**Reads**	**OTUs**	**Rare**	**Coverage**	**Shannon's**	**Chao1**	**SRA Accession**
Corn	2010	1	4359	653	139	0.928	5.03	1231.1	SRX969981
Corn	2011	1	3904	681	157	0.914	5.19	1257.8	SRX970055
Corn	2012	1	4602	733	198	0.925	5.18	1219.4	SRX970864
Corn	2010	2	6206	727	239	0.947	4.76	1262.4	SRX970019
Corn	2011	2	4016	692	208	0.917	5.12	1200.6	SRX970059
Corn	2012	2	3941	635	179	0.927	5.10	1059.7	SRX970866
Corn	2010	3	4652	601	170	0.942	4.56	1015.1	SRX970030
Corn	2011	3	4391	721	193	0.924	5.18	1196.3	SRX970061
Corn	2012	3	2852	598	114	0.903	5.24	935.7	SRX970868
Corn	2010	4	3370	545	93	0.924	4.86	892.2	SRX970034
Corn	2011	4	2741	532	83	0.909	5.14	889.8	SRX970064
Corn	2012	4	3051	532	85	0.921	5.12	845.6	SRX970870
Corn	2010	5	3417	490	82	0.934	4.70	773.8	SRX970038
Corn	2011	5	3013	572	112	0.914	5.12	912.1	SRX970066
Corn	2012	5	2383	495	93	0.898	4.97	853.6	SRX970872
Prairie	2010	1	4158	719	222	0.914	5.11	1274.7	SRX970874
Prairie	2011	1	4462	715	205	0.922	5.07	1211.1	SRX971228
Prairie	2012	1	4211	750	211	0.914	5.15	1226.1	SRX971275
Prairie	2010	2	4121	711	178	0.918	5.22	1175.1	SRX970878
Prairie	2011	2	3881	690	173	0.914	5.17	1162.5	SRX971233
Prairie	2012	2	3637	626	150	0.918	5.03	1077.6	SRX971282
Prairie	2010	3	4043	673	169	0.919	5.07	1178.9	SRX970881
Prairie	2011	3	2912	583	98	0.905	5.16	960.5	SRX971239
Prairie	2012	3	2574	564	77	0.894	5.26	975.0	SRX971291
Prairie	2010	4	2982	564	193	0.910	5.14	958.6	SRX971225
Prairie	2011	4	3094	581	104	0.908	5.08	1030.7	SRX971259
Prairie	2012	4	2777	510	71	0.911	5.08	917.2	SRX971296
Prairie	2010	5	3548	619	119	0.924	5.17	934.4	SRX970885
Prairie	2011	5	3090	592	110	0.904	5.19	1109.1	SRX971267
Prairie	2012	5	2714	517	77	0.908	4.97	871.9	SRX971300
Total			109102	3717	2554	0.985	5.40	6311.0	
Maximum			6206	750	239	0.947	5.26	1274.7	
Minimum			2383	490	71	0.894	4.56	773.8	

Table S1.4 cont.

			nosZ amplicon sequencing						
System	**Year**	**Block**	**Reads**	**OTUs**	**Rare**	**Coverage**	**Shannon's**	**Chao1**	**SRA Accession**
Corn	2010	1	4173	232	27	0.982	3.74	343.0	SRX970011
Corn	2011	1	5625	270	49	0.981	3.92	386.0	SRX970056
Corn	2012	1	4697	295	47	0.985	3.90	401.9	SRX970865
Corn	2010	2	4733	263	47	0.987	3.79	311.4	SRX970024
Corn	2011	2	4807	266	44	0.981	3.57	347.3	SRX970060
Corn	2012	2	3742	208	22	0.982	3.52	298.2	SRX970867
Corn	2010	3	5769	261	34	0.987	3.65	362.9	SRX970031
Corn	2011	3	1179	235	28	0.984	3.71	317.1	SRX970063
Corn	2012	3	1521	213	40	0.929	4.07	310.9	SRX970869
Corn	2010	4	6681	175	27	0.955	3.34	259.4	SRX970037
Corn	2011	4	4846	155	16	0.942	3.28	264.8	SRX970065
Corn	2012	4	1389	191	28	0.957	3.21	287.3	SRX970871
Corn	2010	5	4902	169	15	0.976	3.38	257.5	SRX970040
Corn	2011	5	5106	181	26	0.948	3.33	312.7	SRX970067
Corn	2012	5	1348	150	24	0.940	3.42	246.1	SRX970873
Prairie	2010	1	5938	296	60	0.986	3.89	406.7	SRX970875
Prairie	2011	1	1595	246	37	0.983	3.62	382.1	SRX971232
Prairie	2012	1	1551	265	43	0.985	3.56	356.4	SRX971277
Prairie	2010	2	4471	248	37	0.985	3.81	349.1	SRX970880
Prairie	2011	2	1641	170	19	0.942	3.69	237.0	SRX971237
Prairie	2012	2	1116	257	36	0.984	3.62	353.9	SRX971287
Prairie	2010	3	1252	222	29	0.934	3.93	359.5	SRX970884
Prairie	2011	3	1212	202	35	0.937	3.83	349.2	SRX971243
Prairie	2012	3	5469	199	26	0.939	3.84	282.0	SRX971293
Prairie	2010	4	1167	164	22	0.943	3.36	235.5	SRX971226
Prairie	2011	4	1834	189	28	0.931	3.44	328.9	SRX971262
Prairie	2012	4	1190	200	27	0.946	3.46	317.9	SRX971299
Prairie	2010	5	1283	262	54	0.985	3.56	356.9	SRX970886
Prairie	2011	5	2518	164	16	0.934	3.35	275.2	SRX971269
Prairie	2012	5	1503	177	20	0.957	3.33	236.4	SRX971301
			94258	1072	641	0.996	3.98	1697.0	
			6681	296	60	0.987	4.07	406.7	
			1116	150	15	0.929	3.21	235.5	

Table S1.4 cont

			Shotgun metagenomic sequencing				
System	**Year**	**Block**	**Size (Gbp)**	**Sequences (millions)**	**COG sequences (millions)**	**16S sequences**	**JGI Taxon Object ID**
Corn	2010	1	1.84	9.32	3.76	5285	3300002110
Corn	2011	1	1.86	9.23	3.66	5328	3300002109
Corn	2012	1	1.35	6.63	2.65	2151	3300001694
Corn	2010	2	2.02	10.09	4.04	7924	3300002117
Corn	2011	2	1.86	9.34	3.62	6220	3300002112
Corn	2012	2	1.34	6.54	2.58	2833	3300001693
Corn	2010	3	1.84	9.21	3.74	6352	3300002108
Corn	2011	3	1.47	7.30	2.84	2995	3300001698
Corn	2012	3	1.37	6.70	2.63	2718	3300001695
Corn	2010	4	1.86	9.32	3.72	6296	3300002111
Corn	2011	4	1.39	6.82	2.67	2798	3300001696
Corn	2012	4	1.48	7.27	2.87	3211	3300001697
Corn	2010	5	1.96	9.80	4.03	8135	3300002115
Corn	2011	5	1.56	7.72	3.01	2993	3300001703
Corn	2012	5	1.27	6.22	2.46	2426	3300001691
Prairie	2010	1	1.87	9.40	3.67	5561	3300002113
Prairie	2011	1	1.33	6.50	2.62	2213	3300001692
Prairie	2012	1	1.53	7.51	2.93	2970	3300001700
Prairie	2010	2	1.95	9.89	3.82	5569	3300002116
Prairie	2011	2	1.54	7.55	2.94	3021	3300001701
Prairie	2012	2	1.53	7.65	2.94	2831	3300001702
Prairie	2010	3	1.79	8.97	3.52	4187	3300002107
Prairie	2011	3	1.57	7.75	2.98	2977	3300001704
Prairie	2012	3	1.28	6.14	2.41	2427	3300001690
Prairie	2010	4	1.61	7.91	3.23	3110	3300002106
Prairie	2011	4	1.61	7.99	3.11	3136	3300001705
Prairie	2012	4	1.32	6.48	2.54	2420	3300001745
Prairie	2010	5	1.90	9.47	3.65	5701	3300002114
Prairie	2011	5	1.51	7.52	2.88	3050	3300001699
Prairie	2012	5	1.47	7.19	2.77	2802	3300001747
Total			48.28	239.43	94.29	119640	
Maximum			1.27	6.14	2.41	8135	
Minimum			2.02	10.09	4.04	2151	

Table S1.5 Soil physical and chemical properties, 0-10 cm

Year	System	Percent carbon (w/w)	Percent nitrogen (w/w)	Bulk density (g cm^{-3})
2009	Corn	2.35 ± 0.11 *n.s.*	0.221 ± 0.012 *n.s.*	1.38 ± 0.01 *c*
	Prairie	2.31 ± 0.10 *n.s.*	0.212 ± 0.011 *n.s.*	1.37 ± 0.02 *c*
2013	Corn	2.28 ± 0.11 *n.s.*	0.225 ± 0.011 *n.s.*	1.23 ± 0.03 *b*
	Prairie	2.38 ± 0.10 *n.s.*	0.228 ± 0.012 *n.s.*	1.14 ± 0.02 *a*

Unpublished data from GR Sanford and RD Jackson. Values are mean ± s.e. and derived from 0-10 cm soil samples. Unlabeled groups and groups sharing a letter are not significantly different ($P > 0.05$).

Table S1.6 Abundances of major bacterial phyla and classes

	Amplicon		**Shotgun**		***F* values**		
	Corn	Prairie	Corn	Prairie	Method	System	Interaction
Gram-positive	0.161	0.171	0.317	0.316	7465	0	21
Actinobacteria	0.107	0.113	0.280	0.248	9021	1	108
Firmicutes	0.054	0.057	**0.036**	**0.067**	36	2	241
Other Gram-positives	<0.001	<0.001	0.001	0.002			
Gram-negative	0.797	0.795	0.639	0.637	7262	0	2
Acidobacteria	0.192	0.215	0.124	0.137	2272	2	0
Bacteroidetes	0.061	0.063	**0.011**	**0.018**	2690	2	58
Proteobacteria	**0.300**	**0.273**	**0.313**	**0.293**	70	6	8
α-Proteobacteria	0.136	0.126	0.168	0.154	274	1	0
β-Proteobacteria	0.038	0.039	0.051	0.054	219	0	0
γ-Proteobacteria	0.077	0.060	0.055	0.045	370	2	2
Other Proteobacteria	0.049	0.058	0.055	0.063			
Verrucomicrobia	**0.042**	**0.052**	**0.044**	**0.055**	9	6	0
Other Gram-negatives	0.202	0.192	0.147	0.134			

Bolded values indicate statistically significant cropping system differences within a sequencing methodology ($P < 0.05$). Amplicon refers to amplicon sequencing of 16S rRNA; shotgun refers to shotgun metagenomic sequencing reads annotated as 16S rRNA. Groups comprising > 5% of reads for at least one system-method combination F values are drawn from ANOVA of a logistic general linear model with terms of sequencing method, cropping system, and their interaction.

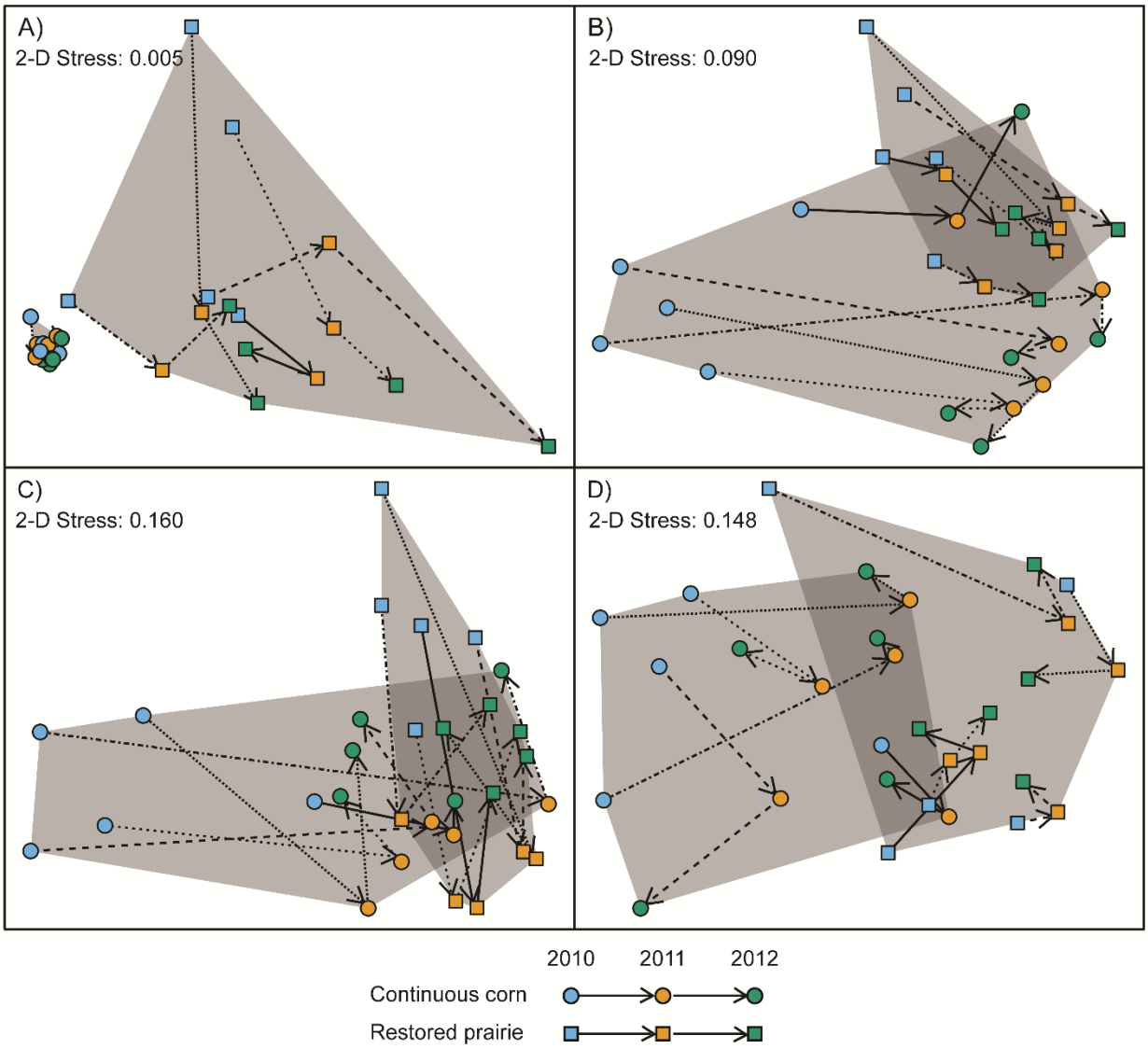

Figure 1.1 NMDS ordinations of soil microbial communities based on A) PLFA profiling (see Table S1.1 for list of PLFAs); B) shotgun sequenced functional gene abundances; C) 16S rRNA amplicon sequence OTUs; and D) *nosZ* amplicon clusters. Line patterns link individual plots sampled over successive years. Grey areas represent ordination hulls for the two cropping systems.

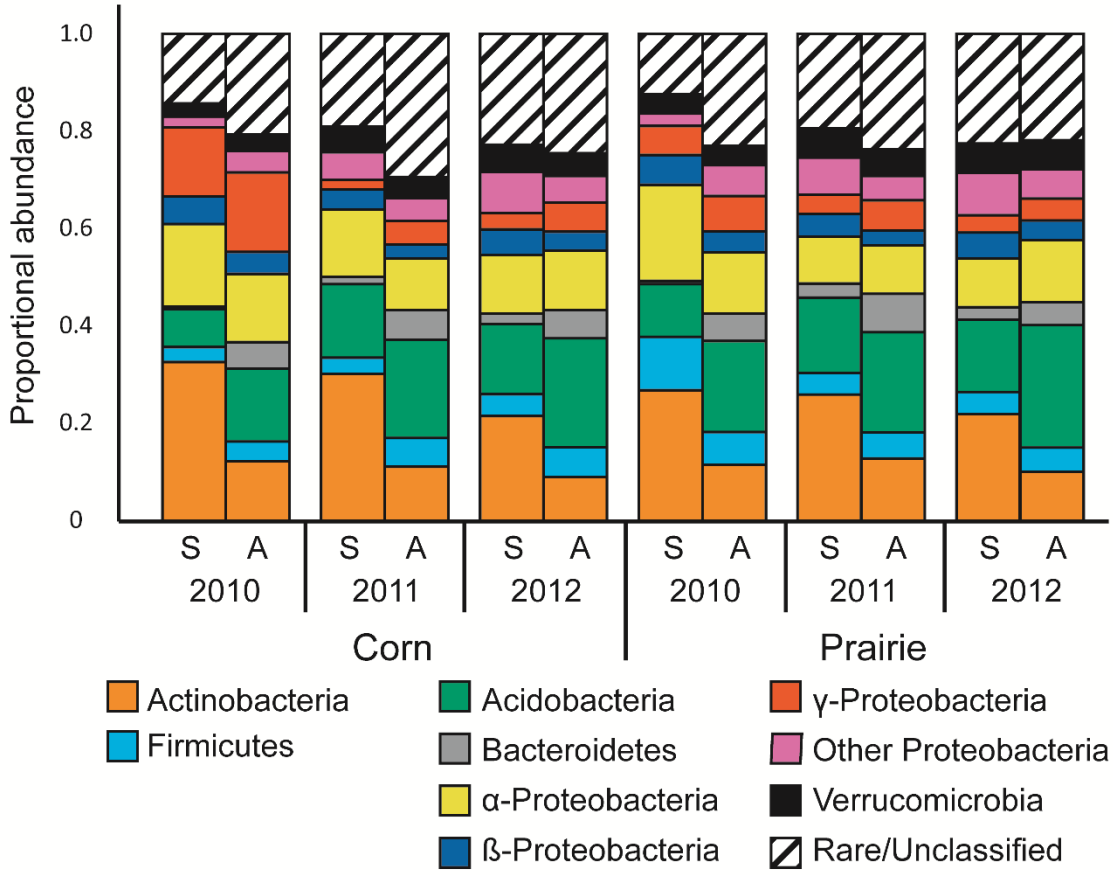

Figure 1.2 Proportion of reads assigned to select bacterial phyla and classes detected in 16S rRNA sequences from S) shotgun metagenomic and A) amplicon sequence data. Bars represent arithmetic means of five independent samples. Rare taxa are defined as contributing < 5% of the total reads for both sequencing types.

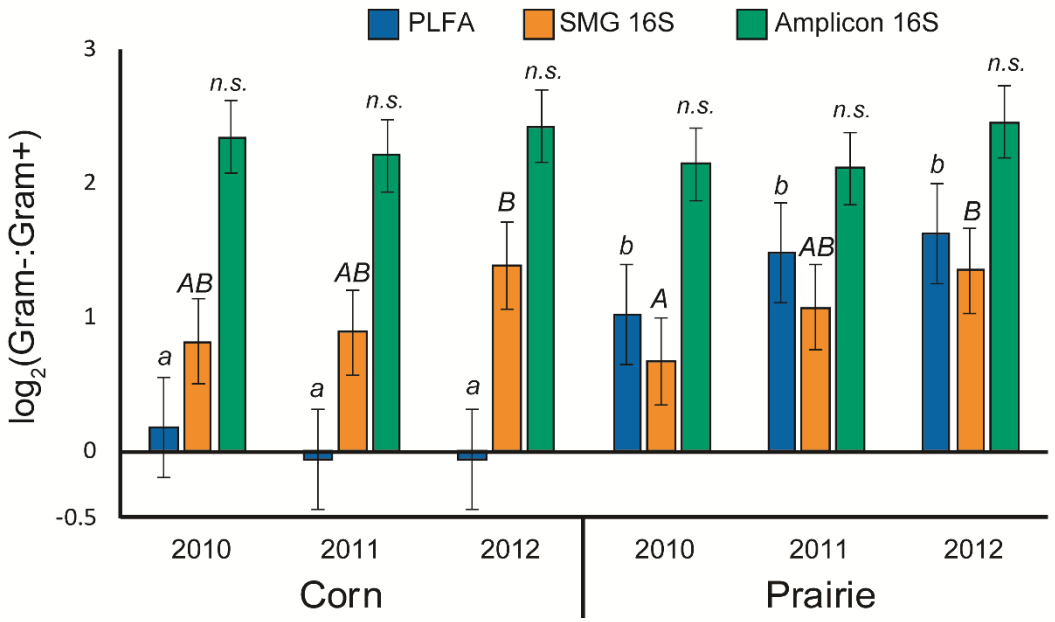

Figure 1.3 Ratios of biomarkers associated with Gram-negative and Gram-positive phyla. The PLFA ratio is based on indicator lipid biomass (see Table S1 for list of PLFAs); DNA-based ratios are derived from phylum-level assignments of 16S rRNA gene reads from either shotgun metagenomic (SMG) or amplicon sequencing (phylum details in Table S1.6). Values are arithmetic means (± 1 standard error) of $\log_2$-transformed ratios. Within a method, groups sharing a letter are not significantly different ($P > 0.05$).

CHAPTER 2

Correlating annual N_2O emissions to soil microbial taxonomic, functional gene, and membrane lipid biomarkers in two ecologically contrasting cropping systems using elastic net modeling

David S. Duncan*, Randall D. Jackson

Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA

DOE-Great Lakes Bioenergy Research Center, Department of Bacteriology, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

* Corresponding author. Email: dsduncan@wisc.edu

Abstract

Soil microbial communities may provide insights into the drivers of variability in nitrous oxide (N_2O) emissions from soils. We used elastic net modelling to explore the relationship between cumulative annual N_2O emissions ins and four microbial biomarker types (membrane lipids, 16S rRNA, nitrous oxide reductase (*nosZ*) genes, and functional genes from shotgun sequencing). We conducted this experiment on two ecologically distinct cropping systems, conventionally managed no-tillage corn, and nonfertilized, harvested assemblages of tallgrass prairie species, located on a research farm in south-central Wisconsin, USA. The elastic net modeling approach reduces the risk of overfitting when performing regression with large numbers of predictors through a combination of coefficient shrinkage and term removal and employs a mixing parameter (alpha) that regulates the weight given to both components. All biomarkers except for membrane lipids formed credible elastic net models with corn system N_2O emissions, while only 16S rRNA operational taxonomic units (OTUs) did so for the prairie system. Functional genes and 16S rRNA OTUs captured both interannual and within-year variability in corn system N_2O emissions, while *nosZ* OTUs only reflected within-year variability. Models employing 16S rRNA OTUs and functional genes, but not those using *nosZ* OTUs, responded to values of the alpha parameter. Individual biomarkers identified through this modelling approach are unlikely to reflect microbial causes of variability in N_2O emissions, but may instead indicate microbial taxa that are sensitive to environmental conditions relevant to N_2O production. With appropriate knowledge of organismal level physiology and ecology, this approach may provide insight into overlooked environmental conditions that drive variability in N_2O emissions.

2.1 Introduction

The production of nitrous oxide (N_2O) from soils is one of the key mechanisms by which the agricultural sector in the United States contributes to global climate change (U.S. Environmental Protection Agency, 2014). Multiple biotic and abiotic factors influence rates of N_2O production. Microbially-mediated processes, notably nitrification and denitrification, are the ultimate source of N_2O production in soils (Braker and Conrad, 2011). Abiotic factors such as temperature, oxygen availability, and substrate concentrations heavily influence the rate at which these processes occur, and thus are frequently linked to N_2O flux rates (Hénault et al., 2005). Microbial processes can also reduce availability of inorganic nitrogen substrates (Luxhøi et al., 2006) or consume N_2O (Cavigelli and Robertson, 2001), further complicating matters. As a result, predictive modeling of soil N_2O fluxes remains a major challenge (Hu et al., 2015; Roelandt et al., 2005).

Efforts to understand and model variability in N_2O fluxes have largely focused on abiotic factors (Hénault et al., 2005), as their relationship to N_2O production is clearer and they have historically been easier to measure than microbial communities. However, the key abiotic drivers of N_2O flux variability differ among environments (Dechow and Freibauer, 2011), limiting the scope inference that can be drawn from purely abiotic data. Incorporating soil microbial information should, in principle, lead to models that are more accurate and can be applied over a broader range of environments. Some studies link microbial communities to processes underlying N_2O production (Harter et al., 2014; Morales et al., 2010; Németh et al., 2014), but microbial data do not necessarily improve model performance (Graham et al., 2014). This is likely part of the broader issue with relating environmental factors and processes: relationships are clear and straightforward only when the factor of interest heavily constrains the process, while constraint by other factors obscures any relationship (Hiddink and Kaiser, 2005). Soil N_2O emissions may be primarily related to proximal environmental controls except in cases where the soil microbial community's functional capacity becomes limiting (Braker and Conrad, 2011; Wallenstein et al., 2006).

Although microbes are primarily conceptualized and interpreted as causal agents, they may also serve as sensitive indicators of environmental factors influencing N_2O production. Soil microbial

communities contain substantial genetic and functional diversity (Torsvik and Øvreås, 2002). Soil microbes experience the small-scale heterogeneity of soil, which may be reflected by community compositions with a nuance that would be lost to standard soil monitoring techniques. Microbiota could serve as sensitive, integrative biomarkers of environmental health (Ritz et al., 2009; Schloter et al., 2003). Ecological indicators can be useful for informing management even if they cannot be directly linked to underlying drivers or processes (Contamin and Ellison, 2009). Overall microbial community patterns, such as indices created from principal components analysis (PCA) axes, may relate to ecological processes in ways that the abundance of individual taxa or other markers would not (Balser and Firestone, 2005). We may be able to gain insights into soil N_2O dynamics by focusing on the organisms most directly involved in its production, but there is also much we can learn from organisms that are highly responsive to environmental factors influencing N_2O production.

In this study, we combined previously-published data on N_2O emissions (Oates et al., 2016) and soil microbial community compositions (Duncan et al., 2016) from a bioenergy feedstock cropping systems study. We focused on the two most ecologically dissimilar systems: a corn (*Zea mays* L.) monoculture under conventional no-till management and an annually-harvested assemblage of native prairie species receiving no agronomic inputs (Sanford et al., 2016). From 2009 to 2011, the corn system emitted considerably more N_2O on an annual basis, reflecting the greater nitrogen inputs it received and the reduced efficiency with which the plant and microbial components of the system immobilized and cycled nitrogen (Oates et al., 2016). Daily N_2O fluxes in the corn system could be reasonably well modeled from soil temperature, moisture, and concentrations of NO_3^- and NH_4^+, while fluxes from the prairie system showed no relationship to these environmental factors. Soil microbial community membrane lipid profiles, taxonomies of 16S rRNA and the nitrous oxide reductase (*nosZ*) gene, and functional gene profiles differed starkly between the two systems (Duncan et al., 2016). Our objective with the present study was to explore whether differences in individual microbial biomarkers correlated to cumulative annual N_2O fluxes.

2.2 Materials and methods

2.2.1 Site description and sampling

This experiment was conducted on the Bioenergy Cropping Systems Experiment (BCSE) at the University of Wisconsin-Madison Arlington Agricultural Research Station (43° 17' 45" N, 89° 22' 48" W, 315 m a.s.l.) described in detail by Sanford et al. (2016). The site is on highly productive Plano silt-loam soils (Fine-silty, Mixed, Superactive, Mesic Typic Argiudolls) and from 1981 to 2010 experienced mean annual precipitation of 869 mm and mean annual air temperature minima and maxima of -14.6 and 27.6°C. The BCSE consisted of eight potential bioenergy feedstock cropping systems grown in 27 × 43-m (0.12 ha) plots, arranged in a randomized complete block design with five replicates. The corn system received no-tillage management, with nutrient (NPK) applications based on University of Wisconsin Extension nutrient recommendations (Laboski et al., 2012) and annual soil tests. The prairie system was an assemblage of species indigenous to the North American tallgrass prairie (listed in Oates et al., 2016) receiving no nutrient inputs or agronomic management aside from an annual post-senescence harvest. Prior to establishment of the BCSE in 2008, the site had been in agricultural production, primarily corn, soybean (*Glycine max* L.) and alfalfa (*Medicago sativa* L.). The blocking structure accounted for differences in prior land use (Sanford et al., 2016).

Soil sampling for microbial community characterization was conducted in mid-August from 2010 to 2012. Soils were sampled to 15 cm (3.7 cm diameter) matching the depth of annual agronomic soil tests. All five replicates were sampled, with five cores taken in a staggered transect from each plot, and homogenized to produce a single composite sample per plot. We sampled soils for microbial community characterization in mid-August from 2010 to 2012. We sampled to a depth of 15 cm to match the sampling depth for routine agronomic measurements taken at the site. For each plot, we took 5 cores in a staggered transect and homogenized them to produce a single composite sample per plot.

2.2.2 Calculating N_2O emissions

We reported full details on N_2O emissions measurement and calculation in Oates et al. (2016). Briefly, N_2O fluxes were measured using static chambers with an effective headspace volume of ~ 10 L (17 cm

height) and an insertion depth of 5 cm. Measurements were taken twice monthly throughout the year, except when soil temperatures were < 0°C, with additional sampling following fertilization and heavy rainfall events. Daily flux measurements were aggregated by linear interpolation to estimate calendar-year N_2O emissions. Emissions were log-transformed prior to analysis to mitigate over-dispersion.

2.2.3 Microbial community characterization

We reported full details for biomarker extraction, sequence analysis, and repository locations for raw data in Duncan et al. (2016). For membrane lipid characterization, we employed a combined phospholipid fatty acid and fatty acid methyl ester extraction method (Balser and Firestone, 2005; Kao-Kniffin and Balser, 2007) with fatty acid identification using MIDI's EUKARY method database. Fatty acids with carbon chain lengths ≥20 were considered plant-derived and excluded from analysis.

Environmental DNA was extracted from 1 g soil samples using SDS, phenol, and bead beating for cell lysis, phenol-chloroform extraction, and precipitation in sodium acetate-ispopropanol (Stevenson and Weimer, 2007), with subsequent use of a Power Soil Cleanup Kit (Mo Bio Laboratories, Carlsbad, CA) following manufacturer instructions with the optional ethanol rinse.

Amplicon sequencing of 16S rRNA and *nosZ* employed a Roche 454 GS Junior Titanium sequencer. We used primers described by de Oliveira (2013) to target the V6-V8 region of the 16S subunit and the nosZF (Kloos et al., 2001) and nosZ1622R (Throbäck et al., 2004) primers to target the catalytic subunit of *nosZ*. Processing of 16S rRNA reads was done with mothur v 1.33.3 (Schloss et al., 2009) with a 200-bp length cutoff, alignment against the SILVA reference database (Pruesse et al., 2007) and OTU definition at 95% sequence identity. Following quality filtration in mothur, *nosZ* amplicons were analyzed using the Functional Gene Pipeline (Fish et al., 2013): translation with Framebot (Wang et al., 2013), alignment with Aligner (Cole et al., 2014), and clustering with mcClust (Fish et al., 2013) with OTU definition at 90% amino acid sequence similarity.

The Department of Energy Joint Genome Institute (JGI) conducted shotgun sequencing of environmental DNA using the Illumina HiSeq Platform with paired-end 150-bp reads and multiplexing 12 samples per lane. Assembly and annotation were carried out through the standard JGI pipeline, using the

updated clusters of orthologous groups (COGs) database for functional gene annotation (Galperin et al., 2015; Tatusov et al., 2003).

2.2.4 Statistical analyses

Analyses were conducted in the R statistical environment (v. 3.3.0, R Core Team, 2016). N_2O emissions were analyzed with the 'lmer' function in the lme4 package (v1.1-12, Bates et al., 2015), with block as a random effect. Principal components analysis (PCA) was calculated with the 'rda' function in the vegan package (v2.3-5 Oksanen et al., 2013) using raw values for lipid and functional gene data, and square root transformations for amplicon abundances (following Duncan et al., 2016).

We used regularization to correlate log-transformed N_2O fluxes to individual microbial community biomarkers. This approach imposes a penalty during model fitting to achieve a sparse solution from a large number of potential predictor terms, attempting to optimize the tradeoff between predictive power and model bias (Zou and Hastie, 2005). Two forms of regularization are ridge regression, which retains all terms but greatly reduces coefficient magnitude, and the lasso, which simultaneously drops terms while selectively reducing their coefficients. We used elastic net modeling, which hybridizes the two methods by using a mixing parameter ('alpha' in glmnet) to weight the combined penalty term toward either ridge regression or the lasso, as implemented by the glmnet package (v2.0-5, Friedman et al., 2010). To identify the appropriate weight for the combined penalty term ('lambda' in glmnet), we used leave-one-out cross validation with the 'cv.glmnet' function. This function calculates cross-validated error across the full range of lambda values; we selected the most stringent penalty term within one standard error of the term that gave the smallest error ('lambda.1se' in glmnet) for evaluating model terms and fit. Model strength was evaluated using the ratio of explained to null deviance ('dev.ratio' in glmnet). To evaluate the strength of these deviance ratios, we generated distributions of deviance ratios from permuted datasets, permuting flux values within year and running the data through the procedure outlined above.

We tested for correlations using individual microbial biomarkers (microbial lipid biomass, 16S rRNA and *nosZ* amplicon OTU abundances, and COG-annotated shotgun sequencing read abundances).

We tested the effect of variable selection stringency by varying the alpha parameter from 1.0 (lasso method) to 0.5 (equal weighting), selecting the lower bound based on Helbling et al. (2015) and a need to avoid excessive variable retention. To remove rare biomarkers that formed binary (presence-absence) rather than abundance gradients, we excluded any marker that was not present in at least one replicate in all study years.

2.3 Results

2.3.1 Sources of variation in N_2O emissions

Aggregate annual N_2O emissions were substantially higher from the corn system than the prairie system (Fig. 2.1). Annual mean N_2O emissions differed only between 2010 and 2011 in corn, while there were no statistically significant interannual differences in the prairie system (Fig. 2.1).

Emissions within cropping systems differed substantially among years, although these differences were not statistically significant in the prairie system, with rank order for years differing between systems (Fig. 2.1). We tested how elastic net modeling represented this interannual variability with models that had year as the only factor. Across all alpha levels, the model of the corn system retained a term distinguishing 2011 from 2010, with a deviance ratio of ~0.3 (i.e. explaining approximately 30% of deviance in aggregate emissions). In the prairie, by contrast, the year variable was dropped, leaving only a null model. Consecutive-year fluxes from individual plots were weakly correlated in the corn system (Pearson r = 0.21 for 2010-11 and 0.47 for 2011-12), but more strongly correlated in the prairie (r = 0.69 for 2010-11 and -0.63 for 2011-12).

System-level differences in N_2O emissions were matched by system-level differences detected by all four microbial biomarker types (Duncan et al., 2016). This risked effectively creating a two-point regression (e.g. Morales et al., 2010), providing little information beyond system-level differences. To avoid this, we analyzed each system separately, thus focusing on biomarker correlations to within-system variability. We analyzed years together, as interannual differences were less consistent than those between systems. However, we included terms for years as potential model predictors.

2.3.2 Correlation of N_2O emissions to microbial biomarkers

In the corn system, N_2O emissions could be correlated to all four microbial community biomarkers used, while in the prairie system only 16S rRNA OTUs produced viable models (Fig. 2.2). We explore individual marker types in detail below. Overall, DNA-based biomarkers led to models with greater explanatory power. The alpha parameter, which modulated coefficient shrinkage and removal, did not affect all measures equally, as is described below. Across alpha levels, models fell within the 85^{th}-95^{th} percentiles of deviance ratios achieved with permuted data. Thus, the explanatory power of these models was within the range of what was observed when microbial and aggregate emissions data were paired within year, albeit at the upper end of that range.

2.3.2.1 Membrane lipids

Microbial lipids generated the weakest models, producing deviance ratios that were lower than those obtained using year as the sole factor (see Section 3.1) and that fell well within the range of values obtained from permuted data (Fig. 2.2). These models retained a term for year, as well as two lipids, *cis*16:1ω7 and 16:1 2OH, which were included for alpha values up to 0.8, above which only 16:1 2OH was retained (Table S2.1).

2.3.2.2 16S rRNA amplicon OTUs

The strongest model for the corn system, as well as the only non-null models for the prairie system, emerged from 16S rRNA amplicon OTU data (Fig. 2.2). Both systems dropped the term for year, but there was no overlap in the OTUs retained for each system (Table 2.1).

Models of the prairie system retained 2 to 6 OTUs, with the full set of OTUs representing 1.6% (s.d. ± 0.4%) of all reads (Table 2.1). Dropping model terms in more lasso-like models did not adversely impact model performance. All OTUs were bacterial, with two *Acidobacteria*, one each of *Actinobacteria*, *Chloroflexi*, and *Gemmatimonadetes*, and one unclassified bacterium. Relative abundance of the *Gemmatimonadetes* OTU differed significantly between cropping systems, while three OTUs differed in abundance among years (Table 2.1). Interannual differences strongly influenced multivariate distributions of these OTUs across both systems.

Models of the corn system retained 13 unique OTUs, with 7 to 12 OTUs retained in any one model (Table 2.1), representing 2.8% (s.d. ± 1.1%) of all reads. In contrast to the prairie system, alpha levels influenced model performance, with more lasso-like models performing better up to alpha = 0.9 but with a sharp drop in performance for the fully lasso model (Fig. 2.2). Retained OTUs included three representatives from *Planctomycetes* and *Proteobacteria*, one from *Armatimonadetes*, *Chloroflexi*, and *Gemmatimonadetes*, three unclassified bacteria, and the archaeon *Nitrososphaera gargensis* (Table S2.2). Relative abundances of eight OTUs differed among years, with six showing consistent interannual differences for both cropping systems, while four OTUs were differently abundant between cropping systems. Cropping system and year separately and significantly influenced overall abundance patterns for retained OTUs retained in the corn system.

2.3.2.3 nosZ amplicon OTUs

In the corn system, amplicons of the nitrous oxide reductase gene *nosZ* resulted in models nearly as effective as those from 16S rRNA (Fig. 2.2). Alpha levels had almost no effect on these models, with only one of eight retained OTUs dropping out (Table 2.2). In contrast to the 16S rRNA models, *nosZ* amplicon models retained the term for year. Non-null models could still be constructed without the year term, but deviance ratios dropped to ~0.3 at all alpha levels. The cumulative abundance of retained OTUs varied substantially among samples, from 1.2 to 8.0% of all *nosZ* amplicon reads. All but two OTUs were negatively correlated to N_2O emissions; the two exceptions were present in the fewest samples and had the lowest relative abundance. Six OTUs most closely resembled uncultured bacteria sequenced from soils. The closest cultured homologs for five OTUs were from the family *Rhizobiales*, with another two OTUs from other *Alphaproteobacteria*. The remaining OTU matched a *Betaproteobacteria* of the *Achromobacter* genus and was one of the two that were positively correlated to N_2O emissions (Table S2.3). Relative abundances of two OTUs differed significantly among cropping systems and none differed significantly among years, but overall abundance patterns were influenced by both cropping system and year.

2.3.2.4 Functional gene abundances

The alpha parameter heavily influenced models built from functional gene profiles (Fig. 2.2). The number of retained COGs dropped from 23 at alpha = 0.5 to 6 at alpha = 1.0, with deviance ratios decreasing as terms were dropped (Table 2.3). Increasing the alpha parameter also increased the proportion of permuted datasets exceeding the deviance ratio of the real model, indicating the greater model explanatory power at low alpha values was not simply an artifact of retaining more terms in the model. Retained COGs covered a broad range of functions, with 13 of the 25 COG functional categories represented (Table S2.4). The only parameter with obvious links to N_2O production was COG3256, annotated as the nitric oxide reductase large subunit (*norB*). This COG was retained only at the lowest alpha level and with a very small coefficient. Curiously, average copy number of this COG by itself correlated strongly to N_2O emissions in the corn system (Fig. S2.1). Average copies per cell differed among years for 15 of the 25 retained COGs; interannual effects for 11 of these were consistent among years (Table 2.3). Eleven COGs differed in average copy number between cropping systems. Overall patterns for retained COGs were extremely different among systems and years, with only 34% of multivariate variability occurring within these groups.

2.4 Discussion

We found that DNA-based measures of soil microbial community composition could explain a large proportion of the variability in aggregate annual N_2O fluxes in a continuous corn cropping system, but not in a restored prairie. We characterized microbial communities using multiple biomarker types to explore how different aspects of the microbial community mapped onto variability in soil N_2O emissions. The four biomarker types resulted in distinct dynamics, most notably in how the balance between ridge regression and lasso penalization terms influenced model performance (discussed in Section 4.2). A far more important dynamic, however, was the substantial difference in the extent to which the biomarkers we used were able to model the corn and prairie cropping systems (discussed in Section 4.1).

Before exploring our results at greater depth, we must address the interpretation of individual biomarkers retained through the modeling process. We sampled microbial communities in mid-August,

while the major N_2O flux events that drove interannual variability occurred earlier in the season (Oates et al., 2016). We do not know the extent of microbial community turnover that occurred over that time, but it could potentially be very high (Schmidt et al., 2007). If turnover was high, then the community we sampled might not be indicative of the community present during periods of high N_2O flux and thus the markers identified would not indicate taxa that drove the N_2O production patterns we observed. In this scenario, the markers identified through elastic net modeling might instead reflect organisms that were particularly responsive to environmental conditions. Further investigation would be needed to determine whether soil microbes cause variability in N_2O emissions or simply reflect the conditions that drive it. If microbes function as sensitive integrators of N_2O-relevant environmental conditions, they could facilitate investigation of environmental processes that contribute to variability in N_2O production.

2.4.1 Microbial-N_2O emission correlations differ fundamentally between corn and prairie

The most consistent trend in our study was the complete dissimilarity between the corn and prairie systems, particularly the difficulty with generating meaningful models for the prairie. This dynamic follows prior observations that key abiotic drivers of N_2O flux differ among agroecosystems (Dechow and Freibauer, 2011; Oates et al., 2016). Fundamental differences in nitrogen cycling and availability likely drove this difference in N_2O dynamics. The corn system received large pulses of inorganic nitrogen which, if combined with high soil moisture and reasonably high temperature, could result in high-intensity, low-duration flux events (Molodovskaya et al., 2012) during which N_2O production might be limited by the soil microbial community's collective metabolism. In contrast, the prairie system received no exogenous nitrogen, so all inorganic nitrogen availability depended on mineralization of organic nitrogen governed by complex soil-plant-microbe interactions (Gliessman, 2007). The focus of N_2O emissions as responses may complicate the matter. More fundamental ecosystem properties such as potential denitrification (Yin et al., 2014) or N_2:N_2O ratios (Domeignoz-Horta et al., 2015) may respond more directly to microbial activity, exploring whether cropping systems differences extend beyond the level of nitrogen inputs.

Despite the limitations in our approach, our findings suggest interesting and potentially fundamental differences in the nitrogen cycling processes underlying N_2O emissions variability in the two systems. Both systems produced workable models using 16S rRNA OTUs, allowing us to infer coarse patterns in nitrogen cycling from the taxonomic identity of retained OTUs. Variability in N_2O emissions in the prairie system appeared heavily linked to growth conditions benefitting certain denitrifying taxa, as all retained OTUs with taxonomic information were potential denitrifiers. Contrast this to the corn system, which had its share of denitrifiers but also retained OTUs from organisms potentially capable of nitrification and dissimilatory nitrate reduction to ammonia (DNRA). This suggests a more complex ecology of nitrogen cycling in the corn system, at least inasmuch as related to N_2O production. This possibility runs counter to the dominant view of agricultural management, and exogenous nutrient addition in particular, as a force that reduces ecological complexity in agroecosystems (Cassman et al., 2002; Gliessman, 2007), and illustrates how microbial indicators may inform more nuanced views of ecosystem function.

2.4.2 Interpreting individual biomarker predictors

The three biomarker types that generated reasonable models for the corn system differed in their retention of an explicit term for interannual differences and in their response to the mixing parameter alpha. In both cases, the 16S rRNA OTUs and COGs behaved differently from *nosZ* OTUs. This is perhaps unsurprising, as prior studies show correlations at the community level between compositional similarity and functional gene abundances patterns (Duncan et al., 2016; Fierer et al., 2012). Moreover, the *nosZ* amplicons captured a much narrower subset of the microbial community than the other two biomarker types, ignoring denitrifiers lacking the *nosZ* gene (Philippot et al., 2011) as well as those possessing atypical forms of the gene (Orellana et al., 2014). This did not, however, prevent generation of reasonably good models from *nosZ* OTU data. In contrast, lipid membrane profiles failed to generate credible correlations. This was surprising, given the greater responsiveness of lipid data to environmental influences (Duncan et al., 2016; Liang et al., 2016). The limited phylogenetic specificity of this biomarker

likely restricted it to reflecting broad impacts on the microbial community, while the environmental drivers that influenced N_2O production appear to be visible with more narrowly-defined groups.

Both the 16S rRNA OTUs and COGs captured interannual variability without needing to use year as a discrete term. These biomarker types exhibited greater overall interannual variability in this system (Duncan et al., 2016), suggesting taxonomic composition and functional gene profiles respond to interannual variability. Many, but not all, markers retained in the models differed significantly among years. However, partitioning interannual effects among multiple terms proved more efficient than representing them with a single term, strongly suggesting these factors were not simply proxies for interannual variability. This contrasts sharply with the *nosZ* OTU model, which needed to retain an explicit term for interannual variability and for which retained terms did not differ in abundance among years. The retained *nosZ* OTUs thus reflected within-year variability in corn N_2O emissions, rather than interannual effects.

Biomarker types also differed in how their model performance responded to alpha values, which determined how strongly the elastic net penalization terms resembled those from either ridge regression or the lasso. A key dynamic to keep in mind is the lasso's tendency to retain single exemplars from groups of correlated terms while discarding the rest, in contrast to the greater inclusion of terms with more ridge regression-like penalization (Zou and Hastie, 2005). Terms retained at high alpha values represented largely independent gradients of community composition, and may thus have reflected similarly independent environmental factors. This further complicates interpretation of individual terms, as it is unclear whether they reflect a unique dynamic or are simply the best example of a constellation of similarly-behaving terms. Changes in term retention and their effect on model performance across multiple alpha values may help with interpretation. With *nosZ* OTUs, alpha levels barely influenced model structure or performance. The COGs were completely different, as increasing alpha values sharply decreased both the number of terms and model performance. The 16S rRNA OTUs provide perhaps the most interesting case. Removing terms improved model performance until the last alpha value, where impactful terms were removed and model performance dropped sharply. Overall, this approach suggests

that environmental gradients driving variability in corn system N_2O production can be approximated by abundances f relatively few N_2O reducers or bacterial taxa. Encompassing these gradients with COG abundances appears to require a larger number of weakly-related terms. This may reflect a lack of specificity in COG data, as multiple taxa with distinct ecological niches may all contribute to the abundance of a specific COG.

This study serves primarily to determine that some linkage exists between microbial community composition and variability in N_2O production. Exploring that linkage in detail requires detailed knowledge of the physiology and ecology of the taxa underlying that linkage. Generating this information lies beyond the scope of this study, but it is possible to illustrate how this approach might work using taxa from the 16S rRNA models. One of the OTUs from the corn system shared strong homology with *Nitrososphaera gargensis*, an ammonia oxidizing *Archaea* (AOA) with some flexibility in its ammonia sources (Spang et al., 2012). As a group, AOA are frequently less responsive than their bacterial counterparts to cropping system and fertilizer effects, but function at lower ammonia concentrations (Carey et al., 2016). A more ecologically-responsive AOA might serve to track nitrification-relevant ammonia or pH values. The prairie retained OTUs for *Acidobacteria* and *Actinobacteria*, which are frequently capable of N_2O production but not reduction (Ahn et al., 2014; Palmer and Horn, 2012; Shoun et al., 1998; Ward et al., 2009). One of the *Acidobacteria* identified in the prairie (OTU00021, identified as iii1-8) increases in abundance with increasing soil pH (Kim et al., 2014). Soil pH exerts strong, if frequently overlooked, controls over denitrification (Liu et al., 2013; Russenes et al., 2016) and such pH-sensitive organisms may reflect biologically-relevant trends in that regard. Far more organismal knowledge is clearly needed to fully interpret patterns like those we observed, but this approach may be useful to generate hypotheses and identify organisms for further, focused study.

2.5 Conclusions

We used elastic net modeling to correlate aggregate annual soil N_2O fluxes from two ecologically distinct bioenergy cropping systems to multiple biomarkers commonly used to characterize microbial communities. Strong relationships were observed for DNA-based biomarkers in the corn system, while

only a weak relationship with 16S rRNA OTUs was observed in the prairie system, suggesting both that biomarker types reflected different functional aspects of microbial community composition and that the linkage between microbial communities and N_2O emissions differed between the two cropping systems. The extent to which elastic net models employed lasso-like predilections for eliminating terms influenced the performance of models using 16S rRNA OTUs and COG-annotated functional genes, but barely impacted performance of models using *nosZ* amplicons. The biomarkers identified through this approach appear to be best interpreted as indirect indicators of environmental conditions driving variability in N_2O production, rather than as signs of microbial agents directly causing that variability. While this approach cannot directly link microbial community composition to N_2O production in soils, it may provide insights into environmental controls of the process which might not be detectable through other methods.

Acknowledgements

We thank KA Jewell, M Miller, A Neumann, HW Read, S Slater, D Stevenson, G Suen, JO Sustachek, and DA Williams for their assistance in generating microbial community characterization data. Early access to 2012 N_2O data was provided by LG Oates who, alongside S Hamilton, A Dean, J Tesmer, J Sustachek, N Tautges, A Miller, Z Andersen, B Faust, K Kahmark, S VanderWulp, and many others, was instrumental to generating the flux dataset. DSD was funded in part by a DC Smith Agronomy Department Fellowship and a Thomsen Wisconsin Graduate Fellowship. Funding was provided by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830). The sequencing work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Ahn, J.-H., Choi, M.-Y., Kim, B.-Y., Lee, J.-S., Song, J., Kim, G.-Y., Weon, H.-Y., 2014. Effects of water-saving irrigation on emissions of greenhouse gases and prokaryotic communities in rice paddy

soil. Microbial Ecology 68, 271–283.

Balser, T.C., Firestone, M.K., 2005. Linking microbial community composition and soil processes in a California annual grassland and mixed-conifer forest. Biogeochemistry 73, 395–415.

Bates, D., Maechler, M., Bolker, B.M., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1 – 48.

Braker, G., Conrad, R., 2011. Diversity, structure, and size of N_2O-producing microbial communities in soils-what matters for their functioning? Advances in Applied Microbiology 75, 33–70.

Carey, C.J., Dove, N.C., Beman, J.M., Hart, S.C., Aronson, E.L., 2016. Meta-analysis reveals ammonia-oxidizing bacteria respond more strongly to nitrogen addition than ammonia-oxidizing archaea. Soil Biology and Biochemistry 99, 158–166.

Cassman, K.G., Dobermann, A., Walters, D.T., 2002. Agroecosystems, nitrogen-use efficiency, and nitrogen management. AMBIO: A Journal of the Human Environment 31, 132–140.

Cavigelli, M.A., Robertson, G.P., 2001. Role of denitrifier diversity in rates of nitrous oxide consumption in a terrestrial ecosystem. Soil Biology and Biochemistry 33, 297–310.

Cole, J.R., Wang, Q., Fish, J.A., et al., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research 42, D633–D642.

Contamin, R., Ellison, A.M., 2009. Indicators of regime shifts in ecological systems: What do we need to know and when do we need to know it. Ecological Applications 19, 799–816.

de Oliveira, M.N.V., Jewell, K.A., Freitas, F.S., et al., 2013. Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. Veterinary Microbiology 164, 307–314.

Dechow, R., Freibauer, A., 2011. Assessment of German nitrous oxide emissions using empirical modelling approaches. Nutrient Cycling in Agroecosystems 91, 235–254.

Domeignoz-Horta, L.A., Spor, A., Bru, D., Breuil, M.-C., Bizouard, F., Léonard, J., Philippot, L., 2015. The diversity of the N_2O reducers matters for the N_2O:N_2 denitrification end-product ratio across an annual and a perennial cropping system. Frontiers in Microbiology 6, 971.

Duncan, D.S., Jewell, K.A., Suen, G., Jackson, R.D., 2016. Detection of short-term cropping system-

induced changes to soil bacterial communities differs among four molecular characterization methods. Soil Biology and Biochemistry 96, 160–168.

Fierer, N., Leff, J.W., Adams, B.J., et al., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences of the United States of America 109, 21390–21395.

Fish, J.A., Chai, B., Wang, Q., Sun, Y., Brown, C.T., Tiedje, J.M., Cole, J.R., 2013. FunGene: the functional gene pipeline and repository. Frontiers in Microbiology 4.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models viaa coordinate descent. Journal of Statistical Software 33, 1–22.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E. V, 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research 43, D261–D269.

Gliessman, S.R., 2007. The agroecosystem concept, in: Agroecology: The Ecology of Sustainable Food Systems. CRC Press, Boca Raton, pp. 23 – 32.

Graham, E.B., Wieder, W.R., Leff, J.W., et al., 2014. Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. Soil Biology & Biochemistry 68, 279–282.

Harter, J., Krause, H.-M., Schuettler, S., et al., 2014. Linking N_2O emissions from biochar-amended soil to the structure and function of the N-cycling microbial community. The ISME Journal 8, 660–674.

Helbling, D.E., Johnson, D.R., Lee, T.K., Scheidegger, A., Fenner, K., 2015. A framework for establishing predictive relationships between specific bacterial 16S rRNA sequence abundances and biotransformation rates. Water Research 70, 471–484.

Hénault, C., Bizouard, F., Laville, P., Gabrielle, B., Nicoullaud, B., Germon, J.C., Cellier, P., 2005. Predicting in situ soil N_2O emission using NOE algorithm and soil database. Global Change Biology 11, 115–127.

Hiddink, J.G., Kaiser, M.J., 2005. Implications of Liebig's law of the minimum for the use of ecological

indicators based on abundance. Ecography 28, 264–271.

Hu, H.-W., Chen, D., He, J.-Z., 2015. Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates. FEMS Microbiology Reviews 39, 729–749.

Kao-Kniffin, J., Balser, T.C., 2007. Elevated CO2 differentially alters belowground plant and soil microbial community structure in reed canary grass-invaded experimental wetlands. Soil Biology and Biochemistry 39, 517–525.

Kim, H.M., Jung, J.Y., Yergeau, E., et al., 2014. Bacterial community structure and soil properties of a subarctic tundra soil in Council, Alaska. FEMS Microbiology Ecology 89, 465–475.

Kloos, K., Mergel, A., Rösch, C., Bothe, H., 2001. Denitrification within the genus *Azospirillum* and other associative bacteria. Functional Plant Biology 28, 991–998.

Laboski, C.A.M., Peters, J.B., Bundy, L.G., 2012. Nutrient application guidelines for field, vegetable, and fruit crops in Wisconsin, UW-Extension Publication A2809. UW-Madison Cooperation Extension Publications, Madison, WI.

Liang, C., Jesus, E. da C., Duncan, D.S., Quensen, J.F., Jackson, R.D., Balser, T.C., Tiedje, J.M., 2016. Switchgrass rhizospheres stimulate microbial biomass but deplete microbial necromass in agricultural soils of the upper Midwest, USA. Soil Biology & Biochemistry 94, 173–180.

Liu, X., Chen, C.R., Wang, W.J., Hughes, J.M., Lewis, T., Hou, E.Q., Shen, J., 2013. Soil environmental factors rather than denitrification gene abundance control N_2O fluxes in a wet sclerophyll forest with different burning frequency. Soil Biology and Biochemistry 57, 292–300.

Luxhøi, J., Bruun, S., Stenberg, B., Breland, T.A., Jensen, L.S., 2006. Prediction of gross and net nitrogen mineralization-immobilization-turnover from respiration. Soil Science Society of America Journal 70, 1121–1128.

Molodovskaya, M., Singurindy, O., Richards, B.K., Warland, J., Johnson, M.S., Steenhuis, T.S., 2012. Temporal variability of nitrous oxide from fertilized croplands: Hot moment analysis. Soil Science Society of America Journal 76, 1728–1740.

Morales, S.E., Cosart, T., Holben, W.E., 2010. Bacterial gene abundances as indicators of greenhouse gas emission in soils. The ISME Journal 4, 799–808.

Németh, D.D., Wagner-Riddle, C., Dunfield, K.E., 2014. Abundance and gene expression in nitrifier and denitrifier communities associated with a field scale spring thaw N_2O flux event. Soil Biology and Biochemistry 73, 1–9.

Oates, L.G., Duncan, D.S., Gelfand, I., Millar, N., Robertson, G.P., Jackson, R.D., 2016. Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. GCB Bioenergy 8, 539–549.

Oksanen, J., Blanchet, F.G., Kindt, R., et al., 2013. vegan: Community ecology package [WWW Document]. R Package Version 2.0-10. URL http://cran.r-project.org/package=vegan

Orellana, L.H., Rodriguez-R, L.M., Higgins, S., et al., 2014. Detecting nitrous oxide reductase (*nosZ*) genes in soil metagenomes: Method development and implications for the nitrogen cycle. mBio 5, e01193–14.

Palmer, K., Horn, M.A., 2012. Actinobacterial nitrate reducers and proteobacterial denitrifiers are abundant in N_2O-metabolizing palsa peat. Applied and Environmental Microbiology 78, 5584–5596.

Philippot, L., Andert, J., Jones, C.M., Bru, D., Hallin, S., 2011. Importance of denitrifiers lacking the genes encoding the nitrous oxide reductase for N_2O emissions from soil. Global Change Biology 17, 1497–1504.

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

R Core Team, 2016. R: A language and environment for statistical computing [WWW Document]. Version 3.3.0. URL http://www.r-project.org

Ritz, K., Black, H.I.J., Campbell, C.D., Harris, J.A., Wood, C., 2009. Selecting biological indicators for monitoring soils: A framework for balancing scientific and technical opinion to assist policy development. Ecological Indicators 9, 1212–1221.

Roelandt, C., van Wesemael, B., Rounsevell, M., 2005. Estimating annual N_2O emissions from agricultural soils in temperate climates. Global Change Biology 11, 1701–1711.

Russenes, A.L., Korsaeth, A., Bakken, L.R., Dörsch, P., 2016. Spatial variation in soil pH controls off-season N_2O emission in an agricultural soil. Soil Biology and Biochemistry 99, 36–46.

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems and Environment 216, 344–355.

Schloss, P.D., Westcott, S.L., Ryabin, T., et al., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75, 7537–7541.

Schloter, M., Dilly, O., Munch, J.C., 2003. Indicators for evaluating soil quality. Agriculture, Ecosystems and Environment 98, 255–262.

Schmidt, S.K., Costello, E.K., Nemergut, D.R., et al., 2007. Biogeochemical consequences of rapid microbial turnover and seasonal succesion in soil. Ecology 88, 1379–1385.

Shoun, H., Kano, M., Baba, I., Takaya, N., Matsuo, M., 1998. Denitrification by *Actinomycetes* and purification of dissimilatory nitrite reductase and Azurin from *Streptomyces thioluteus*. Journal of Bacteriology 180, 4413–4415.

Spang, A., Poehlein, A., Offre, P., et al., 2012. The genome of the ammonia-oxidizing Candidatus *Nitrososphaera gargensis*: Insights into metabolic versatility and environmental adaptations. Environmental Microbiology 14, 3122–3145.

Stevenson, D.M., Weimer, P.J., 2007. Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75, 165–174.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., et al., 2003. The COG database: An updated version includes eukaryotes. BMC Bioinformatics 4, 1–14.

Throbäck, I.N., Enwall, K., Jarvis, A., Hallin, S., 2004. Reassessing PCR primers targeting *nirS*, *nirK* and

nosZ genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiology Ecology 49, 401–417.

Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: From genes to ecosystems. Current Opinion in Microbiology 5, 240–245.

U.S. Environmental Protection Agency, 2014. Inventory of US greenhouse gas emissions and sinks: 1990-2012. Washington, DC.

Wallenstein, M.D., Myrold, D.D., Firestone, M., Voytek, M., 2006. Environmental controls on denitrifying communities and denitrification rates: Insights from molecular methods. Ecological Applications 16, 2143–52.

Wang, Q., Quensen, J.F., Fish, J.A., 2013. Ecological patterns of *nifH* genes in four terrestrial climatic Zones. mBio 4, 1–9.

Ward, N.L., Challacombe, J.F., Janssen, P.H., et al., 2009. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. Applied and Environmental Microbiology 75, 2046–2056.

Yin, C., Fan, F., Song, A., Li, Z., Yu, W., Liang, Y., 2014. Different denitrification potential of aquic brown soil in Northeast China under inorganic and organic fertilization accompanied by distinct changes of *nirS*- and *nirK*-denitrifying bacterial community. European Journal of Soil Biology 65, 47–56.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320.

Table 2.1 16S rRNA OTUs correlated to soil N_2O emissions through elastic net modeling

			Elastic net coefficients (by alpha value)					Variance partitioning		
System	**Phylum**	**Biomarker**	0.5	0.75	1.0	**Incidence**	**Abundance (% of total reads)**	System	Year	S x Y
Corn	*Armatimonadetes*	OTU0162	-85.8	-93.1	-71.4	12	0.14 ± 0.08	0.01	**0.37**	0.02
	Chloroflexi	OTU0224	6.5			14	0.09 ± 0.04	**0.22**	**0.20**	0.02
	Crenarchaeota	OTU0476		-66.3		8	0.04 ± 0.02	0.02	0.01	0.04
	Gemmatimonadetes	OTU0354	-228.1	-206.6	-217.0	10	0.05 ± 0.03	0.00	0.03	0.03
	Planctomycetes	OTU0114	-2.1			14	0.16 ± 0.11	0.03	**0.23**	0.02
		OTU0152	27.8			15	0.14 ± 0.09	**0.11**	**0.31**	0.11
		OTU0235	77.5	99.4	90.9	11	0.15 ± 0.12	**0.26**	**0.26**	**0.20**
	Proteobacteria	OTU0016	4.8	4.8		15	1.12 ± 0.56	0.03	**0.39**	0.04
		OTU0032	24.2	31.4	30.1	15	0.62 ± 0.56	0.02	**0.33**	**0.17**
		OTU0098	-119.5	-167.2	-138.2	15	0.23 ± 0.13	0.07	**0.62**	0.03
	Unclassified	OTU0164	-119.7	-188.4	-83.1	14	0.11 ± 0.06	0.00	0.09	0.01
		OTU0245	-140.0	-166.0	-80.7	9	0.07 ± 0.04	**0.15**	0.01	0.09
		OTU0430	-41.7	-28.4		8	0.04 ± 0.02	0.02	0.12	0.04
		Total					2.76 ± 1.10	**0.10**	**0.26**	0.05
Prairie	*Acidobacteria*	OTU0021	7.8	4.8	3.3	15	0.86 ± 0.27	0.03	**0.22**	0.05
		OTU0120	90.6	87.9	93.2	15	0.17 ± 0.07	0.09	**0.20**	0.03
	Actinobacteria	OTU0383	8.7			10	0.04 ± 0.03	0.06	0.05	0.03
	Chloroflexi	OTU0091	23.9	18.2	16.1	15	0.20 ± 0.14	0.00	**0.24**	**0.17**
	Gemmatimonadetes	OTU0043	6.3	1.7		15	0.33 ± 0.11	**0.28**	0.11	0.08
	Unclassified	OTU0393	-391.2	-406.1	-438.1	8	0.03 ± 0.03	0.00	0.14	0.01
		Total					1.63 + 0.45	0.06	**0.17**	0.05

Incidences and abundances were calculated within cropping systems, with abundance presented as mean ± sd. Variance ratios are the proportion of total variance attributable to a factor, for bolded values $P < 0.05$ (not corrected for multiple comparisons). Coefficients over broader alpha values and full taxonomic information are presented in Table S2.2.

Table 2.2 *nosZ* OTUs correlated to corn system soil N_2O emissions through elastic net modeling

Family	Biomarker	Elastic net coefficients (by alpha value)			Incidence	Abundance (% of total reads)	Variance partitioning		
		0.50	0.75	1.00			System	Year	S × Y
	Year (2011)	-0.3	-0.4	-0.5					
Burkholderiales	OTU0027	181.3	217.1	258.4	11	0.17 ± 0.07	0	0.18	0.01
Rhizobiales	OTU0010	-29.1	-24.8	-20.8	14	0.40 ± 0.21	0.07	0	0.15
	OTU0021	-5.3	-6.2	-10.9	15	1.19 ± 0.83	**0.23**	0.02	0.03
	OTU0040	-6.0			15	0.57 ± 0.30	0.01	0.03	0
	OTU0074	-18.6	-13.1	-7.1	14	0.49 ± 0.23	**0.48**	0.1	0.03
	OTU0085	-7.9	-4.9	-2.8	15	0.86 ± 0.58	0.02	0.17	0.06
Rhodobacterales	OTU0165	148.3	168.0	159.7	12	0.11 ± 0.06	0.01	0.01	0.08
Rhodospirillales	OTU0158	-38.0	-36.5	-27.8	14	0.18 ± 0.19	0.03	0.12	0.04
	Total					3.83 ± 1.64	**0.15**	**0.12**	0.07

Family is for the closest BLASTX homolog with taxonomic identity; the closest homolog to most OTUs was uncultured. Incidences and abundances were calculated for the corn system, with abundance presented as mean ± sd. Variance ratios are the proportion of total variance attributable to a factor, for bolded values $P < 0.05$ (not corrected for multiple comparisons). Coefficients over broader alpha values and full taxonomic information are presented in Table S2.3.

Table 2.3 COGs correlated to corn system soil N_2O emissions through elastic net modeling

		Elastic net coefficients (by alpha value)				Variance partitioning		
Functional category	**Biomarker**	0.50	0.75	1.00	**Abundance (copies cell^{-1})**	System	Year	S × Y
Translation, ribosomal structure and biogenesis	COG0013	0.08			0.99 ± 0.02	0.00	**0.50**	0.04
	COG0220	0.9	0.3		0.50 ± 0.03	**0.08**	**0.48**	**0.14**
	COG1576	0.5	0.0		0.36 ± 0.04	0.03	**0.55**	0.03
	COG3642	-1.0	-0.1		0.22 ± 0.02	**0.37**	**0.23**	0.02
Signal transduction	COG2114	-0.1	-0.1		6.79 ± 0.52	**0.31**	**0.32**	**0.11**
	COG2197	-0.004			12.10 ± 0.40	**0.26**	**0.34**	0.04
	COG2206	-0.04			1.68 ± 0.08	0.00	0.11	**0.26**
	COG5170	-1.7			0.01 ± 0.00	**0.25**	**0.25**	0.03
Cell wall/membrane biogenesis	COG3065	2.0	2.4	1.9	0.07 ± 0.02	0.00	**0.40**	0.03
Energy production and conversion	COG0374	-4.4	-4.4	-2.7	0.14 ± 0.01	0.10	0.03	**0.25**
	COG1141	-0.8	-0.7	-0.1	0.55 ± 0.07	**0.10**	**0.35**	0.06
Nucleotide transport and metabolism	COG0563	-1.1	-0.9		1.38 ± 0.05	0.02	0.03	0.13
Carbohydrate transport and metabolism	COG0120	2.3	2.7	2.7	0.38 ± 0.03	0.00	**0.36**	0.04
Coenzyme transport and metabolism	COG3165	1.9	1.2		0.05 ± 0.01	0.08	0.12	0.02
Lipid transport and metabolism	COG1260	-1.0			0.70 ± 0.02	0.06	0.12	0.06
Inorganic ion transport and metabolism	COG1055	2.6	3.1	2.4	0.30 ± 0.03	0.00	**0.32**	0.07
	COG1393	0.02			0.60 ± 0.08	0.00	**0.50**	0.05
	COG3256	0.0002			0.24 ± 0.03	**0.05**	**0.60**	**0.10**
Secondary metabolite biosynthesis and transport	COG2312	1.6	0.8		0.27 ± 0.02	**0.14**	**0.22**	0.00
General prediction	COG3694	-0.6	-0.2		0.26 ± 0.03	**0.20**	**0.46**	0.01
	COG4589	-0.6			0.29 ± 0.02	**0.30**	0.02	0.14
Function unknown	COG0700	-5.9	-9.0	-13.6	0.11 ± 0.01	**0.14**	0.12	0.12
	COG1315	-3.9			0.01 ± 0.00	0.07	0.10	0.06

Copies per cell were estimated by normalizing by COG model length, then normalizing against 37 single-copy housekeeping genes, presented as mean ± sd. Variance ratios are the proportion of total variance attributable to a factor, for bolded values $P < 0.05$ (not corrected for multiple comparisons). Coefficients over broader alpha values and CG names are presented in Table S2.4.

Table S2.1 Coefficients and taxonomic information for microbial membrane lipids correlated to soil N_2O emissions in prairie samples through elastic net modeling

		alpha values								
System	**Biomarker**	0.5	0.6	0.7	0.8	0.9	1.0	**Incidence**	**Abundance (mol%)**	**Biological interpretation**
Corn	Year (2011)	-0.4	-0.4	-0.4	-0.4	-0.4	-0.4			
	16:1 2OH	89.6	93.4	87.8	73.6	55.0	11.6	13	1.2 + 0.5	None
	16:1ω7c	10.0	9.0	6.7	3.2			15	6.1 + 2.1	Gram-negative bacteria

Lipid inerpretations based on Balser et al., 2000, referenced in manuscript. Incidence and abundance are calculated for corn samples, abundance is a percentage of all microbial lipids, presented as mean ± sd.

Table S2.2 Taxonomic information for 16S rRNA OTUs correlated to soil N_2O emissions in prairie samples through elastic net modeling

System	Marker	Kingdom	Phylum	Class	Order
Corn	Otu00016	Bacteria(100)	Proteobacteria(100)	Betaproteobacteria(100)	unclassified(100)
	Otu00032	Bacteria(100)	Proteobacteria(100)	Betaproteobacteria(100)	unclassified
	Otu00098	Bacteria(100)	Proteobacteria(100)	Gammaproteobacteria(100)	Legionellales(100)
	Otu00114	Bacteria(100)	Planctomycetes(100)	Planctomycea(100)	Pirellulales(100)
	Otu00152	Bacteria(100)	Planctomycetes(100)	Planctomycea(100)	Gemmatales(100)
	Otu00162	Bacteria(100)	Armatimonadetes(100)	CH21(100)	unclassified(100)
	Otu00164	Bacteria(100)	unclassified	unclassified	unclassified
	Otu00224	Bacteria(100)	Chloroflexi(100)	Thermomicrobia(100)	Sphaerobacterales(100)
	Otu00235	Bacteria(100)	Planctomycetes(100)	Planctomycea(100)	Gemmatales(100)
	Otu00245	Bacteria(100)	unclassified	unclassified	unclassified
	Otu00354	Bacteria(100)	Gemmatimonadetes(98)	Gemmatimonadetes(98)	Gemmatimonadales(98)
	Otu00430	Bacteria(100)	unclassified(92)	unclassified(92)	unclassified(92)
	Otu00476	Archaea(100)	Crenarchaeota(100)	Thaumarchaeota(100)	Nitrososphaerales(100)
Prairie	Otu00021	Bacteria(100)	Acidobacteria(100)	Holophagae(99)	DS-18(99)
	Otu00043	Bacteria(100)	Gemmatimonadetes(100)	Gemmatimonadetes(100)	Gemmatimonadales(100)
	Otu00091	Bacteria(100)	Chloroflexi(99)	Anaerolineae(89)	unclassified(89)
	Otu00120	Bacteria(100)	Acidobacteria(98)	Acidobacteria(98)	Acidobacteriales(98)
	Otu00383	Bacteria(100)	Actinobacteria(100)	Actinobacteria(100)	0319-7L14(100)
	Otu00393	Bacteria(100)	unclassified(90)	unclassified(90)	unclassified(90)

Table S2.3 Taxonomic information for *nosZ* OTUs correlated to corn system soil N_2O emissions through elastic net modeling

Biomarker	Representative sequnce ID	Closest BLASTX hit in NCBI with taxonomic information	Percent identity	Class	Family	Species
OTU0010	IC3ZW2301CM10T	DE0RYD0K015	97%	Alphaproteobacteria	Rhizobiales	*Bradyrhizobiaceae bacterium*
OTU0021	ICWNDJH01AL55P	DE0Y2Y8C01R	100%	Alphaproteobacteria	Rhizobiales	*Bradyrhizobium japonicum*
OTU0027	ICWNDJH01DE30O	DE17YZR1015	76%	Betaproteobacteria	Burkholderiales	*Achromobacter sp.*
OTU0040	ICWNDJH01BKTTL	DEJ6DMTU014	85%	Alphaproteobacteria	Rhizobiales	*Bradyrhzizobium oligotrophicum*
OTU0074	IC3ZW2301DRKQO	EE6J5360014	93%	Alphaproteobacteria	Rhizobiales	*Microvirga vignae*
OTU0085	ICWNDJH01B3F0V	DEJ7NTGX014	93%	Alphaproteobacteria	Rhizobiales	*Sinorhizobium meliloti*
OTU0158	ICWNDJH01BCN0T	DEJ8T0WZ014	98%	Alphaproteobacteria	Rhodospirillales	*Skermanella aerolata*
OTU0165	IC3ZW2301DFH6S	DEJN1UYY01R	78%	Alphaproteobacteria	Rhodobacterales	*Paracoccus sp*

Table S2.4 Coefficients and functional information for COGs correlated to corn system soil N2O emissions through elastic net modeling

Biomarker	Function class	COG name
COG0013	J	Alanyl-tRNA synthetase
COG0120	G	Ribose 5-phosphate isomerase
COG0220	J	tRNA G46 methylase TrmB
COG0374	C	Ni,Fe-hydrogenase I large subunit
COG0563	F	Adenylate kinase or related kinase
COG0700	S	Spore maturation protein SpmB (function unknown)
COG1055	P	Na+/H+ antiporter NhaD or related arsenite permease
COG1141	C	Ferredoxin
COG1260	I	Myo-inositol-1-phosphate synthase
COG1315	S	Uncharacterized conserved protein, DUF342 family
COG1393	P	Arsenate reductase and related proteins, glutaredoxin family
COG1576	J	23S rRNA pseudoU1915 N3-methylase RlmH
COG2114	T	Adenylate cyclase, class 3
COG2197	T	DNA-binding response regulator, NarL/FixJ family, contains REC and HTH domains
COG2206	T	HD-GYP domain, c-di-GMP phosphodiesterase class II (or its inactivated variant)
COG2312	Q	Erythromycin esterase homolog
COG3065	M	Starvation-inducible outer membrane lipoprotein
COG3165	H	Ubiquinone biosynthesis protein UbiJ, contains SCP2 domain
COG3256	P	Nitric oxide reductase large subunit
COG3642	J	tRNA A-37 threonylcarbamoyl transferase component Bud32
COG3694	R	ABC-type uncharacterized transport system, permease component
COG4589	R	Predicted CDP-diglyceride synthetase/phosphatidate cytidylyltransferase
COG5170	T	Serine/threonine protein phosphatase 2A, regulatory subunit

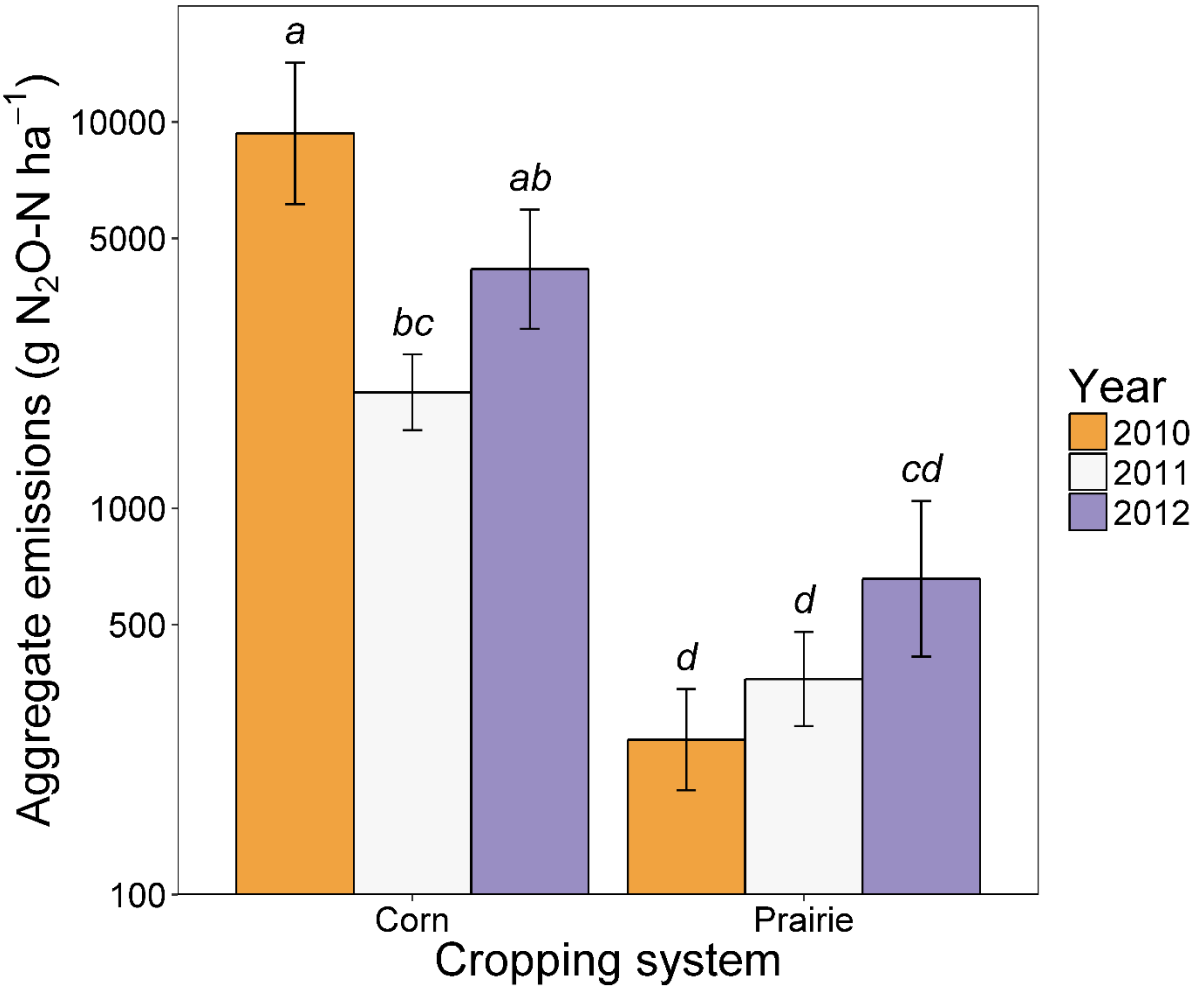

Figure 2.1 Aggregate annual N_2O emissions by cropping system and year. Data are presented on a logarithmic scale. Groups sharing a letter are not significantly different ($P > 0.05$ following a Tukey multiple comparison adjustment)

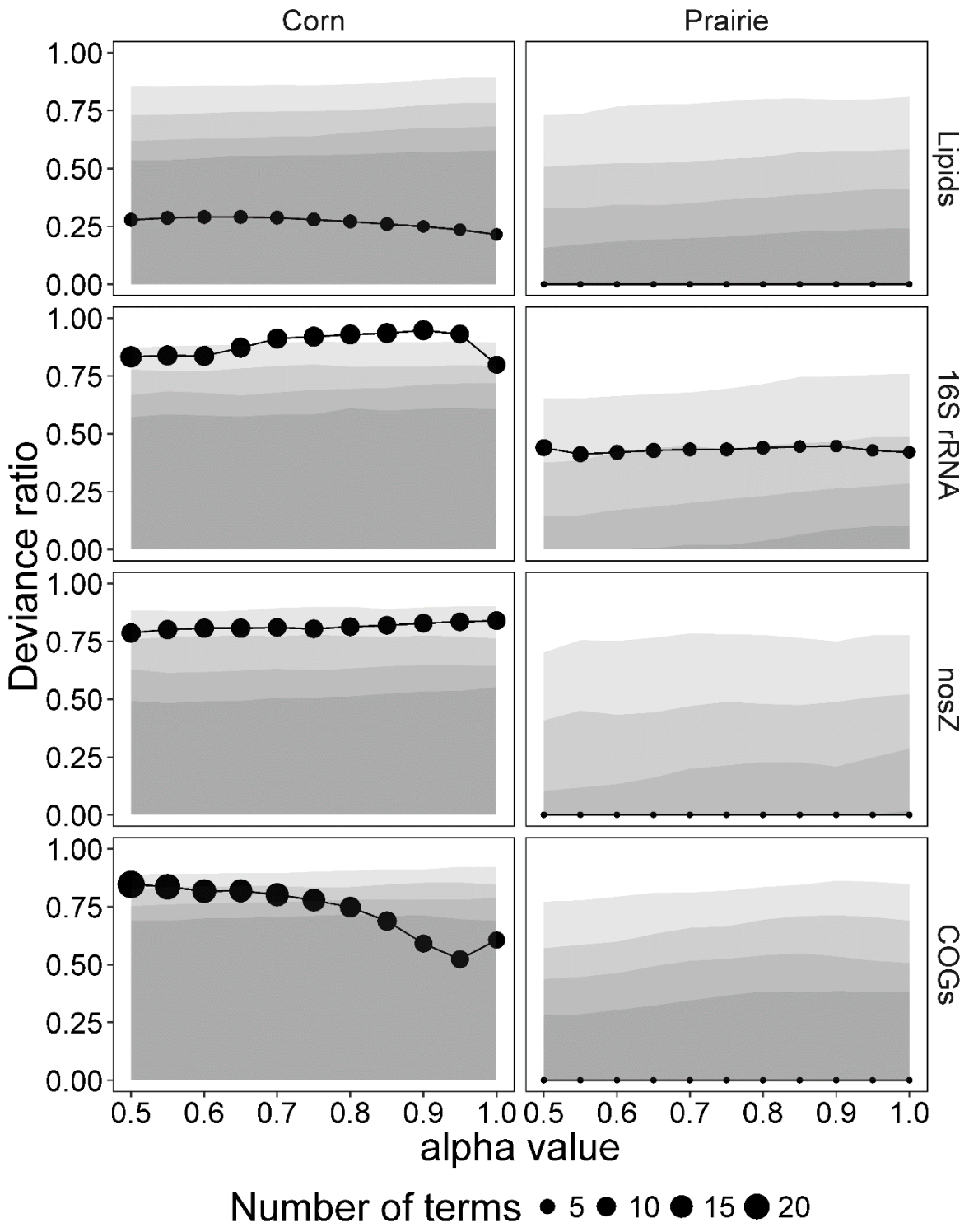

Figure 2.2 Performance of elastic net models correlating microbial biomarkers to N_2O fluxes in corn and prairie cropping systems. Points indicate deviance ratios obtained with different weightings of ridge and lasso regression (alpha values). Point size indicates the number of terms retained by the model, including an intercept. Progressively darker shades of gray indicate deviance ratios for the 95th, 90th, 85th, and 80th percentiles for 1000 within-year data permutations.

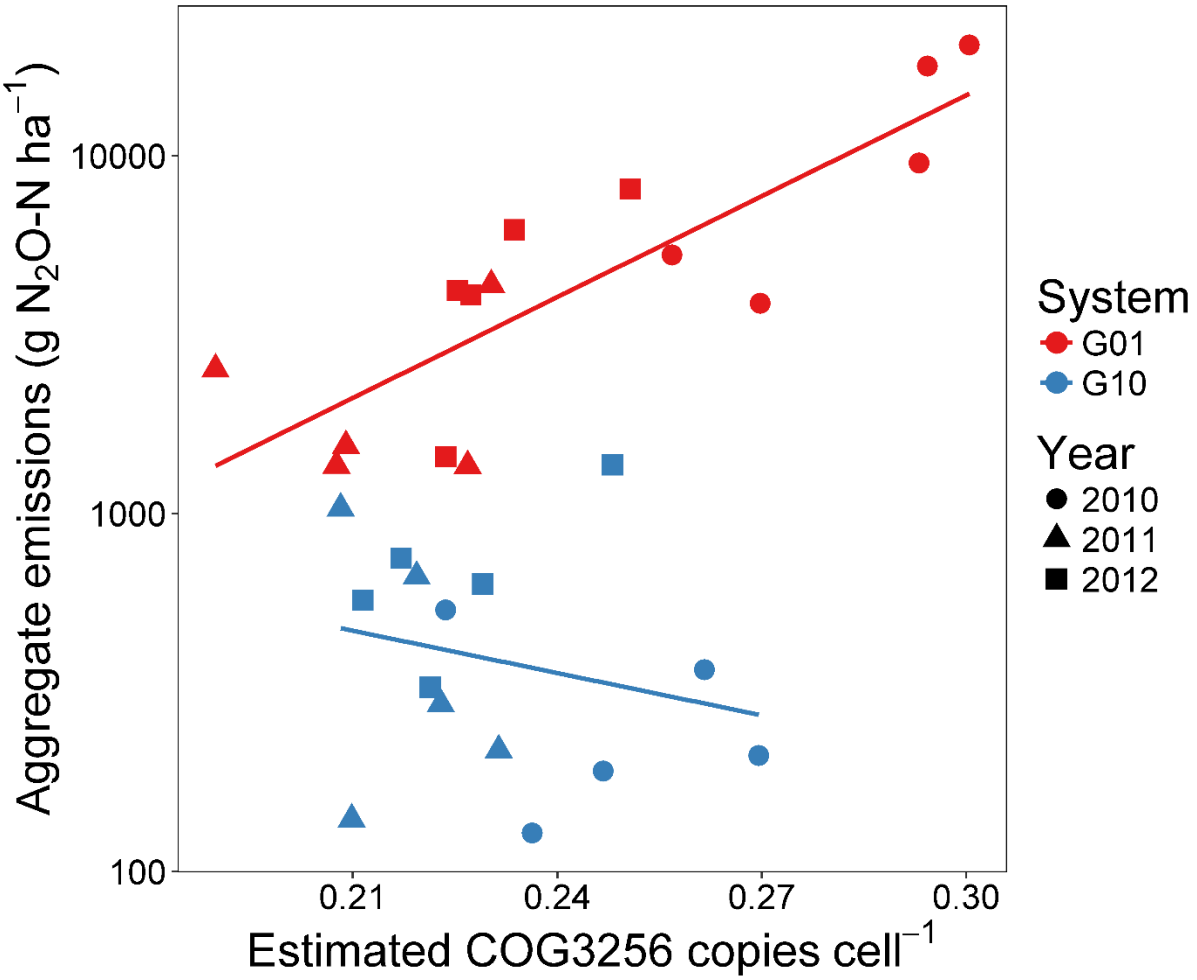

Figure S2.1 Relationship between aggregate N_2O emissions and abundance of COG3256 (nitrous oxide reductase, *norB*).

CHAPTER 3

Bioenergy feedstock cropping systems exert significant but inconsistent effects on soil microbial functional gene abundances

David S. Duncan, Randall D. Jackson

Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA

DOE-Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

Abstract

Soil microbial communities regulate key nutrient cycles in agricultural systems, making them critical for healthy bioenergy feedstock cropping systems. We used functional gene abundance profiles measured from shotgun metagenomic sequencing to study how soil microbial communities responded to eight bioenergy feedstock cropping systems: continuous corn, miscanthus, hybrid poplar, a mixture of native grasses, in addition to both fertilized and nonfertilized switchgrass and restored tallgrass prairie. These systems were grown in agronomic trials in south-central Wisconsin (ARL) and southwest Michigan (KBS). We sampled ARL annually from 2010 to 2012 and KBS in 2012. Microbial community differences existed between sites, but not among years at ARL, and among cropping systems in all cases except for 2011 at ARL. Functional gene profile dissimilarities rarely matched ecological differences among cropping systems, while relative relationships among cropping shifted over site-years. The corn system at ARL had significantly lower average abundances for all function categories considered, suggesting smaller average genome sizes, and at both sites exhibited substantial variability in denitrification pathway gene abundances. Nitrogen fertilization effects were not apparent in functional gene profiles, abundances within function categories, or abundances of denitrification pathway genes. Our findings suggest that within-system variability in microbial functional gene abundances outweighs systematic cropping system influences.

3.1 Introduction

Soil microbes drive many key ecosystem dynamics, influencing biogeochemical cycles and the composition and productivity of plant communities (Schnitzer et al., 2011; van der Heijden et al., 2008; Wagg et al., 2014). Microbially-driven dynamics may be particularly important to cropping systems that produce biomass feestocks for bioenergy and bioproducts. In contrast to the virtually exclusive focus placed on production in conventional grain and forage cropping systems, ecological services and sustainability form a core element of the impetus for bioenergy feedstock cropping systems (Landis et al., n.d.; Robertson et al., 2008). These systems are valued and evaluated not only for their productivity (Sanford et al., 2016), but also for their capacity to provide a number of ecosystem services including carbon storage (Gelfand et al., 2011; Zenone et al., 2013), mitigation of greenhouse gas emissions (Hudiburg et al., 2015; Oates et al., 2016), and provision of habitat for desirable animal species (Robertson et al., 2011, 2012; Werling et al., 2014). Much of the rationale for the sustainability and desirability of bioenergy feedstock cropping systems is predicated on their reduced requirements for exogenous nutrients, agricultural chemicals, and other management inputs (Dale et al., 2014; Tilman et al., 2006). This combination of broader ecological demands and reduced inputs requires greater resilience and resource-use efficiency, which in turn relies heavily on the interactions of plants, microbes, and management activities (Bardgett and McAlister, 1999).

Understanding the forces that shape soil microbial communities is critical for learning how to manage them. In principle, ecosystem properties should exert clear directional effects on soil microbial community composition. Differences in plant species and variety can select directly for particular taxa (Berg and Smalla, 2009; Kowalchuk et al., 2002) or more indirectly influence the selective environment via their influence on energy and nutrient cycles (Butler et al., 2003; Pathan et al., 2015). Exogenous nutrients, particularly nitrogen, greatly impact particular taxa (Fierer et al., 2012a; Leff et al., 2015). At the same time, soil microbial communities can prove exceedingly slow to change, reflecting land use legacies more than present conditions (Jangid et al., 2011). Cropping system effects may also be dampened or magnified by interannual variability (Smith et al., 2015) or by environmental factors

unrelated to cropping system differences (Mao et al., 2013). While high ecological dissimilarity simplifies detection of coarse, systematic differences in microbial community composition (Fierer et al., 2012b), exploring more subtle, yet significant, differences in composition may require systems with less overwhelming ecological differences.

The motivation to study microbial community compositions largely presupposes that compositional changes translate to changes in ecosystem-level functioning. While fully understanding this linkage remains one of the grand challenges of microbial ecology (Torsvik and Øvreås, 2002), there is growing appreciation of the functional differences among microbial communities and of how these differences may influence ecosystem processes. For example, diversity of organisms capable of reducing nitrous oxide (N_2O) influences the ratio of N_2O:N_2 produced during denitrification (Domeignoz-Horta et al., 2015). Community composition and function may be more closely linked in processes conducted by smaller, less diverse groups of microorganisms (Levine et al., 2011), although aggregate community-level differences among ecosystems have also been observed for broader traits, such as microbial growth efficiency (Lee and Schmidt, 2014). While this variability in microbial community functional capabilities is increasingly represented in process-based models, generation of empirical data on these functions remains a major bottleneck (Wieder et al., 2014).

Functional gene characterization provides a powerful tool for exploring and predicting differences in microbial community capabilities, particularly when information on individual taxa and their capabilities is limited. Gene-targeted methods, such as qPCR, allow for determination of the abundances of specific genes of interest (e.g. Mao et al., 2013). Even with recent advances in throughput (Devonshire et al., 2013), this approach still limits the number of genes that can be analyzed at once. Functional gene microarrays (e.g. He et al., 2010) permit simultaneous quantification of large numbers genes, but are still limited in scope to sequences sharing substantial homology to pre-defined sets of probes. In contrast, shotgun metagenomic sequencing largely avoids dependence on gene targeting; as its throughput has increased and its costs have decreased, it is an increasingly attractive alternative for characterizing functional gene abundances (Eisen, 2007; Tringe et al., 2005). Shotgun sequencing can

identify novel or atypical forms of functional genes which would go undetected by targeted methods (Orellana et al., 2014; Sanford et al., 2012). Whole or partial genome reconstruction from metagenomic sequence provides considerable information about both the composition of a community and the genetic capabilities of its constituents (Tyson et al., 2004). Metagenomic assembly remains a major challenge for complex microbial communities, such as those found in agricultural and grassland soils (Howe et al., 2014), but direct annotation of short reads may suffice to infer differences in microbial community function that are consistent with ecological differences among samples (Fierer et al., 2012b; He et al., 2015). One frequently-used annotation framework is the Clusters of Orthologous Groups (COGs) database (Galperin et al., 2015; Tatusov et al., 2003), which identifies genes that are paralogous across multiple microbial lineages. One advantage of this framework is that individual COGs are further classified into functional categories, making it possible to track changes in the relative importance of broad groupings of genetic functions.

We explored how different bioenergy feedstock cropping systems influence the functional gene profiles of their associated soil microbial communities. Our study was conducted in Bioenergy Cropping Systems Experiments (BCSEs) situated Wisconsin and Michigan. We sampled near the end of the growing season, from 2010 to 2012 at ARL and in 2012 at KBS. The BCSEs were set up to compare the productivity and ecological properties of a diverse array of bioenergy cropping systems differing in their perenniality, plant species diversity, and agronomic management intensity (Sanford et al., 2016). These systems exhibit considerable differences in their nitrogen cycles, notably in their nitrous oxide emissions (Duran et al., 2016; Oates et al., 2016). We previously reported significant differences in microbial community taxonomic and functional gene profiles between the two most ecologically dissimilar systems, no-tillage continuous corn (*Zea mays* L.) and a nonfertilzed assemblage of tallgrass prairie species, at the Wisconsin BCSE (Duncan et al., 2016). Here, we extend that study, analyzing functional gene profiles derived from shotgun metagenomic sequencing for eight cropping systems at both sites. We addressed whether functional gene profiles differed systematically among cropping systems, sites, or years and whether these patterns differed for groups of genes with related functions.

3.2 Methods

3.2.1 Study design and soil sampling

The DOE-Great Lakes Bioenergy Research Center's BCSEs were established in 2008. Both BCSEs contained five replicate randomized complete blocks, with 0.12 ha plots (27- × 43-m) managed with standard-sized field equipment (full agronomic details in Sanford et al., 2016). We investigated six systems: continuous no-tillage corn, monocultures of switchgrass (*Panicum virgatum* L.), miscanthus (*Miscanthus* × *giganteus*), and hybrid poplar (*Populus nigra* × *P. maximowiczii*), a native grass mix (five species), and a restored prairie (assemblage of 18 tallgrass prairie species). Full species lists and variety information are presented in Table S1 of Oates et al. (2016). The continuous corn system received annual fertilization based on spring soil tests, with an average of 167 kg N ha^{-1} y^{-1} (5-14-42 NPK granular starter fertilizer and 28-0-0 urea-ammonium nitrate side dress). The poplar received a single nitrogen application in 2010 (210 kg N ha^{-1} as 34-0-0 granular ammonium nitrate). The remaining systems received annual fertilization (56 kg N ha^{-1} as 34-0-0 granular ammonium nitrate) in a single application in the spring. This fertilization regime began in 2010 to limit competition from annual weeds while the perennial crops established. We sampled from a split-plot fertilization experiment superimposed on the switchgrass and restored prairie systems, with a subplot (10- × 43-m) on the western side of the plots receiving a contrasting treatment to the main plot. The switchgrass main plots and restored prairie subplots received annual fertilization while the restored prairie main plot and switchgrass subplot did not.

The Michigan BCSE was established at the Michigan State University W.K. Kellogg Biological Research Station (KBS, 42 23'47" N, 85 22'26"W, 288 m.a.s.l.). Soils at KBS were predominantly Kalamazoo loam (Fine-Loamy, Mixed, Semiactive, Mesic Typic Hapludalfs). Mean annual temperature from 1981 to 2010 was 9.9 °C and mean annual precipitation was 1027 mm (MSCO, 2013). In 2012, the only year this site was sampled, precipitation between May and September was significantly below the 30-year average (Sanford et al., 2016). Prior to BSCE establishment, the previous crop was alfalfa (*Medicago sativa* L.).

The Wisconsin BCSE was established at University of Wisconsin-Madison Arlington Agricultural Research Station (ARL, 43 17'45" N, 89 22'48" W, 315 m.a.s.l). Soils at ARL were predominantly Plano silt-loam (Fine-Silty, Mixed, Superactive, Mesic Typic Arguidolls). Mean annual temperature from 1981 to 2010 was 6.8 °C and mean annual precipitation was 869 mm (NWS, 2013). Of the three years sampled, 2010 had above-average precipitation through August, while 2011 and 2012 were at or below the 30-year average during the growing season (Sanford et al., 2016). Blocking at ARL was designed to control for differences in prior land use, as prior to BCSE establishment blocks A1-A3 were in a corn-soybean (*Glycine max* L.) rotation while blocks A4-A5 were an alfalfa-orchardgrass (*Dactylis glomerata* L.) hay mixture. The miscanthus system was replanted at ARL in 2010 following severe stand loss during the 2008-09 winter.

Including subplot treatments, eight total systems were sampled in this study: continuous corn, fertilized and unfertilized switchgrass, miscanthus, poplar, native grasses, old field, and fertilized and unfertilized prairie. Not all treatments were sampled in all years or at both sites; a full sample list is presented in Table S3.1. Soils were sampled 2010-08-24, 2011-08-21, and 2012-08-30 at ARL, and on 2012-08-27 at KBS. In each plot, five cores (3.7 cm diameter, 15 cm depth) were collected in a staggered transect and composited by sieving to 2 mm. Soils were lyophilized and stored frozen at -20 °C.

3.2.2 DNA extraction and sequencing

DNA was extracted from soils using an adaptation of the approach developed by Stevenson and Weimer (2007) with full details given in Duncan et al. (2016). Soils were ground with liquid nitrogen in a ceramic mortar to disrupt aggregates. Cell lysis was achieved by a combination of bead beating (0.1-mm silica zirconia beads, 2 × 10 min at room temperature), phenol (500 µl in a total liquid volume of 1550 µl), SDS (50 µl) and heat (10 min at 60 °C) in a high salt buffer (100 mM Tris-HCl, 10 mM EDTA, 0.15 M NaCl, pH 8.0). Samples were separated by centrifugation (16,000 × *g* for 10 min), with successive washes with 500 µl phenol, 500 µl 1:1 phenol:chloroform, and 500 µl chloroform. DNA was precipitated with 3 M sodium acetate and isopropanol. All samples were further cleaned with a Power Soil Cleanup Kit (Mo Bio Laboratories, Carlsbad CA).

Sequencing was conducted by the DOE Joint Genome Institute using an Illumina HiSeq platform (Illumina, San Diego, CA) with 12 samples multiplexed per lane. Reads were assembled using SOAPdenovo (v 1.05) and called using FragGenScan (v1.16), prokaryotic GeneMark.hmm (v2.8), Metagenome Annotator (v1.0) and Prodigal (v2.5). For this study, we used annotations from the updated clusters of orthologous groups (COGs) database (Galperin et al., 2015; Tatusov et al., 2003). Samples were submitted for sequencing in three batches, with the first batch containing all continuous corn and nonfertilized prairie samples from ARL. Sequencing data are available through the Genomes Online Database (https://gold.jgi.doe.gov) under Study ID GS0095510. Project ID numbers and batch information for individual samples are given in Table S3.1.

3.2.3 Statistical analysis

COG abundances were obtained from the Integrated Microbial Genomes portal (https://img.jgi.doe.gov/m/), using coverage-based copy number estimates. Assembly rates were extremely low (<1% of reads mapped), thus unassembled reads dominated our analysis, making this effectively a direct annotation of short reads. Consequently, we applied the short-read relativization approach described by He et al. (2015). COG copy numbers were first divided by their consensus sequence length to give coverage per base, then all COGs were divided by the average coverage of a suite of universal single-copy housekeeping genes to give an estimate of copy number $cell^{-1}$ for each COG. Housekeeping COGs were excluded from subsequent analyses. Table S3.2 contains information on COG consensus lengths, housekeeping gene identity, and function categories. COGs involved in denitrification were identified based on information from He et al. (2015) and Wang et al. (2014) and are described in Table S3.3.

Analysis was conducted in the R statistical environment (R Core Team, 2014). Multivariate analysis used the 'vegan' package (Oksanen et al., 2013). Intersample distances were calculated from estimated COG copy number $cell^{-1}$ using Bray-Curtis distances. Permutational multivariate analysis of variance used the ‘adonis’ function. Nonmetric multidimensional scaling (NMDS) ordination was conducted by chaining 10 calls of the ‘metaMDS’ function for each ordination to increase exploration of

the solution space. Intersample distances were corrected for metagenome size and batch effects (see Section 3.1) by using these as predictors in the ‘capscale’ function, and using the residuals from that as the distance matrix for subsequent ordinations. Means differences in copy number $cell^{-1}$ were evaluated using the packages 'lme4' (Bates et al., 2015), 'lsmeans' (Lenth, 2013), and 'pbkrtest' (Halekoh and Højsgaard, 2014). Values were square-root transformed prior to analysis to achieve a uniform distribution of residuals, and back-transformed for display.

3.3 Results

3.3.1 Sequencing effort

We generated 142 shotgun metagenomes from eight cropping systems, spanning 2010 to 2012 at the Wisconsin site (ARL) and 2012 at the Michigan site (KBS). Metagenomes differed in size, ranging from 0.63 to 2.93 Gbp, with a median of 1.40 Gbp (Table S3.1). Metagenome size differed among batches of samples sequenced concurrently ($F_{2,115} = 16.7$, $P < 0.01$); metagenomes from the second batch were ~0.28 Gbp smaller than those from the first and third batches (see Table S3.1 for sample batch information). Metagenome sizes further differed by year ($F_{2,115} = 6.9$, $P < 0.01$), with 2012 having smaller metagenomes than 2010. Assembly was minimal (<1% of reads mapped onto scaffolds). A subset of 12 samples were sequenced in both the second and third batches. Batch effects are clearly visible, although the relative arrangement of samples remains relatively consistent in both batches (Fig S3.1).

3.3.2 Site-year effects

Based on permutational multivariate analysis of variance, differences in metagenome size accounted for 3.8% of intersample distances, while differences among sequencing batches accounted for an additional 25.8% (both $P < 0.01$). Differences between sites accounted for 9.2% ($P < 0.01$), differences among years within ARL accounted for 5.8%, and differences among crops accounted for 8.5% of variability (all $P < 0.01$). Intersample distances (corrected for metagenome size and sequencing batch as detailed in Section 2.3) were reasonably well-represented by two-dimensional nonmetric multidimensional scaling (NMDS, Fig. 3.1). The study sites separated clearly, if imperfectly, while the three years at ARL overlapped. This ordination was moderately correlated to the ordination of these data without corrections for metagenome

size and sequencing batch (Procrustes correlation = 0.75, $P < 0.001$) which showed effectively the same trends (Fig. S3.2).

We used Mantel tests to evaluate whether intersample differences remained consistent across years at ARL. For each pair of years, we built distance matrices from the subset of samples that were present in both years. The least similar years were 2010 and 2011 (Mantel r = 0.48), while 2012 was more similar to both 2011 (Mantel r = 0.63) and 2010 (Mantel r =0.70). These correlations were statistically significant in all cases ($P < 0.01$).

3.3.3 Cropping system effects

We analyzed variance explained by cropping system effects separately within each site-year (Table 3.1). After accounting for metagenome size and sequencing batch effects, cropping systems effects were significant in all cases except for ARL in 2011. Metagenome size and sequencing batch effects differed in importance among site-years but were consistently significant. Their effect was smallest at KBS in 2012; KBS samples were sequenced in only two batches, while all three batches were represented in other cases (Table S3.1).

NMDS ordination revealed cropping systems differences that were inconsistent across site-years, with relatively low stress levels indicating a reasonable depiction of overall intersample distances (Fig. 3.2). ARL in 2010 most resembled the patterns we expected, with the corn and nonfertilized prairie systems on opposite ends of a continuum and the fertilized perennial systems occupying an intermediate space. This pattern was not replicated in other site-years. There were no clear separations among cropping systems in 2011 or 2012 at ARL, although in 2012 some cropping systems, notably fertilized switchgrass and both prairies, formed tighter clusters. KBS in 2012 exhibited the greatest cropping system effects, partially overlapping but still visibly distinct clusters for each system sampled. Notably, ARL in 2010 was the only case where differences in functional gene profiles matched our expectations based on assumed ecological differences among the cropping systems.

Intersample distances uncorrected for metagenome size and sequencing batch effects resulted in very different ordinations (Fig. S3.2). The corn system was a clear outlier in all years at ARL while the

unfertilized prairie was a similar outlier in 2010 and 2012. These two systems, whose combined samples constituted the first sequencing batch (Table S3.1), clustered together at ARL, in sharp contrast to what we observe after correcting for batch effects.

3.3.4 Dynamics of COG function categories by site-year and cropping system

We explored whether COG functional categories followed the same site-year and cropping system dynamics we observed for the full set of functional genes. We focused on six categories which we expected *a priori* would respond to management or other cropping system differences (Table 3.2): changes in forms and availability of carbon and other nutrients were expected to impact genes involved in the transport and metabolism of carbohydrates (G) and inorganic ions (P), as well as energy production and conversion (C), while plant-microbe and microbe-microbe interactions were expected to be reflected by genes involved in defense (V), signal transduction (T), and processes involving secondary metabolites (Q). Intersample distances for each category were highly correlated to distances in overall gene profiles (Table 3.2).

We looked further into within site-year cropping systems dynamics (Table 3.3). Intersample distances remained highly correlated between the full dataset and function categories, although these correlations were generally weakest for KBS. Cropping systems effects were greater among genes involved in signal transduction across site years (Table 3.3); when these were ordinated there was slightly greater separation of the native grass mixture from the switchgrass and miscanthus systems at ARL in 2010 and KBS in 2012, but otherwise no major changes (Fig. S3.3).

We found few systematic differences in average gene copy numbers within functional categories among cropping systems (Fig. 3.3). The corn system was a clear outlier at ARL, with the lowest copy number for all functional categories, although in some instances it was not significantly different from the unfertilized prairie. In nearly all cases at ARL, the nonfertilized prairie had the second-lowest average copy number, although frequently it was closer to the perennial cropping systems than it was to corn. Patterns at ARL were generally consistent across years and function categories. Cropping system effects

at KBS were much weaker. Copy number averages across all cropping systems did not differ among site-years.

3.3.5 Denitrification pathway genes

Fold changes in copy numbers of 16 COGs involved in denitrification (Table S3.2) exhibited some cropping system patterns (Fig. 3.4). The corn system frequently had individual COG copy numbers that differed from those of other systems within a given site-year. Where function category-level copy number differences between corn and other systems were smaller at KBS than ARL (Fig. 3.3), denitrification COG abundances were, if anything, more likely to vary for the KBS corn system (Fig. 3.4). Contrary to what might have been expected given patterns observed at the level of functional categories (Section 3.4), the corn system did not have uniformly below-average copy numbers for all COGs, notably in 2012 at both sites where certain COGs involved in nitrate reduction were more abundant in corn than other systems. We observed no consistent cropping system-based differences in copy numbers of genes involved in different stages of denitrification. COGs involved in a particular step of denitrification did not generally synchronize their relative abundances, although genes involved in nitrous oxide formation and consumption were broadly above the average abundance at ARL in 2010 and 2012, and below average at KBS. The two key catalytic genes for these processes, nitric oxide reductase (*norB*, COG3256) and nitrous oxide reductase (*nosZ*, COG4263) were generally less variable among cropping systems within a site-year than their accessory genes (COG4548 and COG3420).

3.4 Discussion

The cropping systems we characterized fall along an ecological gradient with continuous corn and unfertilized prairie at its extremes. While our two study sites differed in their geographic location and environmental properties, the sites themselves covered relatively small, homogenous areas, as befits a cropping systems trial (Sanford et al., 2016). We conducted our study over three consecutive years, a short period over which to observe microbial community changes (Allison et al., 2005; Jangid et al., 2011). We thus expected that any cropping system effects we observed should be relatively consistent across years and replicate plots. Instead, we found that cropping systems exerted significant, but limited

and inconsistent influence over functional gene abundance profiles. This suggests that the composition and functional capabilities of soil microbial communities may respond to factors other than those which characterize the differences among cropping systems.

Prior to interpreting our findings, we must acknowledge potential biases due to technical and methodological issues. Our samples were sequenced in three distinct batches (see Section 2.2). We are aware that the IMG assembly and annotation pipeline differed among batches (Torben Nielsen, personal communication), as did average metagenome sizes. Functional gene profiles reflected effects of both batches and metagenome sizes. We corrected these statistically to the extent of our ability, but lingering batch effects may still have increased variability in our dataset or otherwise influenced our results.

3.4.1 Functional gene profiles differed among cropping systems, but not systematically

We observed differences in soil microbial functional gene profiles among cropping systems in all site-years except for 2011 at ARL. These differences were, however, relatively small, and rarely resulted in clearly distinct clusters. Moreover, cropping system effects on functional gene profiles varied across years and were inconsistent with ecological differences among systems. Ecologically coherent and consistent cropping system effects on soil microbial communities are common in the literature (Liang et al., 2013; Mbuthia et al., 2015). In particular, annual agricultural systems reliably differ from grasslands and other perennial systems (Allison et al., 2005; Mao et al., 2013). We only observed differences between the continuous corn and unfertilized prairie systems in 2010 at ARL, although overall the corn system behaved differently from the other systems.

At ARL, the continuous corn system consistently had lower average copy numbers of COGs in the function categories we studied. Copy numbers were lower for the unfertilized prairie as well, but to a lesser extent and far less consistently, indicating this was only partially an effect of sequencing batch. More telling, continuous corn also exhibited the most erratic patterns in denitrification gene abundances at both ARL and KBS, the latter of which did not suffer from confounded sequencing batch effects. Previous reports show corn soil microbial communities consistently differing from those of perennial bioenergy crops (Jesus et al., 2016; Mao et al., 2013). Conventionally-managed annual agricultural

systems such as corn are characterized by fluxes of nutrients and energy that are large and simplified (Gliessman, 2007). These systems create high concentrations of resources present over short periods of time and in a narrow range of forms. These conditions could favor narrower groups of microbial taxa which can rapidly exploit large, brief, homogenous resource pulses (Bastian et al., 2009; Fierer et al., 2012a). The broadly lower abundance of functional genes in the corn system could stem from pared down genomes in these taxa, reported to occur in response to nutrient addition (Leff et al., 2015). Overall, this suggests the difference between annual and perennial cropping systems is likely greater than the differences among perennial systems, as other studies have suggested (Jesus et al., 2016; Mao et al., 2013).

3.4.2 Nitrogen management effects were not reflected in functional gene abundances

Nitrogen fertilizer effects on microbial community composition and activity are broadly consistent in the literature (Bodelier, 2011; Bradley et al., 2006; Fierer et al., 2012a). We observed similar fertilizer responses at the agronomic level, with fertilization increasing nitrous oxide emissions in the switchgrass, native grass, and prairie systems (Duran et al., 2016) and increases soil nitrate concentrations (Chapter 4). Nitrogen fertilization in the switchgrass and prairie systems substantially reduced biomass of arbuscular mycorrhizal fungi and Gram-negative bacteria (Oates *et al.*in revison). Despite this, nitrogen fertilization exerted no visible influence on functional gene abundances, either in overall patterns or specifically in denitrification pathway genes. The lack of change may simply reflect the relatively brief duration of the contrasting fertilization regimes, as Sun et al. (2015) reported diverse responses by nitrogen-cycling genes to long-term chemical fertilization.

3.4.3 COG function categories recapitulated patterns observed for all COGs

We hypothesized that certain COG function categories might be more or less responsive to differences among cropping systems and site-years. For instance, greater importance of plant-microbe interactions in the unfertilized systems might increase the abundance of genes involved in signaling and quorum sensing (Mitter et al., 2013). The only evidence we saw of this was in the continuous corn system, which had lower average abundances than other systems in all of the function categories we considered. Not all

COGs followed category-level trends, however, as we saw several cases where denitrification pathway COGs were more abundant than average in the corn system, even though it had lower average abundance of COGs related to inorganic ion transport and metabolism. While this indicates individual COGs within a category can experience distinct selective pressure, cropping system and site-year differences remained consistent across functional categories. It is plausible that the COG function category framework is too broad to capture the selective pressures in our system, and that a distinct level of resolution is necessary to explore these dynamics.

3.4.4 Taxa rather than individual genes, may reflect key environmental drivers

Metagenomics-based studies frequently take a gene-centric approach to microbial ecology (Kunin et al., 2008). Gene abundances have been linked to environmental factors (Sun et al., 2015; Yoshida et al., 2010) as well as process rates (Petersen et al., 2012; Yin et al., 2014), motivating this approach. Individual microbial genomes contain only a subset of the genes found in their broader taxon (Mira et al., 2010), while lateral gene transfer provides a mechanism to further blur connections between taxonomy and function (Lawrence, 2002). This supports focusing on genes, rather than phylogenetically-classified organisms, as the stuff of selection, a perspective that strongly motivated design of this study. However, in our system functional gene abundances may not capture the key dynamic elements of microbial community composition. Other studies conducted on BCSE systems identified more consistent cropping system effects using PLFA (Oates *et al.*in revision), amplicons of the nitrous oxide reductase gene *nosZ* (Duncan et al., 2016), and targeted assembly of nitrogen-cycling genes (Guo & Tiedje, unpublished data). All of these approaches reflect taxonomic composition, to differing extents, rather than functional gene abundance. It may be that in our study system, selection occurs on overall life strategies, rather than on specific metabolic capabilities. While we conceptualize these cropping systems as possessing fundamental ecological differences, these differences may not impose strong selection for specific genetic functions. The major contribution of studies like ours may be the identification of highly responsive microbial taxa. Isolating representatives of these taxa and studying their physiology could provide insight into the environmental drivers that matter at the microbial scale.

3.5 Conclusions

Bioenergy feedstock cropping systems and site-years influenced soil microbial functional gene profiles to a smaller and less consistent extent than we anticipated. Profiles differed generally between our two study sites, but did not differ among years at the ARL site. Variability among cropping systems was statistically significant in most cases, but relative similarities among cropping systems differed among years and rarely reflected differences in cropping system ecology or management. Notably, nitrogen fertilization had no visible effects. One possible exception was the corn system, which had lower abundances of functional genes and greater variability in abundance of denitrification genes. Interpreting this is difficult, however, given that batches of samples sequenced together shared certain similarities and the corn samples were sequenced in the same batch. In our study system, selective pressures may act more visibly upon taxonomic composition, rather than the abundance of individual genes. Nonetheless, functional gene profiles may provide insight into the environmental dynamics that shape microbial communities at relatively small spatial, temporal, and ecological scales.

Acknowledgments

The sequencing underlying this project would not have happened without the encouragement and support of S Slater. We thank T Glavina del Rio, S Tringe, B Foster, and T Nielsen for facilitating the sequencing, analysis, and troubleshooting process for these data, M Miller, J Sustacheck, and D Stevenson for assisting in the unexpectedly involved DNA extraction process, and S VanderWulp for providing KBS soil samples. DSD was supported in part by a D.C. Smith Agronomy Department Fellowship and a Thomsen Wisconsin Graduate Fellowship. Funding was provided by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830). The sequencing work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Allison, V.J., Miller, R.M., Jastrow, J.D., Matamala, R., Zak, D.R., 2005. Changes in soil microbial community structure in a tallgrass prairie chronosequence. Soil Science Society of America Journal 69, 1412–1421.

Bardgett, R.D., McAlister, E., 1999. The measurement of soil fungal:bacterial biomass ratios as an indicator of ecosystem self-regulation in temperate meadow grasslands. Biology and Fertility of Soils 29, 282–290.

Bastian, F., Bouziri, L., Nicolardot, B., Ranjard, L., 2009. Impact of wheat straw decomposition on successional patterns of soil microbial community structure. Soil Biology and Biochemistry 41, 262–275.

Bates, D., Maechler, M., Bolker, B.M., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1 – 48.

Berg, G., Smalla, K., 2009. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiology Ecology 68, 1–13.

Bodelier, P.L.E., 2011. Interactions between nitrogenous fertilizers and methane cycling in wetland and upland soils. Current Opinion in Environmental Sustainability 3, 379–388.

Bradley, K., Drijber, R.A., Knops, J., 2006. Increased N availability in grassland soils modifies their microbial communities and decreases the abundance of arbuscular mycorrhizal fungi. Soil Biology and Biochemistry 38, 1583–1595.

Butler, J.L., Williams, M.A., Bottomley, P.J., Myrold, D.D., 2003. Microbial community dynamics associated with rhizosphere carbon flow. Applied and Environmental Microbiology 69, 6793–6800.

Dale, B.E., Anderson, J.E., Brown, R.C., et al., 2014. Take a closer look: Biofuels can support environmental, economic and social goals. Environmental Science & Technology 48, 7200–7203.

Devonshire, A.S., Sanders, R., Wilkes, T.M., Taylor, M.S., Foy, C.A., Huggett, J.F., 2013. Application of next generation qPCR and sequencing platforms to mRNA biomarker analysis. Methods 59, 89–100.

Domeignoz-Horta, L.A., Spor, A., Bru, D., Breuil, M.-C., Bizouard, F., Léonard, J., Philippot, L., 2015.

The diversity of the N_2O reducers matters for the N_2O:N_2 denitrification end-product ratio across an annual and a perennial cropping system. Frontiers in Microbiology 6, 971.

Duncan, D.S., Jewell, K.A., Suen, G., Jackson, R.D., 2016. Detection of short-term cropping system-induced changes to soil bacterial communities differs among four molecular characterization methods. Soil Biology and Biochemistry 96, 160–168.

Duran, B.E.L., Duncan, D.S., Oates, L.G., Kucharik, C.J., Jackson, R.D., 2016. Nitrogen fertilization effects on productivity and nitrogen loss in three grass-based perennial bioenergy cropping systems. PLOS ONE 11, e0151919.

Eisen, J.A., 2007. Environmental shotgun sequencing: Its potential and challenges for studying the hidden world of microbes. PLoS Biology 5, e82.

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. The ISME Journal 6, 1007–17.

Fierer, N., Leff, J.W., Adams, B.J., et al., 2012b. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the National Academy of Sciences of the United States of America 109, 21390–21395.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E. V, 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research 43, D261–D269.

Gelfand, I., Zenone, T., Jasrotia, P., Chen, J., Hamilton, S.K., Robertson, G.P., 2011. Carbon debt of Conservation Reserve Program (CRP) grasslands converted to bioenergy production. Proceedings of the National Academy of Sciences of the United States of America 108, 13864–13869.

Gliessman, S.R., 2007. The agroecosystem concept, in: Agroecology: The Ecology of Sustainable Food Systems. CRC Press, Boca Raton, pp. 23 – 32.

Halekoh, U., Højsgaard, S., 2014. A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models - The R package pbkrtest. Journal of Statistical Software 59, 1–30.

He, S., Malfatti, S.A., McFarland, J.W., et al., 2015. Patterns in wetland microbial community composition and functional gene repertoire associated with methane emissions. mBio 6, e00066–15.

He, Z., Deng, Y., van Nostrand, J.D., et al., 2010. GeoChip 3.0 as a high-throughput tool for analyzing microbial community composition, structure and functional activity. The ISME Journal 4, 1167–1179.

Howe, A.C., Jansson, J.K., Malfatti, S.A., Tringe, S.G., Tiedje, J.M., Brown, C.T., 2014. Tackling soil diversity with the assembly of large, complex metagenomes. Proceedings of the National Academy of Sciences 111, 4904–4909.

Hudiburg, T.W., Davis, S.C., Parton, W., Delucia, E.H., 2015. Bioenergy crop greenhouse gas mitigation potential under a range of management practices. GCB Bioenergy 7, 366–374.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Schmidt, T.M., Coleman, D.C., Whitman, W.B., 2011. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. Soil Biology & Biochemistry 43, 2184–2193.

Jesus, E. da C., Liang, C., Quensen, J.F., Susilawati, E., Jackson, R.D., Balser, T.C., Tiedje, J.M., 2016. Influence of corn, switchgrass, and prairie cropping systems on soil microbial communities in the upper Midwest of the United States. GCB Bioenergy 8, 481–494.

Kowalchuk, G.A., Buma, D.S., de Boer, W., Klinkhamer, P.G.L., van Veen, J.A., 2002. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. Antonie van Leeuwenhoek 81, 509–520.

Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K., Hugenholtz, P., 2008. A bioinformatician's guide to metagenomics. Microbiology and Molecular Biology Reviews 72, 557–578.

Landis, D.A., Gratton, C., Jackson, R.D., et al., *in review*. Assessing the potential for biodiversity to enhance multiple ecosystem services in North Central US biomass production landscapes. Biomass and Bioenergy.

Lawrence, J.G., 2002. Gene transfer in bacteria: speciation without species? Theoretical Population Biology 61, 449–460.

Lee, Z.M., Schmidt, T.M., 2014. Bacterial growth efficiency varies in soils under different land management practices. Soil Biology & Biochemistry 69, 282–290.

Leff, J.W., Jones, S.E., Prober, S.M., et al., 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proceedings of the National Academy of Sciences 112, 10967–10972.

Lenth, R. V, 2013. lsmeans: Least-squares means [WWW Document]. Version 1.06-05. URL http://cran.r-project.org/package=lsmeans

Levine, U.Y., Teal, T.K., Robertson, G.P., Schmidt, T.M., 2011. Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. The ISME Journal 5, 1683–1691.

Liang, C., Duncan, D.S., Balser, T.C., Tiedje, J.M., Jackson, R.D., 2013. Soil microbial residue storage linked to soil legacy under biofuel cropping systems in southern Wisconsin, USA. Soil Biology & Biochemistry 57, 939–942.

Mao, Y., Yannarell, A.C., Davis, S.C., Mackie, R.I., 2013. Impact of different bioenergy crops on N-cycling bacterial and archaeal communities in soil. Environmental Microbiology 15, 928–942.

Mbuthia, L.W., Acosta-Martínez, V., DeBryun, J., et al., 2015. Long term tillage, cover crop, and fertilization effects on microbial community structure, activity: Implications for soil quality. Soil Biology and Biochemistry 89, 24–34.

Mira, A., Martín-Cuadrado, A.B., D'Auria, G., Rodríguez-Valera, F., 2010. The bacterial pan-genome: A new paradigm in microbiology. International Microbiology 13, 45–57.

Mitter, B., Petric, A., Shin, M.W., et al., 2013. Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants. Frontiers in Plant Science 4, 1–15.

MSCO, 2013. Michigan State Climatologist's Office: 27 year summary of annual values for Gull Lake (3504) 1981-2010. [WWW Document]. URL http://climate.geo.msu.edu/climate_mi/stations/3504/1981-2010 annual summary.pdf (accessed 6.2.16).

NWS, 2013. National Weather Service: Wisconsin 30 year average temperature and precipitation 1981-2010 [WWW Document]. URL www.crh.noaa.gov/images/mkx/ climate/avg_30_year_precip.png (accessed 6.2.16).

Oates, L.G., Duncan, D.S., Gelfand, I., Millar, N., Robertson, G.P., Jackson, R.D., 2016. Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. GCB Bioenergy 8, 539–549.

Oates, L.G., Duncan, D.S., Sanford, G.R., Liang, C., Jackson, R.D., *in review*. Soil microbial composition responds to establishment and management of biomass cropping systems. Agriculture, Ecosystems & Environment.

Oksanen, J., Blanchet, F.G., Kindt, R., et al., 2013. vegan: Community ecology package [WWW Document]. R Package Version 2.0-10. URL http://cran.r-project.org/package=vegan

Orellana, L.H., Rodriguez-R, L.M., Higgins, S., et al., 2014. Detecting nitrous oxide reductase (*nosZ*) genes in soil metagenomes: Method development and implications for the nitrogen cycle. mBio 5, e01193–14.

Pathan, S.I., Ceccherini, M.T., Hansen, M.A., et al., 2015. Maize lines with different nitrogen use efficiency select bacterial communities with different β-glucosidase-encoding genes and glucosidase activity in the rhizosphere. Biology and Fertility of Soils 51, 995–1004.

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. Environmental Microbiology 14, 993–1008.

R Core Team, 2014. R: A language and environment for statistical computing [WWW Document]. Version 3.1.1. URL http://www.r-project.org

Robertson, B.A., Doran, P.J., Loomis, L.R., Robertson, J.R., Shemske, D.W., 2011. Perennial biomass feedstocks enhance avian diversity. GCB Bioenergy 3, 235–246.

Robertson, B.A., Rice, R.A., Sillett, T.S., et al., 2012. Are agrofuels a conservation threat or opportunity for grassland birds in the United States? The Condor 114, 679–688.

Robertson, G.P., Dale, V.H., Doering, O.C., et al., 2008. Sustainable biofuels redux. Science 322, 49–50.

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems and Environment 216, 344–355.

Sanford, R.A., Wagner, D.D., Wu, Q., et al., 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proceedings of the National Academy of Sciences of the United States of America 109, 19709–19714.

Schnitzer, S.A., Klironomos, J.N., HilleRisLambers, J., et al., 2011. Soil microbes drive the classic plant diversity–productivity pattern. Ecology 92, 296–303.

Smith, A.P., Marín-Spiotta, E., Balser, T.C., 2015. Successional and seasonal variations in soil and litter microbial community structure and function during tropical postagricultural forest regeneration: a multiyear study. Global Change Biology 21, 3532–3547.

Stevenson, D.M., Weimer, P.J., 2007. Dominance of *Prevotella* and low abundance of classical ruminal bacterial species in the bovine rumen revealed by relative quantification real-time PCR. Applied Microbiology and Biotechnology 75, 165–174.

Sun, R., Guo, X., Wang, D., Chu, H., 2015. Effects of long-term application of chemical and organic fertilizers on the abundance of microbial communities involved in the nitrogen cycle. Applied Soil Ecology 95, 171–178.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., et al., 2003. The COG database: An updated version includes eukaryotes. BMC Bioinformatics 4, 1–14.

Tilman, D., Hill, J., Lehman, C.L., 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. Science 314, 1598–1600.

Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: From genes to ecosystems. Current Opinion in Microbiology 5, 240–245.

Tringe, S.G., von Mering, C., Kobayashi, A., et al., 2005. Comparative metagenomics of microbial communities. Science 308, 554–557.

Tyson, G.W., Chapman, J., Hugenholtz, P., et al., 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43.

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. Ecology Letters 11, 296–310.

Wagg, C., Bender, S.F., Widmer, F., van der Heijden, M.G.A., 2014. Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proceedings of the National Academy of Sciences of the United States of America 111, 5266–5270.

Wang, Z., Zhang, X.-X., Lu, X., Liu, B., Li, Y., Long, C., Li, A., 2014. Abundance and diversity of bacterial nitrifiers and denitrifiers and their functional genes in tannery wastewater treatment plants revealed by high-throughput sequencing. PloS One 9, e113603.

Werling, B.P., Dickson, T.L., Isaacs, R., et al., 2014. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. Proceedings of the National Academy of Sciences of the United States of America 111, 1652–1657.

Wieder, W.R., Grandy, A.S., Kallenbach, C.M., Bonan, G.B., 2014. Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model. Biogeosciences 11, 3899–3917.

Yin, C., Fan, F., Song, A., Li, Z., Yu, W., Liang, Y., 2014. Different denitrification potential of aquic brown soil in Northeast China under inorganic and organic fertilization accompanied by distinct changes of *nirS*- and *nirK*-denitrifying bacterial community. European Journal of Soil Biology 65, 47–56.

Yoshida, M., Ishii, S., Otsuka, S., Senoo, K., 2010. *nirK*-harboring denitrifiers are more responsive to denitrification-inducing conditions in rice paddy soil than *nirS*-harboring bacteria. Microbes and Environments 25, 45–48.

Zenone, T., Gelfand, I., Chen, J., Hamilton, S.K., Robertson, G.P., 2013. From set-aside grassland to annual and perennial cellulosic biofuel crops: Effects of land use change on carbon balance. Agricultural and Forest Meteorology 182-183, 1–12.

Table 3.1 Sequencing and cropping system effects on functional gene profiles

Site	Year	Factor	*df*	Pseudo-*F*	R^2	
ARL	2010	Batch	2	13.6	0.410	***
		Size	1	1	0.012	
		Crop	4	3.8	0.229	***
		Residual	23		0.346	
	2011	Batch	2	9.9	0.347	***
		Size	1	1.6	0.028	
		Crop	5	1.3	0.113	
		Residual	29		0.511	
	2012	Batch	2	14.4	0.394	***
		Size	1	2.5	0.034	*
		Crop	6	1.5	0.121	*
		Residual	33		0.451	
KBS	2012	Batch	1	9.9	0.213	***
		Size	1	1.3	0.029	
		Crop	5	2.6	0.284	***
		Residual	22		0.474	

Values were calculated using permutational analysis of variance. Batch indicates the groups of samples that were sequenced together while Size indicates the amount of sequence generated for each metagenome. R^2 values are sequential (Type I) and were calculated in the order listed. Significance was assessed by 9999 unconstrained permutations:
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$

Table 3.2 Properties and sources of variation affecting abundance profiles of COG function categories

Function category	**COGs in category**	**Copies cell^{-1}**	**Correlation to all COGs**	**Proportion of variance explained by:**			
				Batch	Site	Year	Crop
All	4594	0.50 (0.46-0.52)		0.287	0.092	0.056	0.086
V	115	0.50 (0.46-0.52)	0.93	0.363	0.052	0.056	0.078
T	182	1.09 (0.98-1.16)	0.92	0.306	0.090	0.068	0.102
C	274	0.70 (0.65-0.74)	0.91	0.246	0.114	0.056	0.110
G	246	0.57 (0.53-0.60)	0.96	0.328	0.084	0.039	0.092
P	232	0.51 (0.47-0.53)	0.95	0.275	0.114	0.047	0.093
Q	97	0.84 (0.76-0.92)	0.89	0.283	0.118	0.041	0.118

Function category interpretations: V, defensive mechanisms; T, signal transduction; C, energy production and conversion; G, carbohydrate transport and metabolism; P, Inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport, and catabolism. Copy cell^{-1} values are medians with 5th and 95th percentiles in parentheses. Correlations between distance matrices of function categories and all COGs were calculated with Mantel tests. Variance proportions were determined by permutational analysis of variance All values listed are significant at $P < 0.001$.

Table 3.3 Site-year effects on abundance profiles of COG function categories

Function category	Correlation to all COGs				Variance explained by Batch				Variance explained by Crop			
	ARL			KBS	ARL			KBS	ARL			KBS
	2010	2011	2012	2012	2010	2011	2012	2012	2010	2011	2012	2012
All	—	—	—	—	0.410	0.347	0.394	0.213	0.228	**0.116**	0.123	0.283
V	0.980	0.935	0.947	0.934	0.503	0.411	0.481	0.264	0.204	**0.094**	**0.102**	0.251
T	0.970	0.910	0.945	0.918	0.430	0.349	0.451	0.242	0.314	**0.118**	**0.124**	0.294
C	0.937	0.937	0.939	0.929	0.296	0.302	0.355	0.227	0.270	**0.136**	0.134	0.289
G	0.986	0.964	0.968	0.963	0.448	0.390	0.456	0.231	0.222	**0.115**	0.123	0.279
P	0.972	0.966	0.949	0.927	0.369	0.343	0.368	0.242	0.227	**0.127**	**0.118**	0.274
Q	0.947	0.894	0.926	0.842	0.309	0.309	0.404	0.149	0.325	**0.138**	0.151	0.340

Function categories are interpreted in Table 3.2. Variance proportions were calculated using permutational analysis of variance. Bolded values are not statistically significant ($P > 0.05$, 9999 permutations).

Table S3.1 Soil metagenome metadata

Site	Crop	Year	Block	Fertilized	JGI Project ID	Size (Gbp)	Sequences	Batch
ARL	Continuous corn	2010	A1	Yes	1020957	1.84	9.5E+06	B1
ARL	Continuous corn	2010	A2	Yes	1020960	2.02	1.0E+07	B1
ARL	Continuous corn	2010	A3	Yes	1020963	1.84	9.4E+06	B1
ARL	Continuous corn	2010	A4	Yes	1020966	1.86	9.5E+06	B1
ARL	Continuous corn	2010	A5	Yes	1020969	1.96	1.0E+07	B1
ARL	Restored prairie	2010	A1	No	1020972	1.87	9.6E+06	B1
ARL	Restored prairie	2010	A2	No	1020975	1.95	1.0E+07	B1
ARL	Restored prairie	2010	A3	No	1020978	1.79	9.1E+06	B1
ARL	Restored prairie	2010	A4	No	1020981	1.61	8.1E+06	B1
ARL	Restored prairie	2010	A5	No	1020984	1.90	9.6E+06	B1
ARL	Continuous corn	2011	A1	Yes	1020987	1.86	9.4E+06	B1
ARL	Continuous corn	2011	A2	Yes	1020990	1.86	9.5E+06	B1
ARL	Continuous corn	2011	A3	Yes	1020993	1.47	7.4E+06	B1
ARL	Continuous corn	2011	A4	Yes	1020996	1.39	7.0E+06	B1
ARL	Continuous corn	2011	A5	Yes	1020999	1.56	7.9E+06	B1
ARL	Restored prairie	2011	A1	No	1021002	1.33	6.6E+06	B1
ARL	Restored prairie	2011	A2	No	1021005	1.54	7.7E+06	B1
ARL	Restored prairie	2011	A3	No	1021008	1.57	7.9E+06	B1
ARL	Restored prairie	2011	A4	No	1021011	1.61	8.1E+06	B1
ARL	Restored prairie	2011	A5	No	1021014	1.51	7.7E+06	B1
ARL	Continuous corn	2012	A1	Yes	1021017	1.35	6.8E+06	B1
ARL	Continuous corn	2012	A2	Yes	1021020	1.34	6.7E+06	B1
ARL	Continuous corn	2012	A3	Yes	1021023	1.37	6.8E+06	B1
ARL	Continuous corn	2012	A4	Yes	1021026	1.48	7.4E+06	B1
ARL	Continuous corn	2012	A5	Yes	1021029	1.27	6.3E+06	B1
ARL	Restored prairie	2012	A1	No	1021032	1.53	7.7E+06	B1
ARL	Restored prairie	2012	A2	No	1021035	1.53	7.8E+06	B1
ARL	Restored prairie	2012	A3	No	1021038	1.23	6.3E+06	B1
ARL	Restored prairie	2012	A4	No	1021041	1.32	6.6E+06	B1
ARL	Restored prairie	2012	A5	No	1021044	1.47	7.3E+06	B1
ARL	Native grass mix	2010	A1	Yes	1032591	1.50	6.9E+06	B2
ARL	Old field	2010	A1	Yes	1032594	1.13	5.2E+06	B2
ARL	Miscanthus	2010	A1	Yes	1032597	1.06	4.8E+06	B2
ARL	Hybrid poplar	2010	A1	Yes	1032600	1.29	5.9E+06	B2
ARL	Switchgrass	2010	A1	Yes	1032603	1.35	6.2E+06	B2
ARL	Miscanthus	2010	A2	Yes	1032606	1.93	8.9E+06	B2
ARL	Switchgrass	2010	A2	Yes	1032609	1.39	6.5E+06	B2
ARL	Native grass mix	2010	A2	Yes	1032612	1.50	7.3E+06	B2
ARL	Old field	2010	A2	Yes	1032615	1.25	5.7E+06	B2
ARL	Hybrid poplar	2010	A2	Yes	1032618	1.76	8.1E+06	B2

Table S3.1 cont.

Site	Crop	Year	Block	Fertilized	JGI Project ID	Size (Gbp)	Sequences	Batch
ARL	Switchgrass	2010	A3	Yes	1032621	1.36	6.2E+06	B2
ARL	Miscanthus	2010	A3	Yes	1032624	1.78	8.1E+06	B2
ARL	Native grass mix	2010	A3	Yes	1032627	1.31	6.0E+06	B2
ARL	Hybrid poplar	2010	A3	Yes	1032630	1.60	7.4E+06	B2
ARL	Old field	2010	A3	Yes	1032633	1.55	7.1E+06	B2
ARL	Hybrid poplar	2010	A4	Yes	1032636	1.19	5.5E+06	B2
ARL	Old field	2010	A4	Yes	1032639	1.25	5.8E+06	B2
ARL	Native grass mix	2010	A4	Yes	1032642	1.22	5.6E+06	B2
ARL	Miscanthus	2010	A4	Yes	1032645	1.39	6.3E+06	B2
ARL	Switchgrass	2010	A4	Yes	1032648	1.34	6.1E+06	B2
ARL	Hybrid poplar	2010	A5	Yes	1032651	1.24	5.7E+06	B2
ARL	Native grass mix	2010	A5	Yes	1032654	2.93	1.3E+07	B2
ARL	Switchgrass	2010	A5	Yes	1032657	1.23	5.6E+06	B2
ARL	Old field	2010	A5	Yes	1032660	1.07	4.9E+06	B2
ARL	Miscanthus	2010	A5	Yes	1032663	1.33	6.1E+06	B2
ARL	Restored prairie	2011	A1	Yes	1032666	2.02	9.2E+06	B2
ARL	Miscanthus	2011	A1	Yes	1032669	1.51	6.8E+06	B2
ARL	Hybrid poplar	2011	A1	Yes	1032672	1.08	4.9E+06	B2
ARL	Old field	2011	A1	Yes	1032675	1.07	4.9E+06	B2
ARL	Switchgrass	2011	A1	No	1032681	1.13	5.2E+06	B2
ARL	Switchgrass	2011	A2	No	1032684	1.13	5.2E+06	B2
ARL	Restored prairie	2011	A2	Yes	1032687	1.44	6.7E+06	B2
ARL	Old field	2011	A2	Yes	1032690	1.48	6.8E+06	B2
ARL	Hybrid poplar	2011	A2	Yes	1032693	1.21	5.5E+06	B2
ARL	Switchgrass	2011	A3	Yes	1032696	0.90	4.2E+06	B2
ARL	Miscanthus	2011	A3	Yes	1032699	1.45	6.7E+06	B2
ARL	Restored prairie	2011	A3	Yes	1032702	1.06	5.0E+06	B2
ARL	Hybrid poplar	2011	A3	Yes	1032705	1.06	4.8E+06	B2
ARL	Old field	2011	A3	Yes	1032708	1.29	5.9E+06	B2
ARL	Hybrid poplar	2011	A4	Yes	1032711	1.25	5.8E+06	B2
ARL	Old field	2011	A4	Yes	1032714	1.36	6.3E+06	B2
ARL	Restored prairie	2011	A4	Yes	1032717	1.28	5.9E+06	B2
ARL	Miscanthus	2011	A4	Yes	1032720	1.32	6.1E+06	B2
ARL	Switchgrass	2011	A4	Yes	1032723	0.87	4.1E+06	B2
ARL	Hybrid poplar	2011	A5	Yes	1032726	1.36	6.3E+06	B2
ARL	Restored prairie	2011	A5	Yes	1032729	1.32	6.1E+06	B2
ARL	Switchgrass	2011	A5	Yes	1032732	1.36	6.3E+06	B2
ARL	Old field	2011	A5	Yes	1032735	1.53	7.0E+06	B2
ARL	Miscanthus	2011	A5	Yes	1032738	1.33	6.2E+06	B2
ARL	Native grass mix	2012	A1	Yes	1032741	1.22	5.6E+06	B2

Table S3.1 cont.

Site	Crop	Year	Block	Fertilized	JGI Project ID	Size (Gbp)	Sequences	Batch
ARL	Restored prairie	2012	A1	Yes	1032744	1.30	6.1E+06	B2
ARL	Old field	2012	A1	Yes	1032747	1.55	7.3E+06	B2
ARL	Miscanthus	2012	A1	Yes	1032750	1.10	5.1E+06	B2
ARL	Hybrid poplar	2012	A1	Yes	1032753	1.18	5.3E+06	B2
ARL	Switchgrass	2012	A1	Yes	1032756	1.53	6.9E+06	B2
ARL	Switchgrass	2012	A1	No	1032759	0.63	2.9E+06	B2
ARL	Miscanthus	2012	A2	Yes	1032762	1.36	6.2E+06	B2
ARL	Switchgrass	2012	A2	No	1032765	1.47	6.8E+06	B2
ARL	Native grass mix	2012	A2	Yes	1032768	1.32	6.0E+06	B2
ARL	Restored prairie	2012	A2	Yes	1032771	1.18	5.4E+06	B2
ARL	Old field	2012	A2	Yes	1032774	1.36	6.2E+06	B2
ARL	Hybrid poplar	2012	A2	Yes	1032777	1.33	6.0E+06	B2
ARL	Miscanthus	2012	A3	Yes	1032780	1.26	5.7E+06	B2
ARL	Restored prairie	2012	A3	Yes	1032783	1.36	6.2E+06	B2
ARL	Native grass mix	2012	A3	Yes	1032786	1.02	4.7E+06	B2
ARL	Hybrid poplar	2012	A3	Yes	1032789	1.59	7.2E+06	B2
ARL	Old field	2012	A3	Yes	1032792	1.34	6.1E+06	B2
ARL	Hybrid poplar	2012	A4	Yes	1032795	1.16	5.3E+06	B2
ARL	Old field	2012	A4	Yes	1032798	0.97	4.5E+06	B2
ARL	Native grass mix	2012	A4	Yes	1032801	1.16	5.4E+06	B2
ARL	Restored prairie	2012	A4	Yes	1032804	1.58	7.3E+06	B2
ARL	Miscanthus	2012	A4	Yes	1032807	1.22	5.5E+06	B2
ARL	Switchgrass	2012	A4	Yes	1032810	1.25	5.6E+06	B2
ARL	Switchgrass	2012	A4	No	1032813	1.32	6.0E+06	B2
ARL	Hybrid poplar	2012	A5	Yes	1032816	1.45	6.5E+06	B2
ARL	Restored prairie	2012	A5	Yes	1032819	1.30	6.0E+06	B2
ARL	Native grass mix	2012	A5	Yes	1032822	1.17	5.3E+06	B2
ARL	Switchgrass	2012	A5	Yes	1032825	2.04	9.4E+06	B2
ARL	Switchgrass	2012	A5	No	1032828	1.89	8.7E+06	B2
ARL	Old field	2012	A5	Yes	1032831	1.94	8.8E+06	B2
ARL	Miscanthus	2012	A5	Yes	1032834	1.76	8.1E+06	B2
KBS	Continuous corn	2012	K2	Yes	1032837	2.08	9.5E+06	B2
KBS	Continuous corn	2012	K4	Yes	1032840	2.00	9.2E+06	B2
KBS	Switchgrass	2012	K2	Yes	1032843	1.66	7.7E+06	B2
KBS	Switchgrass	2012	K3	Yes	1032846	1.69	7.8E+06	B2
KBS	Switchgrass	2012	K4	Yes	1032849	1.20	5.6E+06	B2
KBS	Miscanthus	2012	K3	Yes	1032852	1.40	6.4E+06	B2
KBS	Native grass mix	2012	K2	Yes	1032855	1.13	5.2E+06	B2
KBS	Native grass mix	2012	K3	Yes	1032858	1.08	4.9E+06	B2
KBS	Native grass mix	2012	K4	Yes	1032861	1.06	4.9E+06	B2

Table S3.1 cont.

Site	Crop	Year	Block	Fertilized	JGI Project ID	Size (Gbp)	Sequences	Batch
KBS	Native grass mix	2012	K5	Yes	1032864	1.08	5.0E+06	B2
KBS	Hybrid poplar	2012	K2	Yes	1032867	0.88	4.3E+06	B2
KBS	Hybrid poplar	2012	K3	Yes	1032870	0.94	4.5E+06	B2
KBS	Hybrid poplar	2012	K4	Yes	1032873	0.92	4.4E+06	B2
KBS	Hybrid poplar	2012	K5	Yes	1032876	0.78	3.7E+06	B2
KBS	Old field	2012	K1	Yes	1032879	0.89	4.3E+06	B2
KBS	Old field	2012	K2	Yes	1032882	0.65	3.1E+06	B2
KBS	Old field	2012	K3	Yes	1032885	0.84	4.0E+06	B2
KBS	Old field	2012	K4	Yes	1032888	0.86	4.1E+06	B2
KBS	Restored prairie	2012	K1	No	1032891	0.73	3.5E+06	B2
KBS	Restored prairie	2012	K5	No	1032894	0.80	3.8E+06	B2
KBS	Continuous corn	2012	K1	Yes	1040325	1.60	7.3E+06	B3
KBS	Switchgrass	2012	K1	Yes	1040328	1.58	7.3E+06	B3
KBS	Miscanthus	2012	K1	Yes	1040331	1.63	7.4E+06	B3
KBS	Native grass mix	2012	K1	Yes	1040334	1.47	6.6E+06	B3
KBS	Hybrid poplar	2012	K1	Yes	1040337	1.47	6.7E+06	B3
KBS	**Continuous corn**	**2012**	**K1**	**Yes**	**1040340**	**1.38**	**6.2E+06**	**B3**
KBS	Miscanthus	2012	K2	Yes	1040343	1.38	6.3E+06	B3
KBS	Restored prairie	2012	K2	No	1040346	1.66	7.5E+06	B3
KBS	Continuous corn	2012	K3	Yes	1040349	1.46	6.6E+06	B3
KBS	Miscanthus	2012	K4	Yes	1040352	1.56	7.1E+06	B3
KBS	Restored prairie	2012	K4	No	1040355	2.06	9.3E+06	B3
KBS	Continuous corn	2012	K5	Yes	1040358	1.47	6.6E+06	B3
KBS	Switchgrass	2012	K5	Yes	1040361	2.19	9.9E+06	B3
KBS	Miscanthus	2012	K5	Yes	1040364	1.47	6.7E+06	B3
KBS	Old field	2012	K5	Yes	1040367	1.56	7.1E+06	B3
ARL	**Native grass mix**	**2010**	**A5**	**Yes**	**1040370**	**1.14**	**5.1E+06**	**B3**
ARL	Switchgrass	2011	A2	Yes	1040373	1.58	7.1E+06	B3
ARL	Miscanthus	2011	A2	Yes	1040376	1.40	6.3E+06	B3
ARL	**Switchgrass**	**2011**	**A3**	**Yes**	**1040379**	**1.69**	**7.7E+06**	**B3**
ARL	Switchgrass	2011	A3	No	1040382	2.21	1.0E+07	B3
ARL	Switchgrass	2011	A4	No	1040385	1.86	8.5E+06	B3
ARL	**Restored prairie**	**2011**	**A4**	**Yes**	**1040388**	**1.59**	**7.3E+06**	**B3**
ARL	**Switchgrass**	**2011**	**A5**	**Yes**	**1040391**	**1.74**	**7.9E+06**	**B3**
ARL	**Miscanthus**	**2011**	**A5**	**Yes**	**1040394**	**1.36**	**6.2E+06**	**B3**
ARL	**Hybrid poplar**	**2011**	**A5**	**Yes**	**1040397**	**1.50**	**6.8E+06**	**B3**
ARL	**Old field**	**2011**	**A5**	**Yes**	**1040400**	**1.37**	**6.2E+06**	**B3**
ARL	**Switchgrass**	**2012**	**A1**	**Yes**	**1040403**	**2.06**	**9.3E+06**	**B3**
ARL	**Old field**	**2012**	**A1**	**Yes**	**1040406**	**1.91**	**8.5E+06**	**B3**
ARL	Switchgrass	2012	A2	Yes	1040409	1.66	7.6E+06	B3

Table S3.1 cont.

Site	Crop	Year	Block	Fertilized	JGI Project ID	Size (Gbp)	Sequences	Batch
ARL	**Switchgrass**	**2012**	**A2**	**No**	**1040412**	**1.71**	**7.7E+06**	**B3**
ARL	**Miscanthus**	**2012**	**A2**	**Yes**	**1040415**	**1.61**	**7.4E+06**	**B3**
ARL	Switchgrass	2012	A3	Yes	1040418	1.36	6.2E+06	B3
ARL	Switchgrass	2012	A3	No	1040421	2.03	9.3E+06	B3

Boldface indicates technical replicates.

Table S3.2 Metadata for clusters of orthologous groups (COG) annotations
NOTE: This table contains 4873 records and would cover over 140 printed pages. Only the first 24 records are shown here

COG	Function category	COG name	COG model length (bp)	Single-copy	Outdated
COG0001	H	Glutamate-1-semialdehyde aminotransferase	1296	FALSE	FALSE
COG0002	E	N-acetyl-gamma-glutamylphosphate reductase	1047	FALSE	FALSE
COG0003	P	Anion-transporting ATPase, ArsA/GET3 family	966	FALSE	FALSE
COG0004	P	Ammonia channel protein AmtB	1227	FALSE	FALSE
COG0005	F	Purine nucleoside phosphorylase	786	FALSE	FALSE
COG0006	E	Xaa-Pro aminopeptidase	1152	FALSE	FALSE
COG0007	H	Uroporphyrinogen-III methylase (siroheme synthase)	732	FALSE	FALSE
COG0008	J	Glutamyl- or glutaminyl-tRNA synthetase	1416	FALSE	FALSE
COG0009	J	tRNA A37 threonylcarbamoyladenosine synthetase subunit TsaC/SUA5/YrdC	633	FALSE	FALSE
COG0010	E	Arginase family enzyme	915	FALSE	FALSE
COG0011	S	Uncharacterized conserved protein YqgV, UPF0045/DUF77 family	300	FALSE	FALSE
COG0012	J	Ribosome-binding ATPase YchF, GTP1/OBG family	1116	FALSE	FALSE
COG0013	J	Alanyl-tRNA synthetase	2637	FALSE	FALSE
COG0014	E	Gamma-glutamyl phosphate reductase	1251	FALSE	FALSE
COG0015	F	Adenylosuccinate lyase	1314	FALSE	FALSE
COG0016	J	Phenylalanyl-tRNA synthetase alpha subunit	1005	TRUE	FALSE
COG0017	J	Aspartyl/asparaginyl-tRNA synthetase	1305	FALSE	FALSE
COG0018	J	Arginyl-tRNA synthetase	1731	FALSE	FALSE
COG0019	E	Diaminopimelate decarboxylase	1182	FALSE	FALSE
COG0020	I	Undecaprenyl pyrophosphate synthase	735	FALSE	FALSE
COG0021	G	Transketolase	1989	FALSE	FALSE
COG0022	C	Pyruvate/2-oxoglutarate/acetoin dehydrogenase complex, dehydrogenase (E1) component	972	FALSE	FALSE
COG0023	J	Translation initiation factor 1 (eIF-1/SUI1)	312	FALSE	FALSE
COG0024	J	Methionine aminopeptidase	765	FALSE	FALSE

COGs fitting into multiple function categories are counted in all of them. Single-copy denotes the 37 universal single-copy genes used for normalization across samples. Outdated indicates COGs removed in the 2014 update of the database; these were not included in the study.

Table S3.3 Denitrification pathway COGs

Process	COG	COG name	Function category	Gene symbol	Ref
Nitrate reduction	COG2180	Nitrate reductase assembly protein NarJ, required for insertion of molybdenum cofactor	CPO	narJ/narW	1
	COG2181	Nitrate reductase gamma subunit	CP	narI/narV	1
	COG2132	Multicopper oxidase with three cupredoxin domains (includes cell division protein FtsP and spore coat protein CotA)	DPM		1
	COG5013	Nitrate reductase alpha subunit	CP	narG/narZ/nxrA	1,2
	COG1140	Nitrate reductase beta subunit	CP	narH/narY/nxrB	1,2
	COG3043	Nitrate reductase cytochrome c-type subunit	CP	napB	1
	COG3005	Tetraheme cytochrome c subunit of nitrate or TMAO reductase	C		1,2
	COG2223	Nitrate/nitrite transporter NarK	P	narK	2
Nitrite reduction	COG1251	NAD(P)H-nitrite reductase, large subunit	C		1
	COG3303	Formate-dependent nitrite reductase, periplasmic cytochrome c552 subunit	P		1
	COG2146	Ferredoxin subunit of nitrite reductase or a ring-hydroxylating dioxygenase	PQ		1
	COG3301	Formate-dependent nitrite reductase, membrane component NrfD	P		1
Nitric oxide reduction	COG3256	Nitric oxide reductase large subunit	P	norB	1
	COG4548	Nitric oxide reductase activation protein	P		1
Nitrous oxide reduction	COG4263	Nitrous oxide reductase	P	nosZ	1,2
	COG3420	Nitrous oxidase accessory protein NosD, contains tandem CASH domains	P	nosD	2

Data from the 2014 update to the COG database. Function categories: C, Energy production and conversion; D, cell cycle control, cell division, chromosome partitioning; M, Cell wall/membrange/envelope biogenesis; O, Posttranslational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolite biosynthesis, transport and catabolism. Gene symbols obtained from the KEGG database. References: 1) S. He *et al.*, *MBio*. **6**, e00066–15 (2015); 2) Z. Wang *et al.*, *PLoS One*. **9**, e113603 (2014).

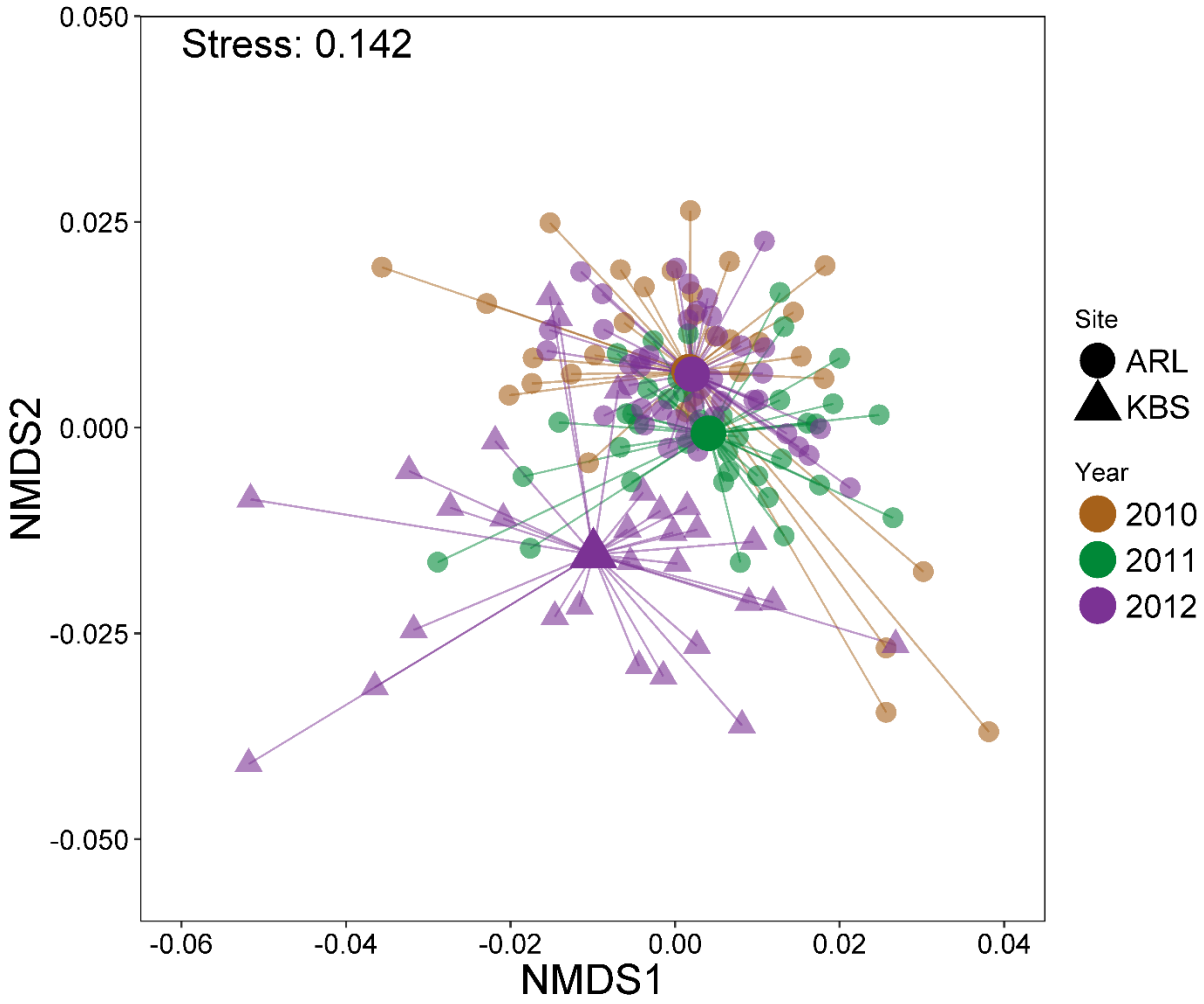

Figure 3.1 Site and year effects on soil microbial functional gene profiles. Intersample distances are corrected for metagenome size and sequencing batch effects. Large opaque symbols indicate mean axis scores for each site-year.

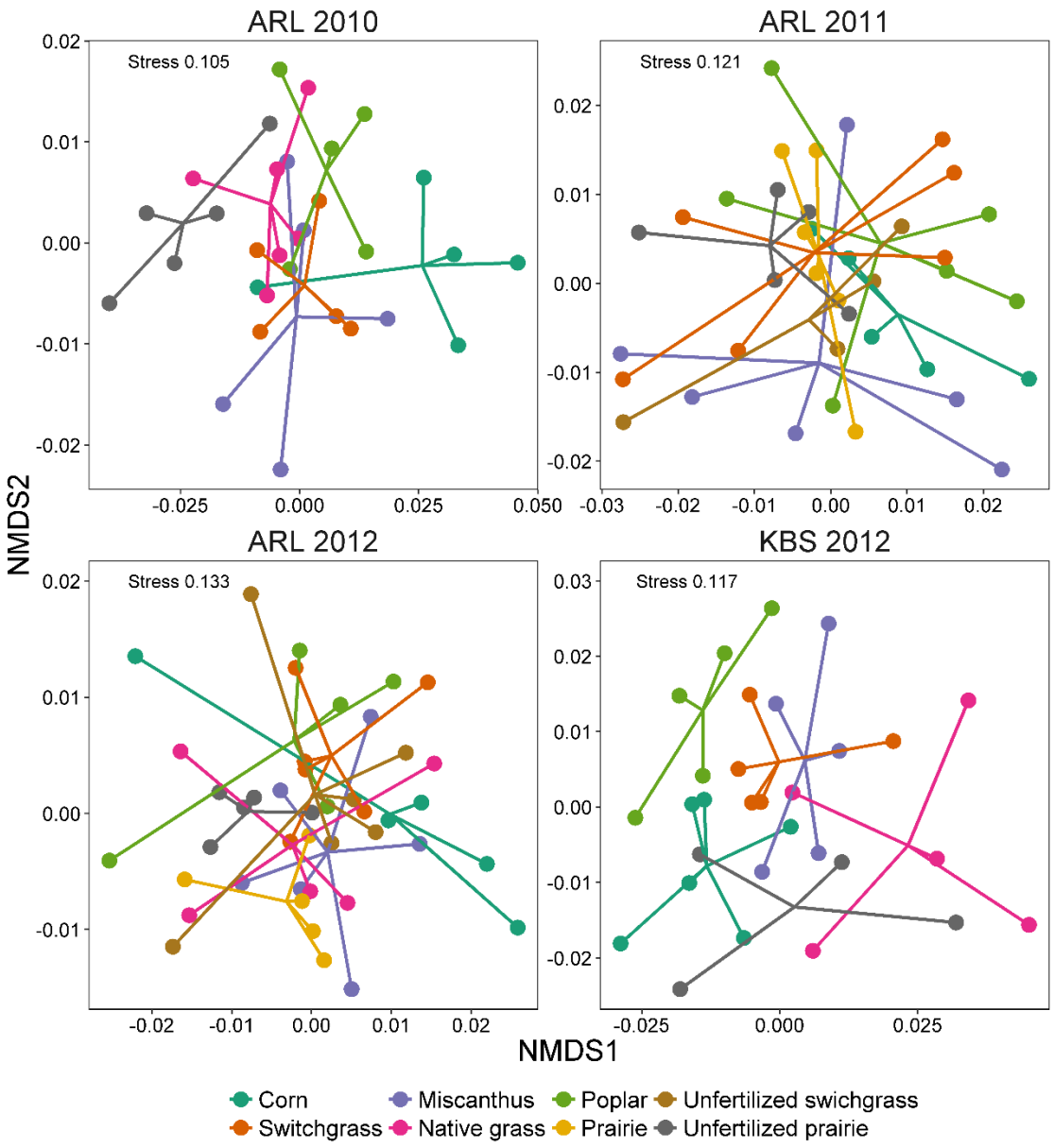

Figure 3.2 Cropping system effects on soil microbial community functional gene profiles. Ordinations conducted separately by site-year. Intersample distances are corrected for effects of metagenome size and sequencing batch.

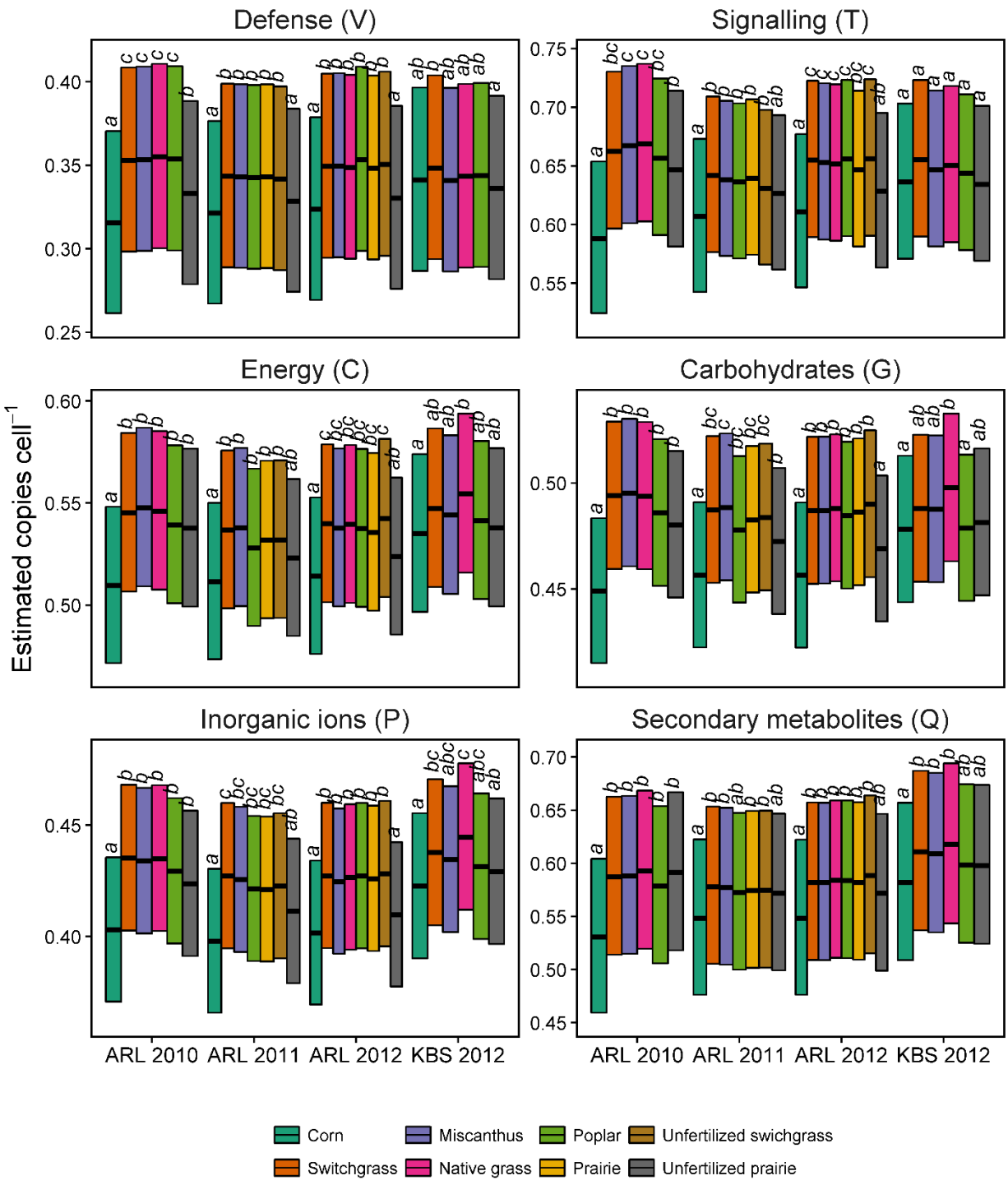

Figure 3.3 Estimated copy numbers for COGs by select function categories. Values were inverse hyperbolic sine (IHS) transformed for analysis, with means ± s.e. back-transformed for display.

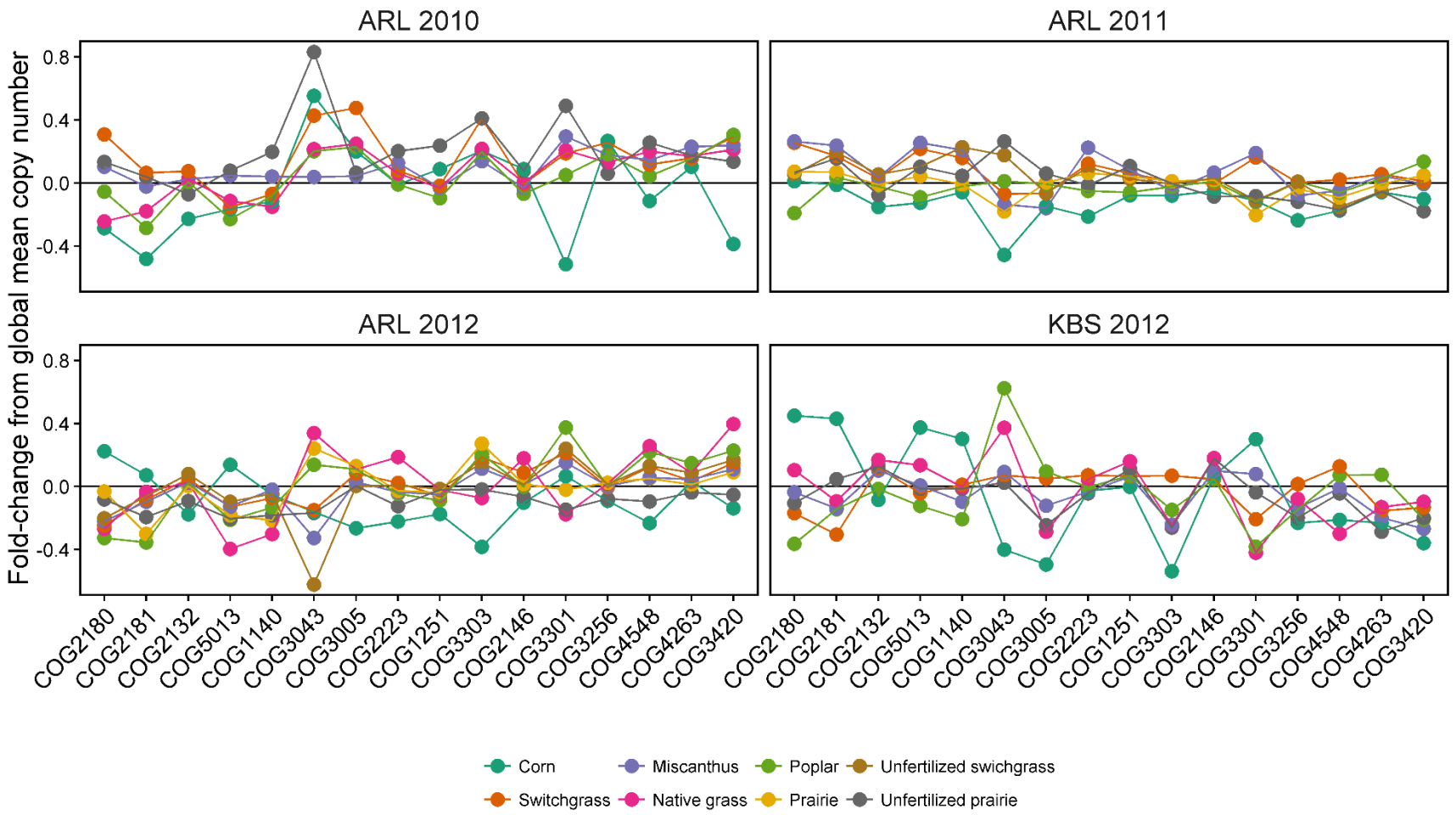

Figure 3.4 Fold change in abundance relative to global means for denitrification-pathway COGs. Fold change calculated as $\log_2$ differences in estimated copy numbers between system-site-year and global means.

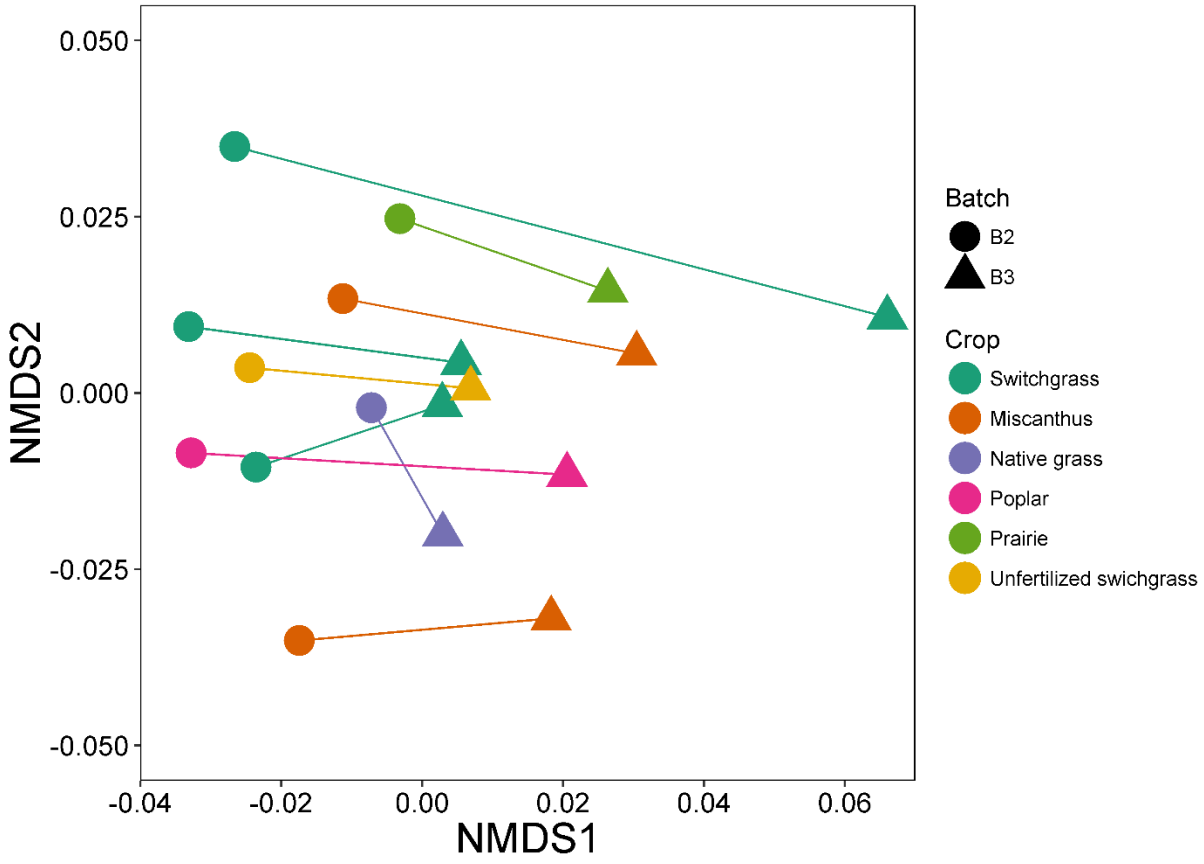

Figure S3.1 Sequencing batch effects on functional gene profiles. Line segments connect technical replicates sequenced in separate batches.

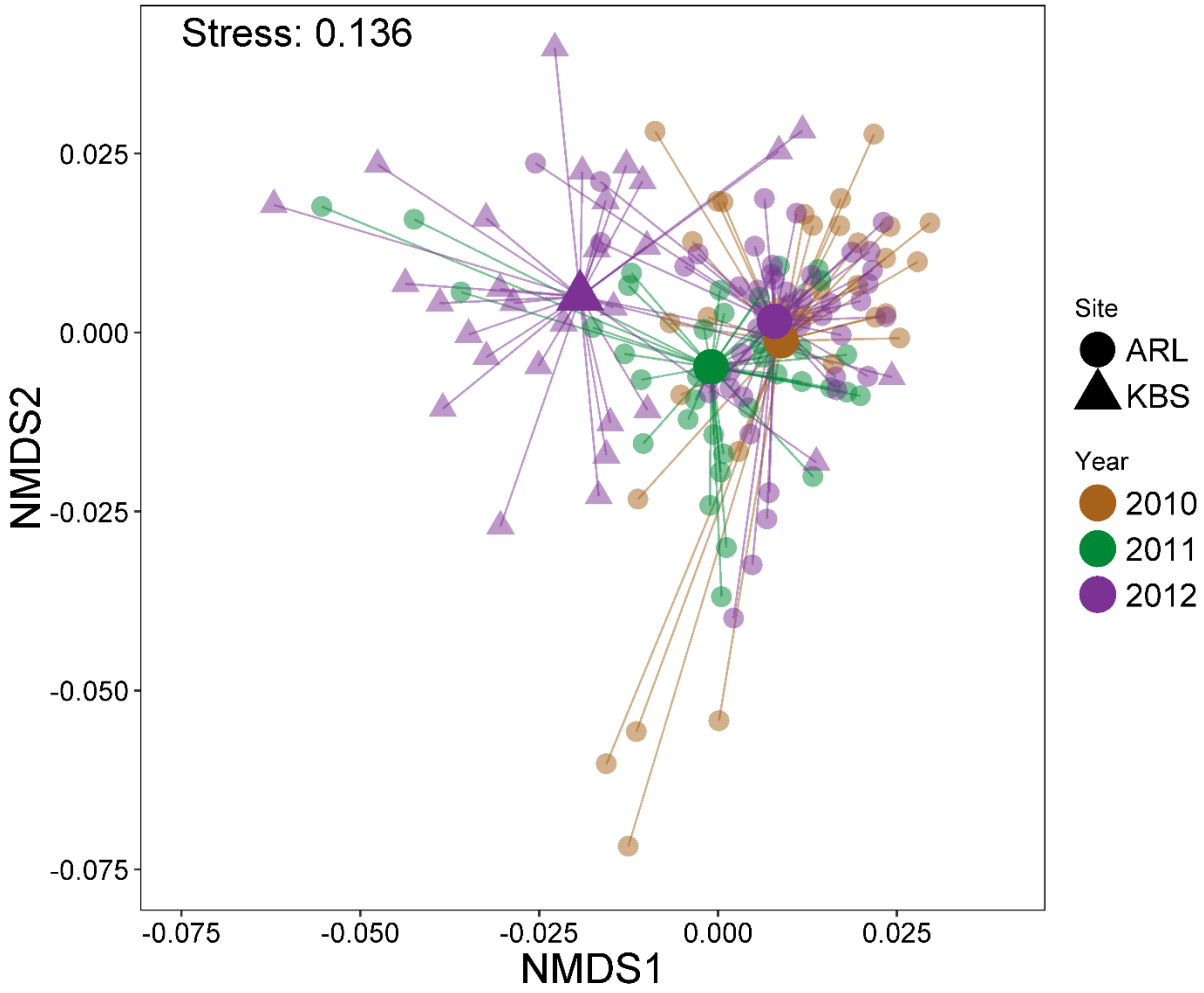

Figure S3.2 Site and year effects on soil microbial functional gene profiles without correction for metagenome size and sequencing batch effects on intersample distances. Large opaque symbols indicate mean axis scores for each site-year.

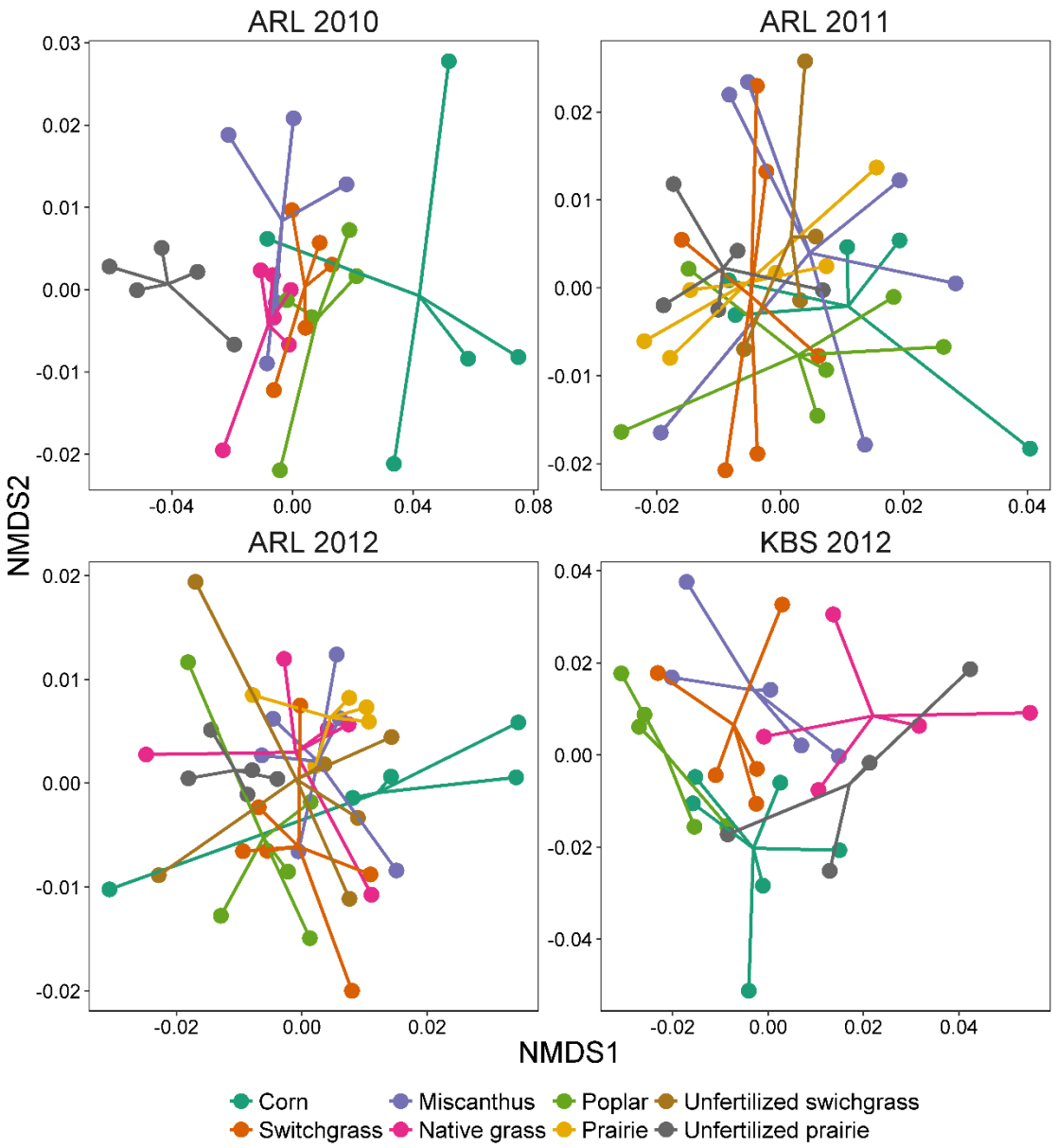

Figure S3.3 Cropping system effects on abundances of genes involved in cell signaling (COG function category V).

CHAPTER 4

Effects of environmental constraints on soil N_2O fluxes differ among potential bioenergy feedstock cropping systems

David S. Duncan [a], Lawrence G. Oates [a], Ilya Gelfand [b], Neville Millar [b], G. Philip Robertson [b], Randall D. Jackson [a]

[a] Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA; DOE-Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

[b] W. K. Kellogg Biological Station & DOE-Great Lakes Bioenergy Research Center, Michigan State University, Hickory Corners, MI 49060; Department of Plant, Soil, and Microbial Sciences, Michigan State University, East Lansing, MI, 48824

Abstract

Nitrous oxide (N_2O) is a potent greenhouse gas and major component of the net global warming potential of bioenergy feedstock production systems. Numerous environmental factors influence soil N_2O production, with effects that may vary among cropping systems. We investigated how soil temperature, water filled pore space (WFPS), and concentrations of soil nitrate and ammonium constrained N_2O production from a range of cropping systems including conventionally-managed annual grain crops, perennial warm-season grasses, hybrid poplar, and polycultures of tallgrass prairie species over six growing seasons at two field sites. We observed higher N_2O fluxes and concentrations of soil nitrate in the annual cropping systems, and in most fertilized perennial systems relative to their nonfertilized counterparts. Measurements of ammonium concentrations, soil temperature, and WFPS within a site had similar distributions across cropping systems. We used quantile regression to evaluate whether levels of the four environmental factors limited the highest fluxes observed. At both sites, all environmental factors were significantly and positively related to the 95th percentile of N_2O fluxes. Both our models and observations found high fluxes could occur at low soil moistures, sub-zero soil temperatures, and minimal nitrogen concentrations. Differences in cropping-system specific quantile regression coefficients did not systematically reflect broad management or other cropping system differences. Overall, environmental conditions appeared to define bounds for N_2O fluxes, but these bounds do not appear to drive cropping system-level differences.

4.1 Introduction

Nitrous oxide (N_2O) is one of the major contributors to global radiative forcing (Forster et al., 2007; Robertson et al., 2000) and is currently the single most important ozone-depleting substance (Ravishankara et al., 2009). In the United States, approximately 75% of N_2O emissions come from fertility management of agricultural soils (U.S. Environmental Protection Agency, 2014). N_2O emissions can counterbalance net carbon balance benefits of fossil fuel displacement for many agricultural systems (Crutzen et al., 2008), making the management and mitigation of N_2O production a major aspect of long-term cropping system sustainability. This consideration is particularly important for the development of bioenergy feedstock production systems from perennial crops and on marginal lands.

By broadening the range of economically viable cropping systems, bioenergy feedstock production offers many ecological advantages (Robertson et al., 2008). Bioenergy feedstock cropping systems can increase diversity in agricultural regions, improving landscape-level provision of ecosystem services (Werling et al., 2014). Some of these systems may function on agronomically and ecologically marginal lands (Gelfand et al., 2013), which could lead to long-term improvement in their soil carbon content and erosion potential (Blanco-Canqui, 2010). While agroecosystems incorporating perenniality and plant species diversity may exhibit more efficient nutrient cycling (Hooper et al., 2005; Hooper and Vitousek, 1998), nitrogen fertilization will likely factor into management of bioenergy feedstock cropping systems (Montross et al., 2013). We have limited knowledge of how perennial and polyculture bioenergy feedstock cropping systems might function under agronomic management (Stehfest and Bouwman, 2006; Trybula et al., 2015), leading to considerable uncertainty about how biotic and abiotic factors in these systems will interact to influence N_2O production dynamics.

Factors influencing N_2O production in soil have received considerable study over the years. We know that most soil N_2O production results from the microbial processes of nitrification (oxidation of ammonia [NH_3] to nitrate [NO_3^-]) and denitrification (reduction of NO_3^- to N_2), both of which can produce N_2O as a side-product (Bremner, 1997; Butterbach-Bahl et al., 2013). Of the two processes, denitrification frequently accounts for the majority of N_2O production, particularly in soils that are moist,

finely-textured, or otherwise oxygen-limited (Mathieu et al., 2006). Despite the microbial nature of these processes, under most conditions abiotic environmental factors typically determine N_2O fluxes (Wallenstein et al., 2006). These factors include concentrations of substrates (NH_3, NO_3^-, and labile carbon), soil oxygen availability (typically driven by soil water-filled pore space [WFPS]), soil temperature, and pH (Wallenstein et al., 2006). With multiple factors influencing N_2O fluxes, standard regression techniques rarely generate clear, consistent relationships between specific factors and flux. Alternative methods, such as quantile regression (Cade and Noon, 2003), may serve to determine the extent to which a particular variable imposes an upper bound on observed N_2O fluxes. Further complicating matters, the effects of these factors may also vary by cropping systems and soils (Dechow and Freibauer, 2011; Lehuger et al., 2009). For instance, soil texture influences the relationship between soil moisture and nitrification rates (Garrido et al., 2002). Similarly, the effect of nitrogen fertilizer application on N_2O emissions differs among cropping systems, although this may reflect immobilization of exogenous nitrogen, in addition to responses to substrate concentrations (Duran et al., 2016; Stehfest and Bouwman, 2006). Efforts to model N_2O fluxes from novel cropping systems such as perennial, polycultural bioenergy feedstocks, should consider cropping system-specific responses to environmental drivers of N_2O production.

Our prior study of N_2O dynamics in this system identified substantial variation in N_2O fluxes among cropping systems and suggested that environmental factors might influence N_2O fluxes in a system-specific manner (Oates et al., 2016). In the present study, we employed a longer data record to examine these relationships in greater detail. Rather than treating environmental factors as direct predictors of N_2O fluxes, as we did in our earlier work and as other studies have done (Castellano et al., 2010; Dechow and Freibauer, 2011; Liu et al., 2013), we interpreted these factors as constraints on maximum N_2O production. This approach is highly appropriate for systems where multiple factors jointly contribute to variability in the property of interest (Cade and Noon, 2003). Through this work, we tested two hypotheses: 1) soil moisture, soil temperature, and concentrations of inorganic nitrogen (NH_4^+, NO_3^-)

independently delineate upper bounds of N_2O production and 2) the values of those upper bounds differ among cropping systems.

4.2 Methods

4.2.1 Experimental design and study sites

We conducted this study on the DOE-Great Lakes Bioenergy Research Center's Bioenergy Cropping Systems Experiment (BCSE), an agronomic trial situated at the Michigan State University W. K. Kellogg Biological Station (KBS, 42 23'47" N, 85 22'26"W, 288 m.a.s.l) and the University of Wisconsin-Madison Arlington Agricultural Research Station (ARL, 43 17'45" N, 89 22'48" W, 315 m.a.s.l. The BCSE consisted of ten treatments. From 2009 to 2011, three treatments consisted of the phases of a no-tillage corn (*Zea mays* L.)-soybean (*Glycine max* L.)-canola (*Brassica napus* L.) rotation; from 2012 to 2014, these treatments consisted of no-tillage continuous corn and the two phases of a corn-soybean rotation, all of which were grown with a rye (*Secale cereale* L.) and Austrian winter pea (*Pisumsativum* L.) cover crop (Table S4.1). The remaining treatments were in place throughout the measurement period and consisted of continuous no-illage corn, monocultures of switchgrass (*Panicum virgatum* L.), miscanthus (*Miscanthus* × *giganteus*), and hybrid poplar (*Populus nigra* × *Populus maximowiczii*), a 5-species native grass mix, an old field recruited from the pre-existing seedbank, and an 18-species restored tallgrass prairie. Species and variety information are presented in Table S1 of Oates et al. (2016).

All treatments were planted in 27 × 43-m plots in a five-replicate randomized complete block design and managed with field-scale equipment. Annual grain systems were managed following recommendations from the university extension programs from their respective states. The poplar system was fertilized in 2010 (210 kg N ha^{-1} as 34-0-0 granular ammonium nitrate) and harvested by coppicing during the 2013-2014 winter. Subplots (10 × 43-m) were established in all other systems to test effects of nitrogen fertilization. The restored prairie subplots and main plots of all other systems received annual spring nitrogen fertilization (56 kg N ha^{-1} as 34-0-0 granular ammonium nitrate), while the main restore prairie plot and subplots of all other systems were not fertilized. Corn systems were harvested for grain and stover separately, with near-complete stover removal. Perennial systems were harvested following the

first frost event in the fall, with a residual stubble height of 10 cm. Poplar systems were harvested once, in 2013, and allowed to coppice. Fertilization dates for all systems are given in Table S4.2, while full details on agronomic management are presented in Sanford et al. (2016).

Soils at KBS were primarily Kalamazoo loam (Fine-Loamy, Mixed, Semiactive, Mesic Typic Hapludalfs). Mean annual temperature from 1981 to 2010 was 9.9 °C and mean annual precipitation was 1027 mm (MSCO, 2013). Prior to BCSE establishment, the field was planted to alfalfa (*Medicago sativa* L.). The switchgrass, native grass mix, and restored prairie treatments at KBS suffered seed loss following flooding in 2008 and were reseeded in 2009. Soils at ARL were predominantly Plano silt-loam (Fine-Silty, Mixed, Superactive, Mesic Typic Arguidolls). Mean annual temperature from 1981 to 2010 was 6.8 °C and mean annual precipitation was 869 mm (NWS, 2013). Pre-BCSE land use differed among blocks: the corn phase of a corn-soybean rotation (Blocks A1-A3) or an alfalfa-orchardgrass (*Dactylis glomerata* L.) hay mixture (Blocks A4-A5). Following severe stand loss during the 2008-2009 winter, miscanthus was replanted at ARL in 2010

4.2.2 Data generation

All field sampling procedures are described in detail in Oates et al. (2016). Systems were sampled biweekly during the growing season, with additional sampling following fertilization and precipitation events, and at reduced frequency during winter, particularly earlier in the study. Static chambers were used to estimate trace gas emissions. Chambers were cylindrical (28.5 cm diameter, ~17 cm headspace, ~10 L volume) and inserted to a depth of ~5 cm. Chamber lids were fitted with a septum for gas extraction and a 2-mm diameter vent tube for pressure equilibration. Headspace gas samples (30 mL) were collected immediately upon chamber closure and at 3×20 min intervals. Samples were placed in glass 5.9-mL Exetainer vials (Labco Limited, Buckinghamshire, UK), using 20 mL to flush the vial before over-pressurizing with the remaining 10 mL. Following gas chromatography, CO_2 concentration was detected using an infrared gas analyzer (IRGA, LiCor 820, Lincoln, NE, USA) and N_2O concentration was detected using an electron capture detector (micro-ECD, Agilent 7890A GC System, Santa Clara, CA, USA).

Prior to estimation of N_2O fluxes, CO_2 accumulation curves were visually inspected for outliers indicating compromised vial integrity or other mechanical errors. In samples with four vials, nonlinearity of fluxes was evaluated using the "HMR" package (v0.3.1, Pedersen, 2015) in the R statistical environment (v3.2.5, R Core Team, 2016). Following this classification, all data were again visually inspected with an emphasis on identifying outliers in N_2O concentrations, particularly those that might drive a nonlinear fit. Nonlineaer flux estimates from the "HMR" function were used for samples that passed this secondary inspection without any data removal and whose nonlinear estimate was outside the 95% confidence interval for the linear flux estimate. For all other samples, linear flux estimates were used.

From 2010 onward, soil cores (3.7 cm diameter, 15 cm depth at ARL, 25 cm depth at KBS) were collected concurrently with trace gas sampling. Inorganic soil nitrogen was extracted from a 10 g field-moist subsample using 2 *M* KCl following Robertson et al. (1999). Ammonium and nitrate concentrations were determined using a Flow Solution 3100 segmented flow injection analyzer (OI Analytical, College Station, TX, USA), using USEPA methods 27200110 and 27190110 respectively. Nitrate values ≤ 0 were assigned a value of 0.05 µg-N g^{-1} soil, which corresponds to the instrument detection limit.

Soil temperature was measured at the time of trace gas sampling using a 15-cm temperature probe (Checktemp 1C, Hanna Instruments, Smithfield, RI, US). Soil moisture was measured at KBS by determining gravimetric water content (GWC) for the soil nitrogen samples. At ARL, moisture was measured as volumetric water content (VWC) within 1 m of the static chamber using a time domain reflectometer with 20-cm rods (FieldScout 300, Spectrum Technologies, Plainfield, IL, US). Bulk density was measured for all plots in 2008 and 2013. We calculated mean bulk density values groups of similarly-managed systems: annual grain crops, poplar, and all other systems. At ARL, blocks were placed into three groups based on their topographic position and soil properties: A1 and A3, A4 and A5, and A2; at KBS all blocks were analyzed together. Water-filled pore space (WFPS) was calculated from bulk density (*Bd*) and soil particle density (*Pd*, assumed to be a constant 2.65 g cm^{-3}):

$$WFPS = VWC \times \left(1 - \frac{Bd}{Pd}\right) = GWC \times Bd \times \left(1 - \frac{Bd}{Pd}\right)$$

4.2.3 Data analysis

N_2O fluxes as well as soil ammonium and nitrate values were inverse hyperbolic sine (IHS) transformed prior to analysis. IHS transformation resembles a log transformation at large input values, but is defined at 0 and allows negative values. We used this transformation to handle negative N_2O fluxes, which are periodically observed (Molodovskaya et al., 2012), and to avoid amplifying measurement errors for values close to the detection limit of our instruments.

All analyses were conducted in the R statistical environment. Graphics were generated using the *ggplot2* package (Wickham, 2009) using default functions to generate boxplot quantiles and LOESS fits. Quantile regression used the "rq" function in the "quantreg" package (Koenker, 2016). We used $\tau = 0.95$, which approximately corresponded to the 95th percentile of the data, and estimated standard errors using the kernel method. Significance of factors was evaluated using model comparison at the $P < 0.05$ significance level.

4.3 Results

Our dataset consisted of 10,572 individual N_2O flux measurements from 2 sites over a 6-year period. We recorded a median of 16 observations per plot per year (range 5 to 22). For a majority of observations, we also had accompanying measurements of soil temperature, water filled pore space (WFPS), and ammonium and nitrate concentrations (Table 4.1). Soil nitrogen data had the sparsest coverage, with no data in 2009 and lower frequency of data collection starting in 2013. Despite this, we generated 4,120 observations with full environmental data.

4.3.1 Distributions of N_2O fluxes and potential environmental drivers, by site and system

Annual cropping systems (described in Table S4.1) generated nearly identical distributions of N_2O fluxes and environmental variables (Fig. S4.1), and were thus grouped for analysis. Soil nitrogen concentrations and N_2O fluxes varied by site, cropping system, and fertilization (Fig. 4.1). In contrast, WFPS differed

only among sites while variability in soil temperature among systems was smaller than the range of interannual variability (Fig. S4.2).

Within a cropping system, N_2O flux observations varied over multiple orders of magnitude (Fig. 4.1). Fluxes above 100 g N_2O-N ha^{-1} day^{-1} occurred in nearly all systems, but were more prevalent in annual systems (Fig. 4.1). There were strong site × cropping system interactions. The poplar and fertilized switchgrass systems produced lower fluxes at KBS than at ARL while fertilized old field and native grass mix systems were virtually identical at both sites. Nitrogen fertilization effects also varied among systems, with strong effects in the switchgrass and old field and minimal effects in the native grass mix and restored prairie.

Different patterns were observed for soil NO_3^- and NH_4^+ concentrations. Sites, cropping systems, and nitrogen fertilization all influenced NO_3^- (Fig. 4.1). Overall NO_3^- concentrations were lower in perennial systems than annual systems, notably in the native grass mix and restored prairie, while within a site NH_4^+ was relatively consistent across systems. NO_3^- concentrations tended to be higher at ARL than in comparable systems at KBS, while NH_4^+ concentrations at KBS were slightly higher and less variable. Data to compare fertilization effects within treatment were only available for ARL, where fertilization greatly increased NO_3^- concentrations in all systems except for the restored prairie, but impacted NH_4^+ to a much smaller degree.

4.3.2 Environmental constraints on N_2O fluxes

We used quantile regression to correlate environmental factors to the upper bound of N_2O fluxes, with τ=0.95 approximately corresponding to the 95th percentile of fluxes at a given value of each constraint. All four environmental factors tested significantly and positively related to this upper bound (Fig. 4.2). Intercepts from these regressions indicated relatively high fluxes even at minimal values of environmental factors; this was particularly unexpected for soil nitrogen, which is essential for N_2O production. We observed, however, that NO_3^- and NH_4^+ are uncorrelated at lower concentrations (Fig. S4.3), indicating that even if one form is nearly absent the other may be available as a substrate. At ARL, including both NO_3^- and NH_4^+ improved the quantile regression model over one using only NO_3^-

($F_{1,3152}$ = 6.5, P < 0.05), although the same was not true at KBS ($F_{1, 1225}$ = 2.7, P =0.10).

In most cases, fitting separate slopes and intercepts for each cropping system significantly improved model performance (all P < 0.05). The sole exception was the NO_3^- relationship at KBS, where only separate slopes were supported. In keeping with the trends observed for the full dataset, most slopes were significantly positive (Fig. 4.3). The most ecologically consistent cropping system effect occurred in the response to temperature, with more positive slopes for annual systems at both sites and nonsignificant slopes in most nonfertilized systems. Similarly, nonfertilized systems at ARL generally had nonsignificant responses to both NO_3^- and NH_4^+, although it should be noted that very few instances of high nitrogen concentrations were observed for these systems (Fig. 4.1, Fig. S4.3). WFPS responses were consistent across systems at ARL and considerably more variable at KBS. Intercepts for soil nitrogen and temperature were generally positive and substantial (Fig. 4.4). WFPS intercepts for most ARL systems were indistinguishable from zero, but were significantly positive for all KBS systems. The miscanthus system at KBS exhibited very curious behavior, with a high intercept and negative slope for NO_3^- and the opposite for NH_4^+. Overall, system-specific quantile regressions replicated the patterns observed for site-level data.

4.4 Discussion

4.4.1 Cropping systems differ in their nitrogen dynamics

We observed substantial cropping system-level differences in soil nitrogen dynamics, notably in the distributions of N_2O fluxes and concentrations of inorganic soil nitrogen. These dynamics were not completely correlated to differences in nitrogen fertilization rates among systems, suggesting systems differed in their capacity to immobilize or transform exogenous nitrogen additions. Species level plant traits such as root morphology and symbiotic associations influence nitrogen uptake capabilities (Craine et al., 2002; De Vries et al., 2015), while at the system level plant species richness can reduce soil inorganic nitrogen concentrations (Oelmann et al., 2007; Palmborg et al., 2005). All of these factors contribute to differences among cropping systems in relationships between exogenous nitrogen inputs, nitrogen pools, and N_2O emissions (Lu et al., 2011; Stehfest and Bouwman, 2006)

In a study of the switchgrass, native grass mix, and restored prairie at ARL in 2011 and 2012, Duran et al. (2016) found all forms of nitrogen loss were lower in diverse plant communities. Over time, these systems should accumulate organic nitrogen, either in soil organic matter or in belowground plant tissues, with unclear long-term consequences for nitrogen cycling. Soil properties such as texture and mineralogy may modify cropping systems effects (Pelster et al., 2011), potentially explaining our observation of site-level differences in cropping system responses. This interaction between cropping systems and soil properties will be of particular importance for predicting the performance of bioenergy feedstocks established on marginal soils, and as such merits further consideration.

4.4.2 Environmental factors constrain N_2O fluxes

Our observation of the importance of soil nitrogen, WFPS, and temperature fits with the general understanding of N_2O production in soils (Robertson and Vitousek, 2009). Nonetheless, this analysis produced some surprising results. Of these, the high intercepts observed with quantile regression were perhaps the most striking. Temperatures below 0 °C might be expected to severely restrict rates of microbial activity; this is an assumption we make when aggregating N_2O emissions on an annual basis. Despite that, we observed fairly high fluxes at these temperatures as well as a relatively muted effect of increasing temperature, particularly at ARL. N_2O fluxes from frozen soils, particularly in response to freeze-thaw events, have been reported elsewhere (Teepe et al., 2000), suggesting it may be inappropriate to discount wintertime microbial activity. Temperature sensitivity may differ for enzymes responsible for N_2O production and consumption, potentially increasing N_2O production at soil temperatures near freezing (Müller et al., 2003). The strength of the quantile regression was particularly evident in the relationship between WFPS and N_2O fluxes at ARL. From the raw data, this appeared to be a log-linear relationship over the entire range of observed WFPS; this contrasts with frequent interpretations of this relationship, which posit that WFPS must exceed some threshold for denitrification as well as a peak in N_2O production near field capacity, rather than near saturation (Hénault et al., 2005; Laville et al., 2011). Our findings suggest WFPS constrains N_2O fluxes even at moisture contents associated with aerobic conditions. The high difference in intercepts between ARL and KBS likely results from differences in

matric potential, the mechanism Castellano et al. (2010) used to explain their observation of greater N_2O fluxes at lower WFPS in more porous soils.

One of our more perplexing observations was the occurrence of relatively high N_2O fluxes at minimal NO_3^- levels. It has been suggested that NO_2^-, rather than NO_3^-, is the most important species for N_2O production (Maharjan and Venterea, 2013). While our measurement of "NO_3^-" concentrations reflects the sum of NO_3^- and NO_2^-, the latter is generally neglected due to its brief residence time in the soil (Butterbach-Bahl et al., 2013). NO_2^- can be produced by both nitrification and denitrification (Butterbach-Bahl et al., 2013), with processes like nitrifier denitrification contributing significantly to N_2O production under conditions that are not conducive to denitrification (Kool et al., 2011). We observed a significant effect of NH_4^+ concentration on N_2O flux upper limits as well as a marginal improvement of quantile regression models that included both NH_4^+ and NO_3^- over those containing only one or the other, all of which suggests nitrification may contribute to N_2O fluxes in these systems. The strong and consistent effect of WFPS we observed appears to contradict this, as nitrification rates are thought to peak at intermediate soil moisture levels (Lehuger et al., 2009), although our findings in Chapter 5 suggest this may not be so clear-cut. The one thing that emerges with clarity from this is the importance of further understanding the importance of nitrification and NH_4^+ concentrations on N_2O production under conditions that are not conducive to denitrification.

4.4.3 Environmental constraints on N_2O fluxes differ by cropping system

Cropping systems differed in their N_2O flux responses to environmental constraints. These differences, however, did not align with our prior classification of ecological differences among systems. Contrast this to the patterns observed in N_2O fluxes and NO_3^- concentrations, whose distributions reflected differences in perenniality, plant species diversity, and fertilization management. The response to soil temperature gave the most ecologically-interpretable pattern: annual systems exhibited the strongest response while nonfertilized systems exhibited the weakest response. This matched the patterns of NO_3^- concentration, and reflected the higher levels of nitrogen addition for the annual systems. Thus, greater nitrogen inputs led to higher soil NO_3^- concentrations, which appear to have allowed N_2O fluxes to respond more strongly

to increases in temperature. We anticipated this mechanism would also apply to moisture and nitrogen availability, but this did not prove to be the case. The different responses to moisture may have reflected enzymatic variability, as oxygen sensitivities for key denitrification enzymes may differ among systems (Cavigelli and Robertson, 2001). Similarly, the relationship between soil nitrogen concentrations and N_2O production varies among systems (Lehuger et al., 2009), as does the proportion of nitrogen inputs that are emitted as N_2O (Duran et al., 2016), leading us to expect different responses to nitrogen concentrations. The limited range of soil nitrogen concentrations we observed for many of our systems, notably nonfertilized ones, may have hampered our detection of system-specific responses. Prior research indicates that denitrification responds less strongly to nitrogen fertilization in nonagricultural systems than in agricultural ones (Lu et al., 2011), suggesting we might have observed greater differentiation among systems at higher soil nitrogen concentrations. These challenges illustrate a major limitation of observational studies, where fully sampling the parameter space may prove extremely difficult. Moreover, the uncertainty in many of our parameter estimates speaks to the exceptionally large volume of data required for this type of approach.

Multiple mechanisms could drive differences among cropping systems in their response to environmental factors. Soil pH exerts important, if frequently overlooked, influence on nitrification and denitrification (Cuhel et al., 2010; Oehler et al., 2010). Carbon available for microbial respiration fuels denitrification (Henry et al., 2008; Senbayram et al., 2012); perennial plants allocate far more carbon belowground and often engage in symbiotic relationships with soil microbiota, greatly increasing soil carbon availability (Warembourg and Estelrich, 2001). Nutrient limitation is thought to increase this dynamic (Bell et al., 2015). At the same time, microbial communities differ in their size, activity, and biochemistry. In particular, they may differ in the proportion of their denitrification that results in N_2O rather than N_2 (Domeignoz-Horta et al., 2015). While the extent and source of these differences in microbial community properties remain unclear, this may be a mechanism by which cropping systems develop different responses to environmental constraints of N_2O fluxes.

4.5 Conclusions

Our study aimed to investigate differences among bioenergy feedstock cropping systems in their production of N_2O and in their response to potential environmental constraints of this process. Cropping systems differed in their distribution of N_2O fluxes and soil nitrate concentrations, with fertilization and annual crops generally increasing levels. Across all systems, soil temperature, WFPS, and concentrations of nitrate and ammonium all correlated to maximum N_2O flux observations, suggesting these factors constrained N_2O production. High fluxes occurred even at low substrate concentrations, temperature, and moisture. Responses to these constraints varied among cropping systems, implying N_2O fluxes may differ among systems under a given set of conditions. Interpretation of our findings needs to be tempered by the limited range of variables measured for some systems, particularly nitrogen concentrations in nonfertilized systems. Nonetheless, this study illustrates environmental constraints over N_2O fluxes and suggests responses to these constraints differ among bioenergy feedstock cropping systems.

Acknowledgements

We thank SK Hamilton, GR Sanford, K Kahmark, JT Simmons, SL VanderWulp, AP Dean, JR Tesmer, AP Butz, JO Sustachek, CE Menick, CC McMinn, NE Tautges, A Miller, Z Andersen, B Faust and many others for providing agronomic management and data collection for this project. S Sippel and S Bohm contributed to database management and the data quality control framework. Funding was provided by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494) and the DOE OBP Office of Energy Efficiency and Renewable Energy (DE-AC05-76RL01830).

References

Bell, C.W., Asao, S., Calderon, F., Wolk, B., Wallenstein, M.D., 2015. Plant nitrogen uptake drives rhizosphere bacterial community assembly during plant growth. Soil Biology and Biochemistry 85, 170–182.

Blanco-Canqui, H., 2010. Energy crops and their implications on soil and environment. Agronomy Journal 102, 403–419.

Bremner, J.M., 1997. Sources of nitrous oxide in soils. Nutrient Cycling in Agroecosystems 49, 7–16.

Butterbach-Bahl, K., Baggs, E.M., Dannenmann, M., Kiese, R., Zechmeister-Boltenstern, S., 2013. Nitrous oxide emissions from soils: How well do we understand the processes and their controls? Philosophical Transactions of the Royal Society B: Biological Sciences 368, 1–13.

Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1, 412–420.

Castellano, M.J., Schmidt, J.P., Kaye, J.P., Walker, C., Graham, C.B., Lin, H., Dell, C.J., 2010. Hydrological and biogeochemical controls on the timing and magnitude of nitrous oxide flux across an agricultural landscape. Global Change Biology 16, 2711–2720.

Cavigelli, M.A., Robertson, G.P., 2001. Role of denitrifier diversity in rates of nitrous oxide consumption in a terrestrial ecosystem. Soil Biology and Biochemistry 33, 297–310.

Craine, J.M., Tilman, D., Wedin, D., Reich, P., Tjoelker, M., Knops, J., 2002. Functional traits, productivity and effects on nitrogen cycling of 33 grassland species. Functional Ecology 16, 563–574.

Crutzen, P.J., Mosier, A.R., Smith, K.A., Winiwarter, W., 2008. N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. Atmospheric Chemistry and Physics 8, 389–395.

Cuhel, J., Simek, M., Laughlin, R.J., Bru, D., Chèneby, D., Watson, C.J., Philippot, L., 2010. Insights into the effect of soil pH on N_2O and N_2 emissions and denitrifier community size and activity. Applied and Environmental Microbiology 76, 1870–8.

De Vries, F.T., Bracht Jørgensen, H., Hedlund, K., Bardgett, R.D., 2015. Disentangling plant and soil microbial controls on carbon and nitrogen loss in grassland mesocosms. Journal of Ecology 103, 629–640.

Dechow, R., Freibauer, A., 2011. Assessment of German nitrous oxide emissions using empirical modelling approaches. Nutrient Cycling in Agroecosystems 91, 235–254.

Domeignoz-Horta, L.A., Spor, A., Bru, D., Breuil, M.-C., Bizouard, F., Léonard, J., Philippot, L., 2015. The diversity of the N_2O reducers matters for the N_2O:N_2 denitrification end-product ratio across an

annual and a perennial cropping system. Frontiers in Microbiology 6, 971.

Duran, B.E.L., Duncan, D.S., Oates, L.G., Kucharik, C.J., Jackson, R.D., 2016. Nitrogen fertilization effects on productivity and nitrogen loss in three grass-based perennial bioenergy cropping systems. PLOS ONE 11, e0151919.

Forster, P., Ramaswamy, V., Artaxo, P., et al., 2007. Changes in atmospheric constituents and in radiative forcing, in: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 129–234.

Garrido, F., Hénault, C., Gaillard, H., Pérez, S., Germon, J.C., 2002. N_2O and NO emissions by agricultural soils with low hydraulic potentials. Soil Biology and Biochemistry 34, 559–575.

Gelfand, I., Sahajpal, R., Zhang, X., Izaurralde, R.C., Gross, K.L., Robertson, G.P., 2013. Sustainable bioenergy production from marginal lands in the US Midwest. Nature 493, 514–517.

Hénault, C., Bizouard, F., Laville, P., Gabrielle, B., Nicoullaud, B., Germon, J.C., Cellier, P., 2005. Predicting in situ soil N_2O emission using NOE algorithm and soil database. Global Change Biology 11, 115–127.

Henry, S., Texier, S., Hallet, S., et al., 2008. Disentangling the rhizosphere effect on nitrate reducers and denitrifiers: insight into the role of root exudates. Environmental Microbiology 10, 3082–3092.

Hooper, D.U., Chapin, F.S., Ewel, J.J., et al., 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecological Monographs 75, 3–35.

Hooper, D.U., Vitousek, P.M., 1998. Effects of plant composition and diversity on nutrient cycling. Ecological Monographs 68, 121–149.

Koenker, R., 2016. quantreg: Quantile regression [WWW Document]. Version 5.21. URL https://cran.r-project.org/package=quantreg

Kool, D.M., Dolfing, J., Wrage, N., Van Groenigen, J.W., 2011. Nitrifier denitrification as a distinct and significant source of nitrous oxide from soil. Soil Biology and Biochemistry 43, 174–178.

Laville, P., Lehuger, S., Loubet, B., Chaumartin, F., Cellier, P., 2011. Effect of management, climate and soil conditions on N_2O and NO emissions from an arable crop rotation using high temporal resolution measurements. Agricultural and Forest Meteorology 151, 228–240.

Lehuger, S., Gabrielle, B., van Oijen, M., Makowski, D., Germon, J.C., Morvan, T., Hénault, C., 2009. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agriculture, Ecosystems and Environment 133, 208–222.

Liu, X., Chen, C.R., Wang, W.J., Hughes, J.M., Lewis, T., Hou, E.Q., Shen, J., 2013. Soil environmental factors rather than denitrification gene abundance control N_2O fluxes in a wet sclerophyll forest with different burning frequency. Soil Biology and Biochemistry 57, 292–300.

Lu, M., Yang, Y., Luo, Y., et al., 2011. Responses of ecosystem nitrogen cycle to nitrogen addition: A meta-analysis. The New Phytologist 189, 1040–1050.

Maharjan, B., Venterea, R.T., 2013. Nitrite intensity explains N management effects on N_2O emissions in maize. Soil Biology & Biochemistry 66, 229–238.

Mathieu, O., Hénault, C., Lévêque, J., Baujard, E., Milloux, M.-J., Andreux, F., 2006. Quantifying the contribution of nitrification and denitrification to the nitrous oxide flux using 15N tracers. Environmental Pollution 144, 933–940.

Molodovskaya, M., Singurindy, O., Richards, B.K., Warland, J., Johnson, M.S., Steenhuis, T.S., 2012. Temporal variability of nitrous oxide from fertilized croplands: Hot moment analysis. Soil Science Society of America Journal 76, 1728–1740.

Montross, M.D., De Bolt, S., Adams, W.C., 2013. Interplay between yield, nitrogen application, and logistics on the potential energetic and greenhouse gas emissions from biomass crops. GCB Bioenergy 5, 664–673.

MSCO, 2013. Michigan State Climatologist's Office: 27 year summary of annual values for Gull Lake (3504) 1981-2010. [WWW Document]. URL http://climate.geo.msu.edu/climate_mi/stations/3504/1981-2010 annual summary.pdf (accessed 6.2.16).

Müller, C., Kammann, C., Ottow, J.C.G., Jäger, H. -J., 2003. Nitrous oxide emission from frozen grassland soil and during thawing periods. Journal of Plant Nutrition and Soil Science 166, 46–53.

NWS, 2013. National Weather Service: Wisconsin 30 year average temperature and precipitation 1981-2010 [WWW Document]. URL www.crh.noaa.gov/images/mkx/ climate/avg_30_year_precip.png (accessed 6.2.16).

Oates, L.G., Duncan, D.S., Gelfand, I., Millar, N., Robertson, G.P., Jackson, R.D., 2016. Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. GCB Bioenergy 8, 539–549.

Oehler, F., Rutherford, J.C., Coco, G., 2010. The use of machine learning algorithms to design a generalized simplified denitrification model. Biogeosciences 7, 3311–3332.

Oelmann, Y., Wilcke, W., Temperton, V.M., et al., 2007. Soil and plant nitrogen pools as related to plant diversity in an experimental grassland. Soil Science Society of America Journal 71, 720–729.

Palmborg, C., Scherer-Lorenzen, M., Jumpponen, A., Carlsson, G., Huss-Danell, K., Högberg, P., 2005. Inorganic soil nitrogen under grassland plant communities of different species composition and diversity. Oikos 110, 271–282.

Pedersen, A.R., 2015. HMR: Flux estimation with static chamber data [WWW Document]. Version 0.4.1. URL http://cran.r-project.org/package=HMR

Pelster, D.E., Chantigny, M.H., Rochette, P., Angers, D.A., Rieux, C., Vanasse, A., 2011. Nitrous oxide emissions respond differently to mineral and organic nitrogen sources in contrasting soil types. Journal of Environmental Quality 41, 427–435.

R Core Team, 2016. R: A language and environment for statistical computing [WWW Document]. Version 3.3.0. URL http://www.r-project.org

Ravishankara, A.R., Daniel, J.S., Portmann, R.W., 2009. Nitrous oxide (N2O): The dominant ozone-depleting substance emitted in the 21st century. Science 326, 123–125.

Robertson, G.P., Dale, V.H., Doering, O.C., et al., 2008. Sustainable biofuels redux. Science 322, 49–50.

Robertson, G.P., Paul, E.A., Harwood, R.R., 2000. Greenhouse gases in intensive agriculture:

Contributions of individual gases to the radiative forcing of the atmosphere. Science 289, 1922–1925.

Robertson, G.P., Sollins, P., Ellis, B.G., Lajtha, K., 1999. Exhangeable ions, pH, and cation exchange capacity, in: Robertson, G.P., Coleman, D.C., Bledsoe, C.S., Sollins, P. (Eds.), Standard Soil Methods for Long Term Ecological Research. Oxford University Press, Oxford, pp. 106 – 114.

Robertson, G.P., Vitousek, P.M., 2009. Nitrogen in agriculture: Balancing the cost of an essential resource. Annual Review of Environment and Resources 34, 97–125.

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems and Environment 216, 344–355.

Senbayram, M., Chen, R., Budai, A., Bakken, L., Dittert, K., 2012. N_2O emission and the $N_2O/(N_2O+N_2)$ product ratio of denitrification as controlled by available carbon substrates and nitrate concentrations. Agriculture, Ecosystems & Environment 147, 4–12.

Stehfest, E., Bouwman, L.F., 2006. N_2O and NO emission from agricultural fields and soils under natural vegetation: Summarizing available measurement data and modeling of global annual emissions. Nutrient Cycling in Agroecosystems 74, 207–228.

Teepe, R., Brumme, R., Beese, F., 2000. Nitrous oxide emissions from frozen soils under agricultural, fallow and forest land. Soil Biology and Biochemistry 32, 1807–1810.

Trybula, E.M., Cibin, R., Burks, J.L., Chaubey, I., Brouder, S.M., Volenec, J.J., 2015. Perennial rhizomatous grasses as bioenergy feedstock in SWAT: parameter development and model improvement. GCB Bioenergy 7, 1185–1202.

U.S. Environmental Protection Agency, 2014. Inventory of US greenhouse gas emissions and sinks: 1990-2012. Washington, DC.

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

Warembourg, F.., Estelrich, H.., 2001. Plant phenology and soil fertility effects on below-ground carbon allocation for an annual (*Bromus madritensis*) and a perennial (*Bromus erectus*) grass species. Soil Biology and Biochemistry 33, 1291–1303.

Werling, B.P., Dickson, T.L., Isaacs, R., et al., 2014. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. Proceedings of the National Academy of Sciences of the United States of America 111, 1652–1657.

Wickham, H., 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.

Table 4.1 Number of N_2O flux and environmental parameter measurements collected, by site and year

Year	Site	N_2O flux	Soil nitrogen	Soil temperature	Water-filled pore space
2009	ARL	692	—	684	663
	KBS	495	—	483	447
2010	ARL	887	640	834	864
	KBS	499	301	316	351
2011	ARL	925	820	922	875
	KBS	634	456	486	483
2012	ARL	1388	1254	1388	1320
	KBS	594	431	460	456
2013	ARL	1387	311	1386	1383
	KBS	733	40	534	565
2014	ARL	1334	130	1229	1194
	KBS	1004	0	659	581

Soil nitrogen measurement methods are incompatible between 2009 and subsequent years. The number of treatments sampled expanded in 2011 at KBS and 2012 at ARL. Frequency of nitrogen data collection was reduced beginning in 2013. Soil nitrogen data were collected in 2014 at KBS but are not yet available.

Table S4.1 Crops and rotational phases for annual cropping systems, by year

Year	System	Rotation	Crop
2009	G01	Continuous corn	Corn
	G02	Corn-canola-soybean	Soybean
	G03	Corn-canola-soybean	Canola
	G04	Corn-canola-soybean	Corn
2010	G01	Continuous corn	Corn
	G02	Corn-canola-soybean	Canola
	G03	Corn-canola-soybean	Corn
	G04	Corn-canola-soybean	Soybean
2011	G01	Continuous corn	Corn
	G02	Corn-canola-soybean	Corn
	G03	Corn-canola-soybean	Soybean
	G04	Corn-canola-soybean	Canola
2012	G01	Continuous corn	Corn
	G02	Continuous corn	Corn (cover crop)
	G03	Corn-soybean	Corn (cover crop)
	G04	Corn-soybean	Soybean (cover crop)
2013	G01	Continuous corn	Corn
	G02	Continuous corn	Corn (cover crop)
	G03	Corn-soybean	Soybean (cover crop)
	G04	Corn-soybean	Corn (cover crop)
2014	G01	Continuous corn	Corn
	G02	Continuous corn	Corn (cover crop)
	G03	Corn-soybean	Corn (cover crop)
	G04	Corn-soybean	Soybean (cover crop)

Cover consisted of rye and Austrian winter pea.

Table S4.2 Dates of annual nitrogen applications, by treatment

		Year					
Site	**Treatment**	2009	2010	2011	2012	2013	2014
ARL	G01	**11-Jun**	**7-Jun**	**28-Jun**	**7-Jun**	**28-Jun**	**16-Jun**
	G02		27-May	**28-Jun**	**7-Jun**	**28-Jun**	**16-Jun**
	G03	18-May	**7-Jun**		**7-Jun**		**16-Jun**
	G04	**11-Jun**		27-May		**28-Jun**	
	G05		27-May	27-May	11-May	30-May	5-Jun
	G06		*Replanted*	27-May	11-May	30-May	5-Jun
	G07		27-May	27-May	11-May	30-May	5-Jun
	G08		21-Apr				
	G09	18-May	27-May	27-May	11-May	30-May	5-Jun
	G10		27-May	27-May	11-May	30-May	5-Jun
KBS	G01	**22-Jun**	**15-Jun**	**13-Jun**	**11-Jun**	**12-Jun**	**18-Jun**
	G02		10-May	**13-Jun**	**11-Jun**	**28-Jun**	**27-Jun**
	G03	15-Jun	**15-Jun**		**11-Jun**		**27-Jun**
	G04	**22-Jun**		20-May		**28-Jun**	
	G05	*Replanted*	10-May	18-May	4-May	16-May	23-May
	G06	22-Jun	10-May	18-May	4-May	16-May	23-May
	G07	*Replanted*	10-May	18-May	4-May	16-May	23-May
	G08		1-Jun				
	G09	15-Jun	10-May	18-May	4-May	16-May	23-May
	G10	*Replanted*	10-May	18-May	4-May	16-May	23-May

Bolded values indicate corn phases From2009-2011, G02-G04 were in a corn-soybean-canola rotation. From 2012-2014, G02 was continuous corn with a cover crop, while G03-G04 were a corn-soybean rotation with a cover crop. Soybean phases received no nitrogen fertilizer.

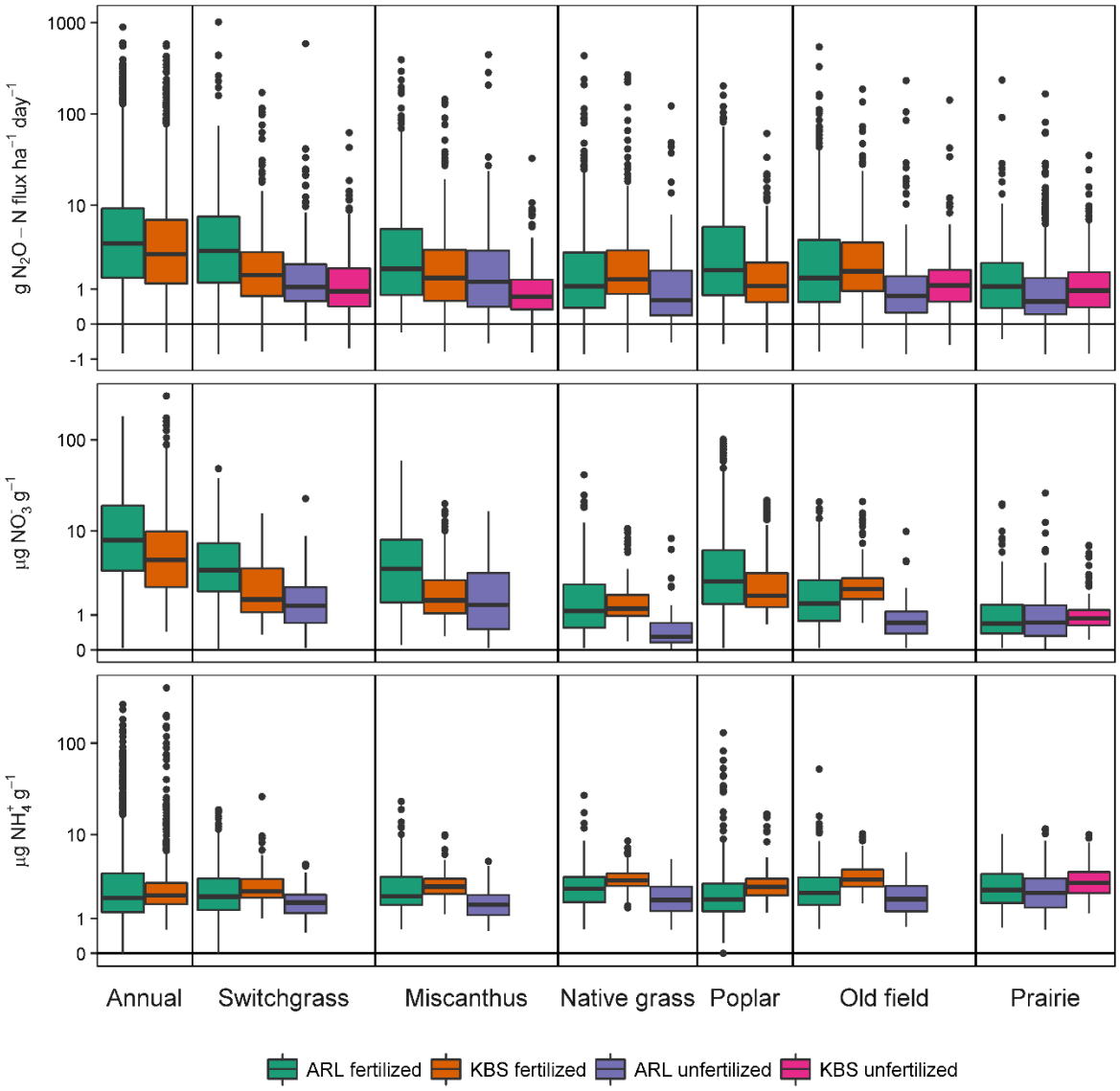

Figure 4.1 Effects of cropping system, site, and fertilization on distributions of N_2O fluxes and concentrations of inorganic soil nitrogen. Values are plotted using an inverse hyperbolic sine (IHS) transformation. Boxplots give 25th, 50th, and 75th percentiles.

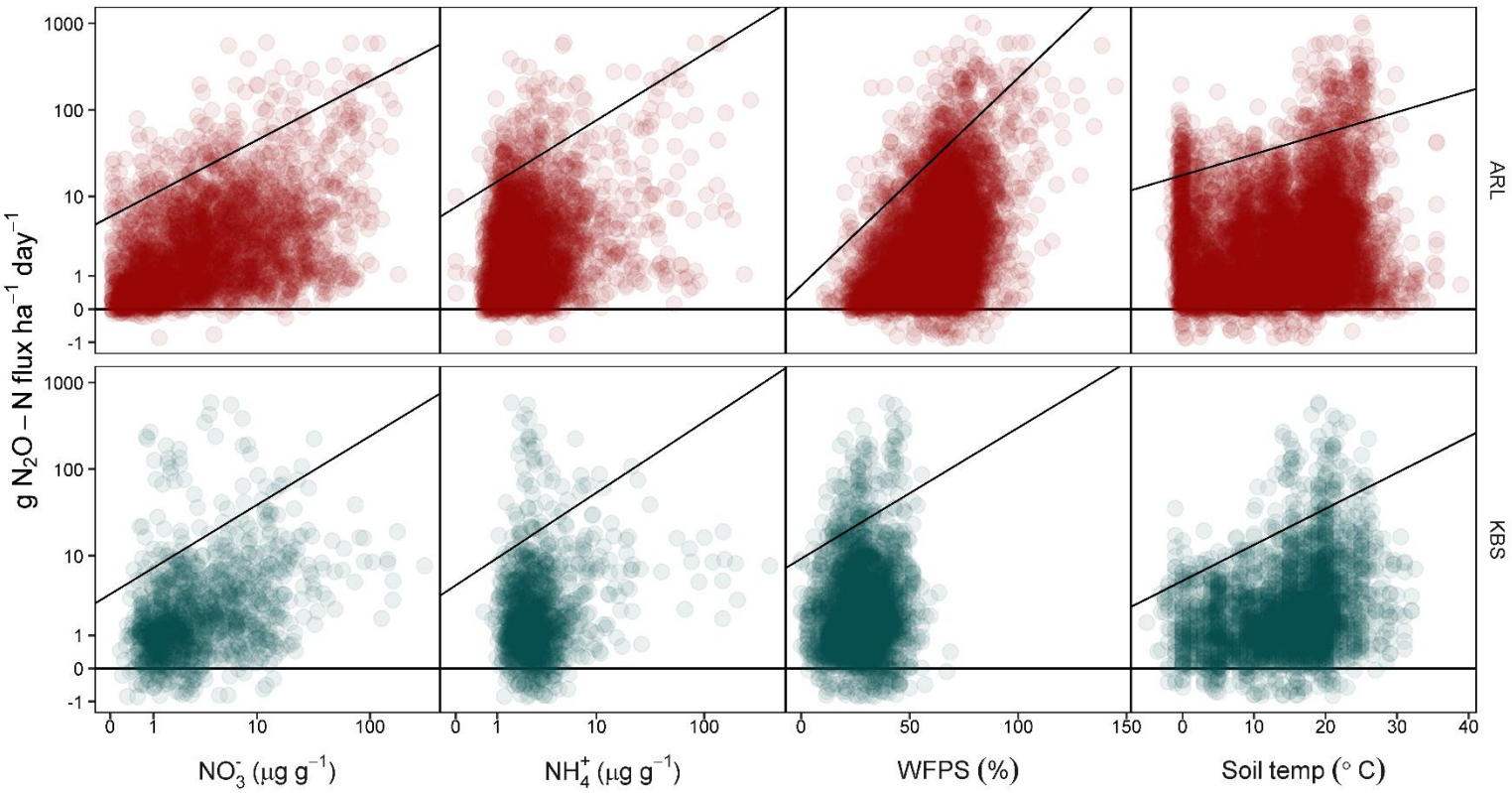

Figure 4.2 Quantile regression between soil N_2O fluxes and environmental parameters. Lines indicate the quantile regression relationship between parameters at $\tau = 0.95$, approximately reflecting the relationship that determines the 95th percentile of fluxes. Regressions were calculated independently for each environmental parameter. Flux data and soil inorganic nitrogen concentration data were inverse hyperbolic sine (IHS) transformed for regression and are presented on a transformed scale. Individual observations are partially transparent to illustrate observation density. All slope and intercept terms are significant at $P < 0.05$.

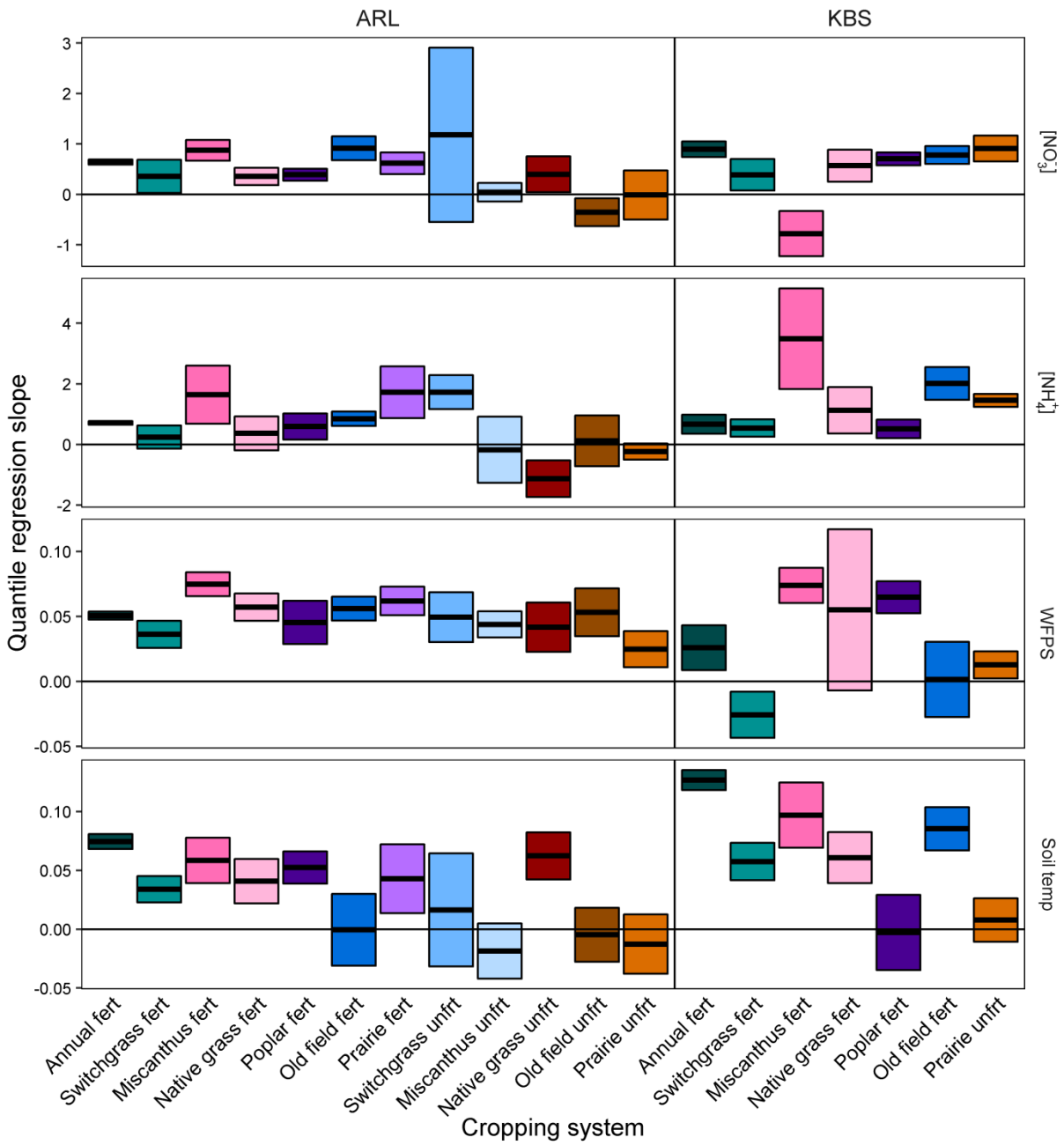

Figure 4.3 Cropping system-specific slopes for quantile regression of soil N_2O fluxes and environmental parameters. Quantile regression was conducted at $\tau = 0.95$. Crossbars indicate slope ± 1 s.e. (based on a kernel estimate). Slope denominators are unit changes in IHS-transformed NO_3^- or NH_4^+ concentrations, percentage points of WFPS, or °C of soil temperature. Corresponding intercepts are given in Figure 4.4.

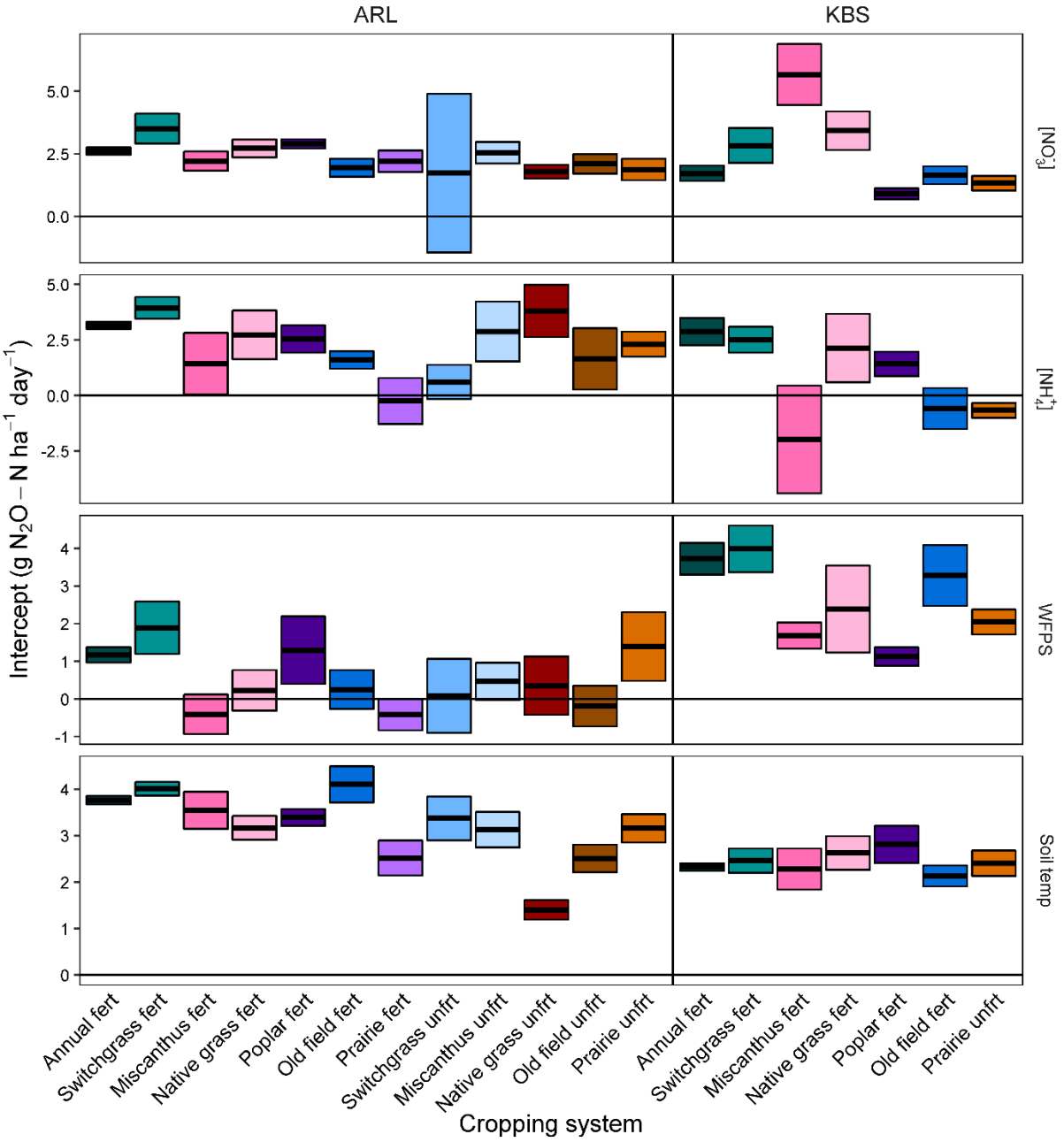

Figure 4.4 Cropping system-specific intercepts for quantile regression of soil N_2O fluxes and environmental parameters. Values approximate the 95th percentile of fluxes at environmental parameter values of 0, ± 1 s.e. (based on a kernel estimate). Corresponding slopes are given in **Figure 4.3**.

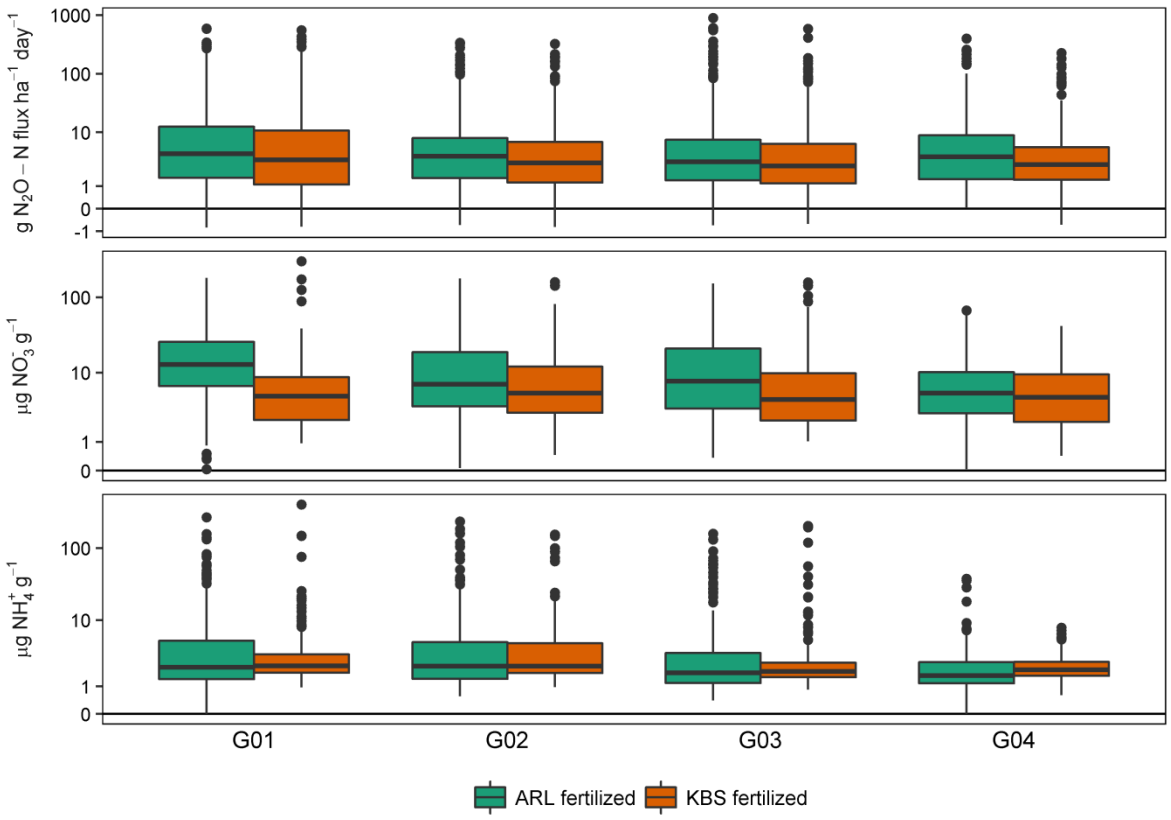

Figure S4.1 Effects of annual cropping system treatments on distributions of soil N_2O fluxes and inorganic nitrogen concentrations. Treatment G01 consists of continuous no-till corn. From 2009 to 2011, treatments G02-G04 were phases of a corn-canola-soybean rotation; from 2012 onward G02 was continuous corn with a cover crop while G03-G04 were phases of a cover-cropped corn-soybean rotation. Values are IHS transformed.

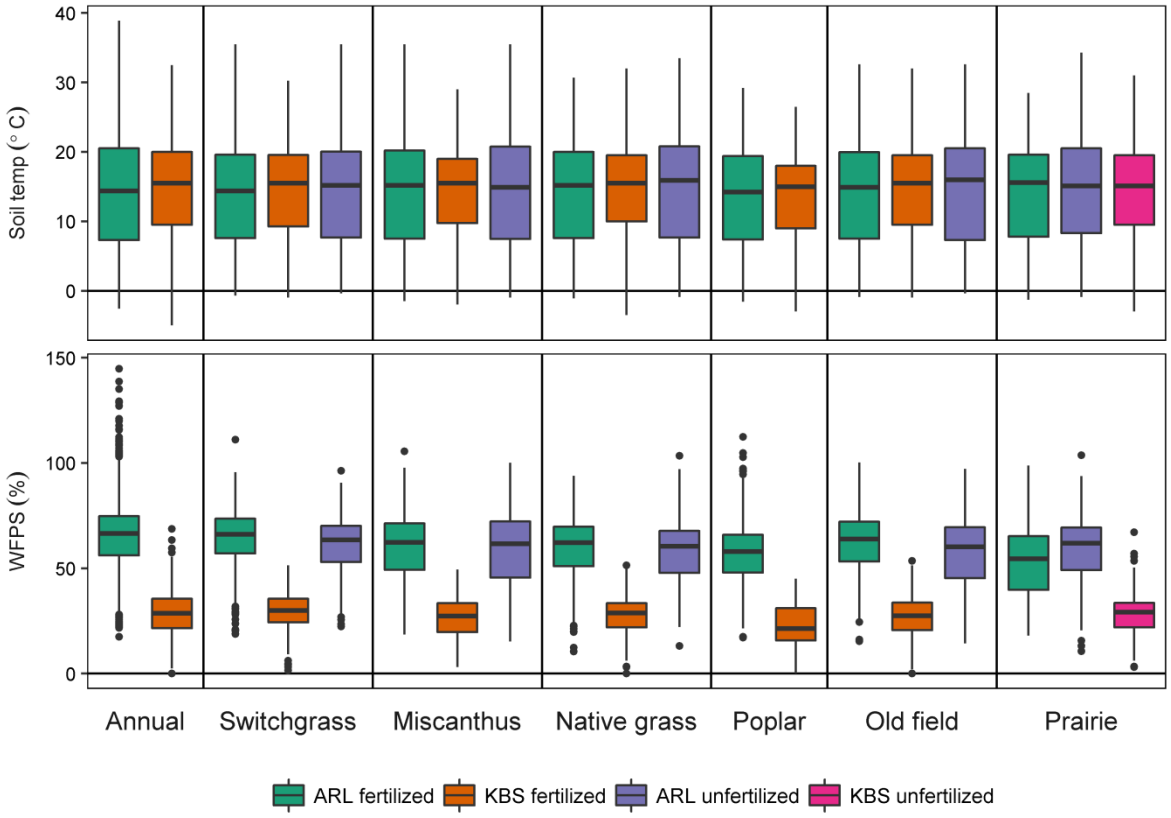

Figure S4.2. Effects of cropping system, site, and fertilization on distributions of soil temperature and WFPS.

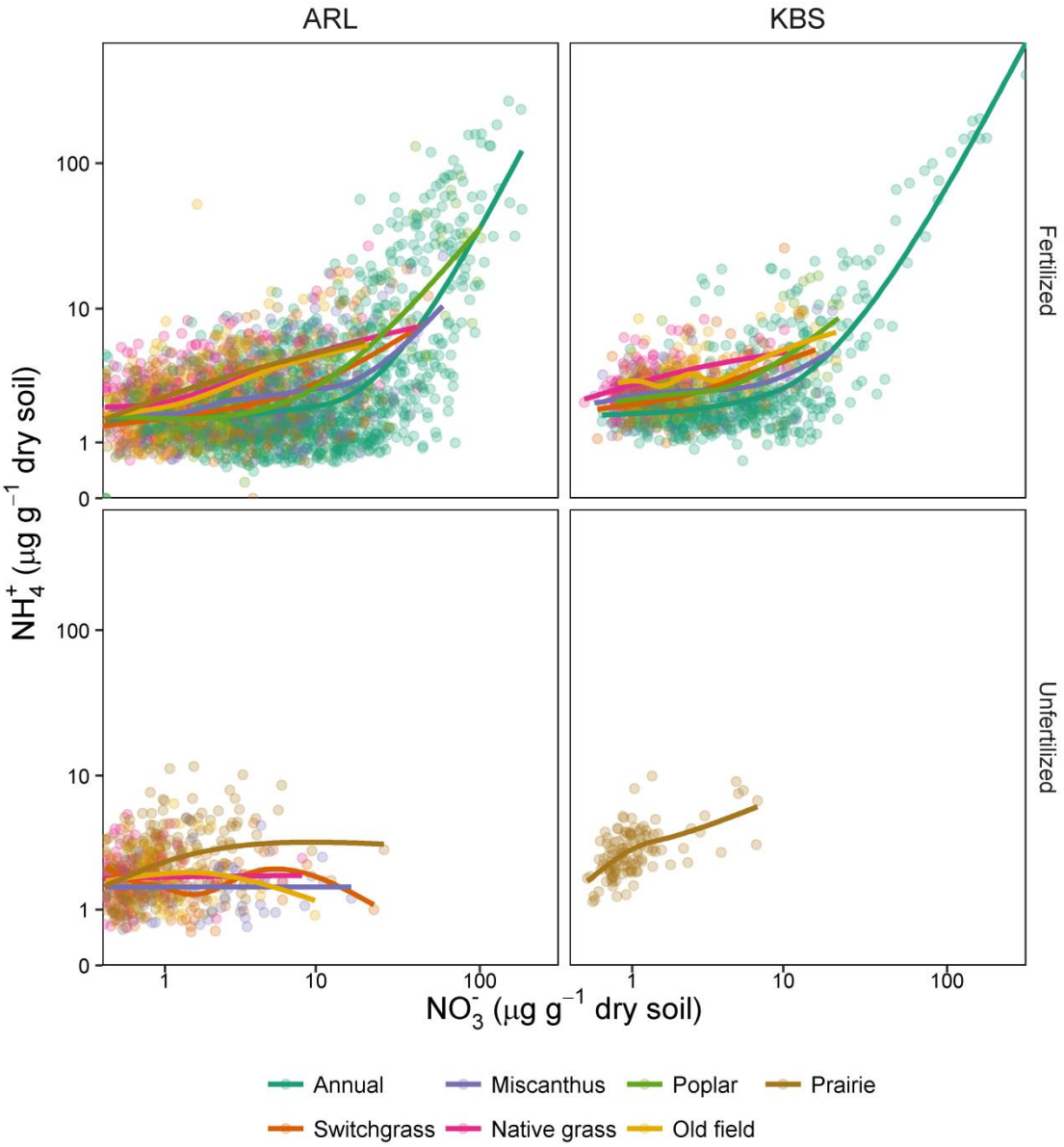

Figure S4.3. Relationship between soil NH_4^+ and NO_3^- concentrations. Lines are loess curves of second-order polynomials with a span of 0.75, plotted by cropping system.

CHAPTER 5

Aggregate annual nitrous oxide emissions correlate to soil microbial functional gene abundance profiles at the plot level and may reflect inherent nitrous oxide production capacities in bioenergy feedstock cropping systems

David S. Duncan, Randall D. Jackson

Department of Agronomy, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA

DOE-Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, 1552 University Avenue, Madison, WI 53726, USA

Abstract

Soil microbial communities likely determine the inherent N_2O production capacity of a system, but microbial roles may be more difficult to observe if environmental factors constrain N_2O fluxes below this inherent capacity. We explored four possible estimators of inherent production capacity: aggregate annual emissions, peak flux events, and potential N_2O production via nitrification (PNR) and denitrification (PDR), as estimated through recalibration of a N_2O production model. Aggregate emissions differed among cropping systems and sites, with nitrogen fertilization generally, but not always, increasing emissions. Within cropping systems, we observed considerable interannual variability and occasionally high intraannual variability. Contrary to our expectations, the timing of peak flux events was not synchronized by fertilization schedules. Likewise, environmental conditions during peak flux events were not more favorable to N_2O production than conditions at other times during the year. N_2O production model recalibration was largely unsuccessful, generating uninformative posterior estimates of PNR and PDR for most samples in our dataset. Aggregate emissions and peak fluxes were highly correlated at both sites, suggesting both measures reflect inherent N_2O production capacity. In contrast, PNR and PDR were uncorrelated to either. We used elastic net modeling to correlate these four estimators to microbial functional gene profiles. This approach explained a high proportion of variability in aggregate emissions but was nonsignificant for peak fluxes, PNR, and PDR. Functional genes retained by the model of aggregate emissions clustered in a small number of functional categories, and few were directly involved in nitrification or denitrification. These correlations indicate functional gene abundances may reflect factors that drive within-system variability in N_2O emissions.

5.1 Introduction

The driving goal of this research project has been to link bioenergy feedstock cropping systems, soil microbial communities, and variability in N_2O production. In Chapter 3, we explored how microbial functional gene profiles differed among the systems of the Bioenergy Cropping Systems Experiment (BCSE). While we found cropping system effects, they were inconsistent and maintained considerable variability among the replicate plots of a system. What effects there were weakly influenced the overall abundance profile, rather than exerting stronger influences on a subset of genetic functions. In Chapter 4, we explored how environmental conditions, such as inorganic nitrogen concentrations, soil moisture, and temperature, influenced N_2O fluxes. We found these factors defined upper limits for N_2O fluxes, with some variation among cropping systems in the nature of this limit. While our analysis in both chapters emphasized cropping system effects, we consistently found considerable plot-level variation in both microbial community composition and N_2O flux dynamics. In this chapter, we focus more on plot-level dynamics, and in particular on the connection between plot-level variability in soil microbial communities and N_2O production.

Soil microbes are both essential to N_2O production and potentially unnecessary for capturing its variability. Virtually all soil N_2O results from microbially-mediated processes, predominantly nitrification and denitrification (Butterbach-Bahl et al., 2013). N_2O production can be dramatically reduced by inhibiting these processes directly (Severin et al., 2016) or by inhibiting microbial growth in general (Mothapo et al., 2013). Despite this, N_2O production can be modeled with reasonable success without incorporating microbial data (Giltrap et al., 2010; Oehler et al., 2010), and incorporating microbial information may not improve empirical models (Graham et al., 2014). Part of this disconnect may reflect the rarity with which microbial activity limits actual N_2O production rates. Both nitrification and denitrification respond strongly to substrate availability, moisture, and pH (Booth et al., 2005; Cuhel et al., 2010). Variation in these environmental factors can be sufficient to explain variation in N_2O production, rendering microbial community characteristics redundant (Attard et al., 2011). At the same time, denitrification gene transcriptional activity can be linked to N_2O production rates (Harter et al.,

2014). Microbial effects are often more visible in properties such as potential denitrification (Morales et al., 2010), N_2O consumption capacity (Jones et al., 2014), or the proportion of N_2O in denitrification products (Domeignoz-Horta et al., 2015). These properties may be more indicative of the inherent N_2O production capacity of a system than of its production at a specific point in time, possibly allowing detection of microbial influences past the obscuring effects of more proximal environmental drivers (Wallenstein et al., 2006).

Inherent, microbially-influenced, capabilities may be a particularly useful property in the context of process-based N_2O production models. Generalized models capable of reflecting N_2O production dynamics across a broad range of environments are highly desirable (Blagodatsky and Smith, 2012), but at present models generally need to be calibrated to a specific cropping system, or even a specific region (Chen et al., 2008). This might be reduced through improved understanding and depiction of factors such as microbial activity that underlie system-specificity of behaviors. Many of extant N_2O models lack an explicit framework for incorporating differences in microbial physiology (compare to Wieder et al., 2014), but include properties such as potential denitrification and N_2:N_2O production ratios (Butterbach-Bahl et al., 2013) which could provide a platform for incorporating microbial community information. For this work, we focused on the nitrous oxide emissions module (NOE) of the CERES model (Hénault et al., 2005). NOE lacks the complexity of more recent models, but only requires information on soil temperature, water-filled pore space (WFPS), and concentrations of ammonium (NH_4^+) and nitrate (NO_3^-), overlapping perfectly with the environmental parameters we measured alongside N_2O fluxes (Chapter 4). Moreover, the authors of the model have published a study where they recalibrated key model parameters to fit differences among field sites (Lehuger et al., 2009). NOE employs terms that reflect potential N_2O production from denitrification and from nitrification, providing a framework by which we could estimate these properties and then relate them to functional gene abundance profiles.

The experiment in this chapter consisted of two steps. First, we attempted to estimate inherent single-year N_2O production capabilities at the plot-year level. These estimates reflected the combination of potential nitrification/denitrification rates and N_2O production ratios, both of which have previously

been linked to microbial functional gene abundances (Domeignoz-Horta et al., 2015; Morales et al., 2010). We also calculated plot-level aggregate annual N_2O emissions, having found in Chapter 2 that these correlated to functional gene profiles, at least in the corn (*Zea mays* L.) system. Finally, we identified peak N_2O flux events, which the constraint-based relationships we identified in Chapter 4 suggested would indicate the best conditions for N_2O production during the year. N_2O production during high-flux events has previously been linked to denitrification gene expression (Németh et al., 2014), suggesting this may be another metric that reflects microbial community influence. After generating and analyzing these four estimators, we attempted to link them to functional gene profiles via elastic net modeling, largely replicating the approach presented in Chapter 2. Through this effort, we attempted to determine whether soil microbial gene abundance patterns reflected within-system variability in estimates of inherent N_2O production capacity.

5.2 Methods

5.2.1 Datasets

We used the N_2O flux and environmental parameter dataset described in Chapter 4 and the microbial functional gene profile dataset described in Chapter 3. Environmental data consisted of N_2O flux point measurements collected from Arlington Agricultural Research Station in Wisconsin (ARL) and W.K. Kellogg Biological Station in Michigan (KBS) from 2009 to 2014 (see any of the preceding chapters for descriptions of site history and agronomic management). Fluxes were estimated using static chambers, as described in Oates et al. (2016). For many of these flux measurements, we had accompanying data on ammonium (NH_4^+) and nitrate (NO_3^-) concentrations, water-filled pore space (WFPS), and soil temperature. While WFPS and temperature were measured systematically throughout the study period, nitrogen data were collected beginning in 2010 at both sites, with a sharp decrease in sampling frequency in 2012 at KBS and 2013 at ARL. Microbial data were collected from 2010 to 2012 at ARL and in 2012 at KBS. Field data are archived in the GLBRC Sustainability Data Catalog (https://data.sustainability.glbrc.org/), while microbial data are available through the Integrated Microbial Genomes database (https://img.jgi.doe.gov/m, with identifying information in Table S3.1). A site-level

breakdown of the number of observations with environmental and microbial data is presented in Table 5.1.

5.2.2 Analysis

All analyses were conducted in the R statistical environment (v3.3.0, R Core Team, 2016). Graphics were generated with the 'ggplot2' package (v2.1.0, Wickham, 2009), and used default settings for boxplot and smoothing summaries.

5.2.2.1 Aggregate annual emissions

Aggregate annual emissions were calculated by linear interpolation (Oates et al., 2016):

$$\sum_{t=0}^{y,T} \frac{F_{y,t} + F_{y,t+1}}{2} (D_{y,t} - D_{y,t+1})$$

In year *y*, *T* flux events were sampled, where $F_{y,t}$ is the flux and $D_{y,t}$ is the date of sample *t*. $D_{y,0}$ corresponds to the last day before $D_{y,1}$ where soil temperature was < 0 °C, while $D_{y,T+1}$ corresponds to the first day after $D_{y,T}$ where soil temperature was < 0 °C. Both $F_{y,0}$ and $F_{y,T+1}$ were assumed to be 0. If temperatures never dropped below 0 °C between $D_{y,T}$ and $Dy+1,1$, $D_{y,T}$ and $D_{y+1,0}$ were taken as December 31 and January 1 of their respective years, while their fluxes were the time-weighted average of $F_{y,T}$ and $F_{y+1,1}$.

There are limitations to this approach. There are indications that N_2O fluxes may not transition linearly over the two-week time period typically used in our study (Molodovskaya et al., 2012), potentially leading to overestimate of aggregate fluxes. Linear interpolation is nonetheless regularly used in the field, and we are unaware of a validated alternative method. Our measurements outside of the growing season were limited, particularly in the earlier years of the experiment. In the absence of measurements, we assumed there were no N_2O fluxes when soil temperatures were < 0 °C, although our findings from Chapter 4 indicate this to be otherwise. Alternative assumptions of unobserved fluxes outside of the growing season minimally impacted aggregate emission rank orders, leading us to continue employing the approach with the least complicated assumptions.

5.2.2.2 Peak fluxes

Based on our findings in Chapter 4, we expected that peak fluxes would occur on days with the fewest environmental constraints on N_2O production. Moreover, we expected that plots within a cropping system should experience similar environmental conditions on a given date. In particular, the timing of fertilizer application is thought to be important (Laville et al., 2011) and was taken into consideration (Table 5.2).

We looked at dates on which peak fluxes occurred as dates on which all plots in a cropping system would be expected to have high N_2O fluxes. For the actual peak flux from a plot, we calculated an average of fluxes from that plot during peak flux days for its cropping system, weighted by the number of plots that peaked on a given date:

$$\sum_{d=1}^{D} \frac{F_d N_d}{P}$$

More than one plot in a system could experience a peak flux event on a given day, thus the number of unique dates with peak flux events, D, would be less than or equal to the total number of plots, P. Flux F_d was recorded for the plot on day d, on which N_d plots from that system recorded peak fluxes.

5.2.2.3 N_2O emissions model calibration

This analysis was based on the NOE Bayesian model recalibration of Lehuger et al. (2009). In the NOE model, N_2O fluxes from nitrification and denitrification are calculated independently, then summed. N_2O production from each process is calculated as a product of the maximum potential rate (P_N, P_D), the ratio of N_2O produced by each process (R_N, R_D), and constraints based on nitrogen substrate concentrations (N_N, D_N), soil moisture (N_W, D_W) and temperature (N_T, D_T):

$$N_2O = P_N R_N N_N N_W N_T + P_D R_D D_N D_W D_T$$

In the original framework, rate potentials and N_2O proportions are determined empirically through laboratory assays (Hénault et al., 2005), whereas we needed to treat these terms as parameters to be calibrated. As we had no means of calculating potential rates and N_2O ratios separately, we calculated their product (PNR, PDR), which we interpreted as the maximum rate of N_2O production via nitrification and denitrification respectively. Our prior distribution for PDR was based on the maximum value of $P_D R_D$

reported in Hénault et al (2005). Their largest value was 5,640 g N ha^{-1} day^{-1}, which we rounded to 10,000 g N ha^{-1} day^{-1} to increase the likelihood that our prior distribution contained any plausible true PDR value (Table 5.3). We used a similarly conservative estimate that PNR would not exceed 10% of PDR, following Matthieu et al. (2006).

We calculated the nitrogen constraint identically to NOE, as a Michaelis-Menten relationship between nitrogen concentration and a half-saturation constant (NK_M, DK_M, units and priors for all fitted parameters given in Table 5.3):

$$N_N = \frac{[NH_4^+]}{NK_M + [NH_4^+]},\ D_N = \frac{[NO_3^-]}{DK_M + [NO_3^-]}$$

In the NOE model, N_T is calculated as a simple Q_{10} relationship, while D_T is calculated as two separate Q_{10} relationships, with a strong response below ~10 °C and a weaker response above that. Due to our inability to handle discontinuous functions (see below), and the relatively small effect this had on relationships (Lehuger et al., 2009), we used the same general function for both N_T and D_T:

$$N_T = \exp\left[\frac{(T-20)\log NQ_{10}}{10}\right], D_T = \exp\left[\frac{(T-20)\log DQ_{10}}{10}\right]$$

We made the most drastic modifications to the soil moisture constraints. In NOE, the constraint for denitrification is a power relationship with a cutoff at 40-80% WFPS below which denitrification is held to be nonexistant. Once again, we needed to avoid discontinuity, and given the significant contribution of denitrification to N_2O production at low WPFS (Bateman and Baggs, 2005), we modeled this relationship more conservatively as an exponential function with a parameter governing the curvature:

$$D_W = \exp[(WFPS - 1)DW_A]$$

The constraint for nitrification required even greater alteration. NOE models this through three terms: upper and lower WFPS bounds, outside of which nitrification does not occur, and an optimal WFPS, with linear slopes connecting the three points. We approximated this using the probability density function of the beta distribution, rescaled to give a maximum value of 1:

$$Nw = \left(\frac{WFPS^{NW_A-1}(1-WFPS)^{NW_B-1}}{B(NW_A, NW_B)}\right)\left(\frac{NW_A + NW_B - 2}{NW_A - 1}\right)$$

We constructed this model using the Stan language and framework as implemented in the 'rstan' R package (Carpenter and et al., n.d.; Hoffman and Gelman, n.d.; Stan Development Team, 2015). Stan differs from other Bayesian modeling software (e.g. BUGS, JAGS) in its use of a No-U-Turn Sampler (NUTS) implementation of Hamiltonian Monte Carlo (Hoffman and Gelman, n.d.). This has the benefit of dramatically reducing the number of iterations needed to appropriately sample the posterior distribution, but also requires continuous functions for its gradient calculations, leading to our need to diverge from the NOE model formulation. To increase our precision, we calculated logarithms for the constraints, summed them, then exponentiated to model the contributions of nitrification and denitrification. Errors were modeled as independent and identically-distributed draws from a normal distribution with a single, site-level variance term. In contrast to most analyses in this project, N_2O fluxes were *not* transformed prior to analysis, as NOE purports to estimate actual flux values.

Individual observations with negative N_2O fluxes (NOE does not accommodate N_2O consumption) or missing environmental data were removed. Only plots with ≥5 valid observations were included in the analysis, with the total number of observations and plot-years given in Table 5.1. A ceiling of 1.0 was set for WPFS values, while floors had been previously set for soil nitrogen concentrations (Chapter 4). Each site was modeled separately. PDR and PNR were calculated at the plot level, while all other parameters were calculated at the site level. Model sampling consisted of 4 separate chains, each with 2000 iterations, of which 500 were discarded as warm ups, for a total of 6000 observations.

5.2.2.4 Elastic net modeling

Our elastic net modeling approach used the 'glmnet' package (v2.0-5, Friedman et al., 2010), largely replicating our approach from Chapter 2. Because of the larger size of this dataset (Table 5.1), leave-one-out cross-validation was intractable. Instead, we used 14-fold cross-validation, removing 2 samples per fold at KBS and 4 to 5 samples per fold at ARL. For each model we conducted 20 independent cross-validations and used the mean ‘lambda.1se’ value as the regularization parameter. Permutation

maintained intact cropping system and year structures. Thus, the marginal effect of the model relative to permuted data reflected explanation of plot-level variability. Potential terms for the model included individual cropping systems, distinguishing between fertilized and unfertilized versions of the same system, individual years (only relevant at ARL, Table 5.1), and a term for the estimated copy number of individual clusters of orthologous groups (COGs, Galperin et al., 2015; Tatusov et al., 2003). Interactions among these terms were not included. We computed models for the full range of alpha values, including both the ridge regression (alpha 0.0) and lasso (alpha 1.0) extremes (Zou and Hastie, 2005).

5.3 Results and Discussion

5.3.1 Dataset description

Our dataset contained 10,679 measurements of soil N_2O fluxes recorded over 6 years (Table 5.1). The basic sampling unit was a plot measured over the course of a single year (plot-year), with 759 plot-years represented in our dataset. The N_2O emissions model recalibration dataset (Section 3.4) comprised 38% of our total observations, with 319 plot-years each having 6 to 19 measurements (median 13.5). Our dataset for elastic net modeling (Section 3.6) was further constrained by availability of microbial data, limiting us to 132 plot-years. We compared distributions of N_2O fluxes and environmental parameters between observations with full environmental and microbial data to those with missing data (Fig. 5.1). Both sets effectively overlapped for ARL, while at KBS the more constrained dataset had slightly lower WFPS and slightly higher temperature and NH_4^+ concentrations. We only collected microbial data from KBS in 2012, a year with an abnormally warm spring and severe drought (Oates et al., 2016; Sanford et al., 2016). This likely directly caused higher temperature and lower WFPS, while the reduced precipitation and soil moisture may have inhibited NH_4^+ movement, leading to higher concentrations. Nonetheless, as these deviations were relatively minor, it appears the data we used for model recalibration and elastic net modeling were likely representative of the full dataset.

5.3.2 Aggregate annual emissions reflect cropping system effects despite high plot-level variability

We linearly interpolated N_2O fluxes to aggregate them into plot level annual emissions. Emissions differed substantially among cropping systems, sites, and fertilizer managements (Fig. 5.2). Emissions

distributions from ARL exceeded those from KBS for all systems except the fertilized native grass and nonfertilized restored prairie which were similar at both sites. The continuous corn (G01) and the corn phase of the rotational system (G03) produced unusually high N_2O emissions at both sites in 2010, while emissions from fertilized switchgrass (*Panicum virgatum* L.) at ARL were lower in 2009 than in subsequent years (Fig. 5.3). Thus, our report of emissions trends based on data from 2009-11 overestimated N_2O emissions from annual systems and underestimated emissions from switchgrass at ARL (Oates et al., 2016), which in the longer-term dataset did not differ from annual systems. The switchgrass system at ARL experiences higher rates of nitrogen loss, notably fertilizer-induced N_2O emissions, relative to the native grass mix and restored prairie (Duran et al., 2016). As a monoculture, the switchgrass system likely has a narrower time window for peak nitrogen uptake than more diverse systems (Oelmann et al., 2007; Palmborg et al., 2005); it may thus be more difficult to match nitrogen applications to the timing and amount of plant demand.

The peculiar behavior of the poplar (*Populus nigra* × *P. maximowiczii*) system merits explanation. At both sites, this system was fertilized once, in 2010 (Table 5.2), likely resulting in the progressive decline in N_2O emissions observed at KBS. The uptick in 2014 coincides with plant coppicing, which likely resulted in release of plant carbon and nitrogen and a temporary decrease in plant uptake. The persistently high N_2O emissions at ARL likely reflect the effects of a *Marssonina spp*. leaf fungus that infected the plants in 2010. This dramatically reduced accumulation of plant biomass in subsequent years (Sanford et al., 2016), and likely curtailed nitrogen uptake.

Though significant, cropping system effects were small relative to interannual, and occasionally intraannual, variability (Fig. 5.3). The within-system variability we observed underscored the extent to which N_2O production capabilities might differ, even among the plots of a cropping system.

5.3.3 Plot level annual peak flux events were not restricted by environmental conditions or fertilizer timing

We expected the variability in aggregate emissions should reflect similar patterns in environmental drivers of N_2O production. We expected this would be particularly evident during extremely high

magnitude flux events (Molodovskaya et al., 2012; Németh et al., 2014). Precipitation events occurring shortly after fertilization seemed likely to generate the highest fluxes recorded during a year (Laville et al., 2011), with accompanying high levels of soil nitrogen, moisture, and temperature, as per our findings in Chapter 4. When we looked at the dates on which these peak flux events occurred, however, we found they were frequently dispersed throughout the year (Fig. 5.4). In some cases, e.g. KBS in 2013 and 2014, many peak flux events followed fertilization events, but much more frequently we encountered large numbers of events prior to fertilization or several months afterward. The strongest example of event-driven synchronization of flux peaks occurred in 2012, where a late July precipitation, the first major one of the growing season, resulted in the highest fluxes observed from nearly all systems, including those receiving no fertilization. Increases in N_2O production are common when rewetting follows a drought (Guo et al., 2014), but the severity of this drought speaks to the magnitude of event required for this degree of synchronization.

While peak flux timings were only weakly determined by major field events, we reasoned that peak fluxes might still require near-ideal environmental conditions. To that end, we contrasted distributions of environmental parameters observed during peak flux events to those observed at all other times (Fig. 5.5). Non-peak fluxes from some years were higher than peak fluxes from others; thus while N_2O fluxes were generally higher during peak flux events, there was still overlap between the distributions. Overlap was much higher for other environmental variables. At ARL, WFPS during peak flux events tended toward the upper end of the distribution, but this was not the case at KBS. For some systems, the distribution of soil nitrogen concentrations was shifted higher during peak flux events relative to other times of year, while for others like the fertilized switchgrass at ARL the two distributions were indistinguishable. It may be that while individual terms were not that different from the norm, on peak flux days all parameters were near the upper end of their distribution, raising the upper limit for N_2O production (Chapter 4). Conditions leading up to a flux event may also matter. For instance, rewetting events after a drought, like those observed in 2012, affect microbial activity in ways that differ

qualitatively from moisture-activity relationships under a more consistent moisture regime (Lawrence et al., 2009; Li et al., 2010). Such interactions might not

5.3.4 Predicting potential nitrification-/denitrification-derived N_2O production through model recalibration

We attempted to recreate the NOE model Bayesian recalibration described by Lehuger et al. (2009), with the modifications described in Section 2.2.3. We estimated the potential for N_2O production from nitrification (PNR) and denitrification (PDR) at the plot-year level, while calibrating other parameters at the site level (Table 5.3). We successfully generated informative posterior distributions for all site-level parameters (Fig. 5.6). However, many parameters pushed up against the edge of their prior ranges, e.g. DK_M and DW_A at ARL) indicating their true value was likely outside of the pre-defined range and indicating that our prior assumptions about their possible values was incorrect. This interpretation is reinforced by inspection of the results reported by Lehuger et al. (2009), who report multiple instances of parameters whose posterior distribution abutted the limits of their priors. Moreover, a large number of their parameters yielded uninformative posterior distributions, similar to what we observed for NQ_{10} at ARL and DK_M at KBS. This suggests the model we used may not be well specified for reflecting the dynamics in our dataset.

The model-derived estimates of environmental constraints over N_2O production suggested extremely dissimilar responses to environmental conditions between our study sites. The model indicated that nitrification was the dominant process at KBS (Fig. 5.6, right panels). The moisture constraint over denitrification, D_W, was miniscule over the range of WFPS values observed at KBS, while N_W, the constraint over nitrification, exhibited considerable variation over that range. The miniscule NK_M estimate indicates nitrification at KBS was limited by soil moisture and temperature, but not NH_4^+ concentrations. The apparent irrelevance of NH_4^+ may agree with our findings in Chapter 4, where the effect of NH_4^+ concentrations on upper bonds of N_2O production could not be separated from their correlation to NO_3^-. These results also likely overstate the extent to which low soil moisture inhibited denitrification at KBS, as research suggests soil matric potential may be more important to determining N_2O production

(Castellano et al., 2010). At ARL, similar soil moisture relationships were modeled for nitrification and denitrification. This runs counter to the general understanding of the two processes, which holds that nitrification occurs best near field capacity (Bateman and Baggs, 2005). The model further posited extremely high saturation constants for both NH_4^+ and NO_3^-, suggesting that even at the high levels we observed both nitrification and denitrification were significantly substrate-limited.

The relationships described by this model should likely be treated with considerable skepticism, particularly in light of the limited extent to which the model captured N_2O flux variability within individual plots (Fig. 5.7). Plot-level root mean squared error (RMSE) appeared to reflect the magnitude of fluxes observed, rather than how well the model reflected their dynamic variability. The recalibrated models in Lehuger et al. (2009) showed similar issues, consistently underestimating large flux events. The NOE model is designed to reflect long-term emissions dynamics rather than individual fluxes (Hénault et al., 2005), likely explaining many of these shortcomings and leaving the smallest ray of hope that it may serve to estimate meaningful inherent capacities.

Alas, estimation of PNR and PDR was no more successful than any other aspect of the process (Fig. 5.8). For the majority of plots, posterior estimates of PDR and PNR were completely uninformative. Results were particularly stark for denitrification at KBS, likely because the miniscule D_W values rendered PDR values irrelevant to the final flux estimate. Nonfertilized systems at ARL also gave largely uninformative posteriors, possibly reflecting the limited range of fluxes and soil nitrogen conditions we observed for them (Fig. 4.1). We did observe considerable variation in PDR for fertilized systems at ARL, notably annual systems, switchgrass, and miscanthus (*Miscanthus* × *giganteus*), as well as in PNR for many systems at KBS. Curiously, there was no systematic variation among systems; instead, plots within some systems captured the entire range of the parameters. In aggregate, it seems likely that the PDR parameter at ARL and the PNR parameter at KBS provided a mechanism for accommodating plot-years with exceptionally large fluxes while largely ignoring systems whose fluxes hovered at more normal levels. It is difficult at this point to interpret the NOE model recalibration approach we attempted here as

anything other than a failure. However, we included PNR and PDR in the subsequent analysis, if only in the spirit of scientific inquiry.

5.3.5 Aggregate N_2O emissions strongly correlated to peak fluxes but not to PDR or PNR

The existence of an inherent, plot-level capacity for N_2O production formed a central tenet of the analyses undertaken in this chapter. The metrics we explored in the previous three sections were all attempts to estimate this capacity. We would thus expect them to correlate to the extent that they reflect similar aspects of their systems. We observed correlation between aggregate annual emissions and peak fluxes, neither of which correlated significantly to PDR or PNR (Fig. 5.9, PDR and PNR correlations to peak fluxes not shown). Prior studies have noted that short-duration, high-intensity flux events contribute significantly to aggregate emissions and to their variability (Molodovskaya et al., 2012; Németh et al., 2014). Despite this, we were unprepared for the strength and generality of this relationship. The influence of peak flux events varies greatly among plots and treatments in our dataset; even the tiny subset from Fig. 5.7 reflects the variety of forms annual flux profiles can take. In our previous analysis of these systems, we found the relative magnitude and importance of peak fluxes varied among cropping systems, with more prominent and influential peaks in fertilized, annual systems than polycultural and unfertilized ones (Oates et al., 2016). This could have resulted in a relationship driven primarily by differences among cropping systems, rather than the within-system relationship we observed (Fig. 5.9). It is worth reflecting on the many assumptions folded into aggregate emissions estimates (Section 2.2.1) as well as the myriad factors influencing the flux we observe during a peak, including short and long term environmental contexts and even a degree of stochasticity in the actual date of sampling. That we see a relationship that is consistent through differences in management, site conditions, and annual weather patterns suggests both of these metrics may reflect some more fundamental cropping system property.

5.3.6 Elastic net modeling relates functional gene profiles to variability in aggregate N_2O emissions

Having generated four estimators of inherent plot-level N_2O production capacity, we used elastic net modeling to evaluate whether they mapped onto aspects of soil microbial functional gene profiles. Given the consistent qualitative differences we observed between ARL and KBS, we analyzed the sites

separately. The set of potential predictors for these models included terms for cropping systems, years, and estimated copy numbers of clusters of orthologous groups (COGs, details in Section 2.2.4). At both sites, we were most successful in modeling aggregate emissions, where they could explain over 75% of the variability in aggregate emissions (Fig. 5.10). Despite the strong correlation between aggregate emissions and peak fluxes (Fig. 5.9), the latter were not as easily modeled with this approach. At ARL, PDR was actually better-modeled than peak fluxes, although PNR at ARL and both metrics at KBS produced effectively null models.

In elastic net modeling, the alpha parameter determines the tradeoff between ridge (alpha=0) and lasso (alpha=1) penalization terms. The distinction can be generalized as a tradeoff between retention of all terms with heavy shrinkage of their coefficients in ridge regression and the retention of a limited number of terms with lower shrinkage in lasso (Friedman et al., 2010; Zou and Hastie, 2005). The effect of alpha differed by site and response variable. Aggregate emissions at ARL and peak fluxes at KBS were largely insensitive to alpha values (Fig. 5.10). In contrast, aggregate emissions modeling at KBS improved with alpha, while dropping alpha from 0 to 0.2 substantially decreased peak flux model quality at KBS. For alpha ≥ 0.5, the number of terms remained relatively constant for most models. These responses contrast with the results we obtained from the ARL corn system in Chapter 2, where we found that increasing alpha above 0.5 substantially decreased both model performance and the number of parameters retained. The differences with Chapter 2 and between sites in this analysis may reflect the relative number of observations for each analysis, where a larger dataset may reduce sensitivity to the particulars of the modeling process.

We used permutation tests to evaluate model performance and guard against overfitting. Of particular concern was the risk of the microbial data simply replicating system-level differences (Fig. 5.2, Fig. 5.5), via cropping system differences in functional gene profiles (Chapter 3). We included terms for cropping systems and years to provide a mechanism for expressing that variability, but also restricted sample permutations to occur within cropping system and year. Thus, cropping system means for the response variables were identical for our real data and all permutations; the only differences occurred

within cropping systems. Thus, the differences between our actual model results and results from permuted data reflect the extent to which plot-level variability in functional gene profiles reflected plot-level variability in estimators of N_2O production capacity. By this metric, functional gene data substantially improved modeling of plot-level variability in aggregate emissions for high values of alpha at KBS and for all values of alpha at ARL (Fig. 5.10). In contrast, the models built to predict peak fluxes from both sites and PDR at ARL performed comparably to results obtained without a plot-level linkage between microbial community data and response variables. From this, we conclude that functional gene data only contributed to modeling plot-level variability of aggregate emissions.

Models of aggregate emissions at ARL and KBS retained few non-microbial terms. At ARL, the exceptions included specific terms for nonfertilized prairie and fertilized corn, whose emissions were systematically lower and higher than those from continuous corn, respectively. At KBS, the only cropping system term retained was poplar, with lower emissions than the continuous corn baseline. All three terms were retained at all alpha values. As in Chapter 2, all year terms were dropped. The terms retained reflect unusual behaviors noted in Section 3.2: at ARL, N_2O emissions were higher in switchgrass than in other fertilized perennial crops while emissions from KBS poplar and ARL nonfertilized prairie were among the lowest from any system in (Fig. 5.3). It should be noted that multiple COG terms were retained alongside the cropping system terms, indicating emissions from these systems differed from what their microbial communities would predict. Switchgrass appears to possess unusual and undesirable nitrogen cycling properties relative to other perennial systems grown at ARL: it emits more N_2O than other systems (Section 3.2), emits a higher proportion of fertilizer nitrogen as N_2O than the restored prairie or native grass mix (Duran et al., 2016) its N_2O fluxes respond differently to environmental drivers than those from virtually every other system (Oates et al., 2016), and it displayed slower nitrogen resorption and reduced resorption proficiency relative to prairie (Jach-Smith and Jackson, 2015). Switchgrass performs like a typical perennial system at KBS, indicating this is not a species-level issue; it is unclear whether the undesirable performance of switchgrass at ARL stems from a mismatch to soil or other growing conditions at the site, or reflects issues with its establishment.

In addition to the cropping system terms, aggregate annual emissions models retained 46 to 38 COGs as terms at ARL and 19-15 at KBS (at alpha from 0.5 to 1). At alpha=1, all COGs retained at KBS belonged to one of six function categories at KBS and only four categories at ARL (Fig. 5.11). Two denitrification-associated COGs were retained in the KBS model: COG5013, the nitrate reductase α subunit, retained at alpha≤0.2, and COG1140, the nitrate reductase β subunit, retained at all alpha levels. By contrast, ARL models at alpha≥0.2 lacked any COGs from function category P, which contains genes involved in inorganic ion transport and metabolism, including all genes directly involved in nitrification and denitrification.

It is important not to over-interpret this observation. At higher alpha levels, individual terms are retained from groups of correlated parameters, with limited discrimination of which term is retained (Helbling et al., 2015). As noted in Chapter 2, terms with limited relevance in an elastic net model may correlate directly to the response variable. This lack of clarity in the importance of whether individual parameters are dropped or retained may reduce the utility of elastic net modeling to identify mechanisms by which the relationships it identifies operate. Moreover, we characterized microbial functional capabilities simply based on gene abundance, whereas other studies suggest variation in N_2O production may be due to physiological differences among microorganisms (Cavigelli and Robertson, 2001; Yoon et al., 2016) or to differences in gene ratios (Palmer et al., 2012; Philippot et al., 2011). Gene expression might provide a better link to microbial activity than gene abundance, but this expression would have to be recorded during conditions that approximated inherent production capabilities (e.g. Uchida et al., 2014). Our study does not serve to establish a causal link between microbial community properties and N_2O production, but we demonstrate that the substantial plot-level variability is, in some way, related.

5.4 Conclusions

We explored four plot-level estimators of inherent N_2O production capacity which we attempted to relate to microbial functional gene abundance profiles. Bayesian recalibration of the NOE N_2O production model failed to generate informative or credible estimates of potential N_2O production by nitrification and denitrification. Aggregate annual emissions and peak annual fluxes, two estimators based on empirical

measurements, correlated strongly to each other despite many potential sources of divergence, and as such may reflect underlying plot-level properties. Despite this correlation, aggregate emissions proved easier to model from functional gene abundance data using elastic net modeling, suggesting the two estimators differ slightly in the system properties they reflect. Models were quite good at capturing variability of plots within cropping system, and did so using genes from relatively few different function categories. Genes involved in denitrification were included in the models for KBS, but no nitrogen cycle genes were included in models for ARL. Overall, our findings strongly suggest a relationship between plot-level variabilities in aggregate N_2O emissions and microbial functional gene abundance profiles, although this relationship may not rest on differences in the microbial community's genetic capacity for N_2O production and consumption.

References

Attard, E., Recous, S., Chabbi, A., et al., 2011. Soil environmental conditions rather than denitrifier abundance and diversity drive potential denitrification after changes in land uses. Global Change Biology 17, 1975–1989.

Bateman, E.J., Baggs, E.M., 2005. Contributions of nitrification and denitrification to N_2O emissions from soils at different water-filled pore space. Biology and Fertility of Soils 41, 379–388.

Blagodatsky, S., Smith, P., 2012. Soil physics meets soil biology: Towards better mechanistic prediction of greenhouse gas emissions from soil. Soil Biology & Biochemistry 47, 78–92.

Booth, M.S., Stark, J.M., Rasteter, E., 2005. Controls on nitrogen cycling in terrestrial ecosystems: A synthetic analysis of literature data. Ecological Monographs 75, 139–157.

Butterbach-Bahl, K., Baggs, E.M., Dannenmann, M., Kiese, R., Zechmeister-Boltenstern, S., 2013. Nitrous oxide emissions from soils: How well do we understand the processes and their controls? Philosophical Transactions of the Royal Society B: Biological Sciences 368, 1–13.

Carpenter, B., et al., *in review*. Stan: A probabilistic programming language. Journal of Statistical Software.

Castellano, M.J., Schmidt, J.P., Kaye, J.P., Walker, C., Graham, C.B., Lin, H., Dell, C.J., 2010.

Hydrological and biogeochemical controls on the timing and magnitude of nitrous oxide flux across an agricultural landscape. Global Change Biology 16, 2711–2720.

Cavigelli, M.A., Robertson, G.P., 2001. Role of denitrifier diversity in rates of nitrous oxide consumption in a terrestrial ecosystem. Soil Biology and Biochemistry 33, 297–310.

Chen, D., Li, Y., Grace, P., Mosier, A.R., 2008. N_2O emissions from agricultural lands: A synthesis of simulation approaches. Plant and Soil 309, 169–189.

Cuhel, J., Simek, M., Laughlin, R.J., Bru, D., Chèneby, D., Watson, C.J., Philippot, L., 2010. Insights into the effect of soil pH on N_2O and N_2 emissions and denitrifier community size and activity. Applied and Environmental Microbiology 76, 1870–8.

Domeignoz-Horta, L.A., Spor, A., Bru, D., Breuil, M.-C., Bizouard, F., Léonard, J., Philippot, L., 2015. The diversity of the N_2O reducers matters for the N_2O:N_2 denitrification end-product ratio across an annual and a perennial cropping system. Frontiers in Microbiology 6, 971.

Duran, B.E.L., Duncan, D.S., Oates, L.G., Kucharik, C.J., Jackson, R.D., 2016. Nitrogen fertilization effects on productivity and nitrogen loss in three grass-based perennial bioenergy cropping systems. PLOS ONE 11, e0151919.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models viaa coordinate descent. Journal of Statistical Software 33, 1–22.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E. V, 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Research 43, D261–D269.

Giltrap, D.L., Li, C., Saggar, S., 2010. DNDC: A process-based model of greenhouse gas fluxes from agricultural soils. Agriculture, Ecosystems & Environment 136, 292–300.

Graham, E.B., Wieder, W.R., Leff, J.W., et al., 2014. Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes. Soil Biology & Biochemistry 68, 279–282.

Guo, X., Drury, C.F., Yang, X., Daniel Reynolds, W., Fan, R., 2014. The extent of soil drying and

rewetting affects nitrous oxide emissions, denitrification, and nitrogen mineralization. Soil Science Society of America Journal 78, 194–204.

Harter, J., Krause, H.-M., Schuettler, S., et al., 2014. Linking N_2O emissions from biochar-amended soil to the structure and function of the N-cycling microbial community. The ISME Journal 8, 660–674.

Helbling, D.E., Johnson, D.R., Lee, T.K., Scheidegger, A., Fenner, K., 2015. A framework for establishing predictive relationships between specific bacterial 16S rRNA sequence abundances and biotransformation rates. Water Research 70, 471–484.

Hénault, C., Bizouard, F., Laville, P., Gabrielle, B., Nicoullaud, B., Germon, J.C., Cellier, P., 2005. Predicting in situ soil N_2O emission using NOE algorithm and soil database. Global Change Biology 11, 115–127.

Hoffman, M.D., Gelman, A., n.d. The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Jourlan of Machine Learning Research.

Jach-Smith, L.C., Jackson, R.D., 2015. Nitrogen conservation decreases with fertilizer addition in two perennial grass cropping systems for bioenergy. Agriculture, Ecosystems & Environment 204, 62–71.

Jones, C.M., Spor, A., Brennan, F.P., et al., 2014. Recently identified microbial guild mediates soil N_2O sink capacity. Nature Climate Change 4, 801–805.

Laville, P., Lehuger, S., Loubet, B., Chaumartin, F., Cellier, P., 2011. Effect of management, climate and soil conditions on N_2O and NO emissions from an arable crop rotation using high temporal resolution measurements. Agricultural and Forest Meteorology 151, 228–240.

Lawrence, C.R., Neff, J.C., Schimel, J.P., 2009. Does adding microbial mechanisms of decomposition improve soil organic matter models? A comparison of four models using data from a pulsed rewetting experiment. Soil Biology & Biochemistry 41, 1923–1934.

Lehuger, S., Gabrielle, B., van Oijen, M., Makowski, D., Germon, J.C., Morvan, T., Hénault, C., 2009. Bayesian calibration of the nitrous oxide emission module of an agro-ecosystem model. Agriculture, Ecosystems and Environment 133, 208–222.

Li, X., Miller, A.E., Meixner, T., Schimel, J.P., Melack, J.M., Sickman, J.O., 2010. Adding an empirical factor to better represent the rewetting pulse mechanism in a soil biogeochemical model. Geoderma 159, 440–451.

Mathieu, O., Hénault, C., Lévêque, J., Baujard, E., Milloux, M.-J., Andreux, F., 2006. Quantifying the contribution of nitrification and denitrification to the nitrous oxide flux using ^{15}N tracers. Environmental Pollution 144, 933–940.

Molodovskaya, M., Singurindy, O., Richards, B.K., Warland, J., Johnson, M.S., Steenhuis, T.S., 2012. Temporal variability of nitrous oxide from fertilized croplands: Hot moment analysis. Soil Science Society of America Journal 76, 1728–1740.

Morales, S.E., Cosart, T., Holben, W.E., 2010. Bacterial gene abundances as indicators of greenhouse gas emission in soils. The ISME Journal 4, 799–808.

Mothapo, N. V., Chen, H., Cubeta, M.A., Shi, W., 2013. Nitrous oxide producing activity of diverse fungi from distinct agroecosystems. Soil Biology & Biochemistry 66, 94–101.

Németh, D.D., Wagner-Riddle, C., Dunfield, K.E., 2014. Abundance and gene expression in nitrifier and denitrifier communities associated with a field scale spring thaw N_2O flux event. Soil Biology and Biochemistry 73, 1–9.

Oates, L.G., Duncan, D.S., Gelfand, I., Millar, N., Robertson, G.P., Jackson, R.D., 2016. Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. GCB Bioenergy 8, 539–549.

Oehler, F., Rutherford, J.C., Coco, G., 2010. The use of machine learning algorithms to design a generalized simplified denitrification model. Biogeosciences 7, 3311–3332.

Oelmann, Y., Wilcke, W., Temperton, V.M., et al., 2007. Soil and plant nitrogen pools as related to plant diversity in an experimental grassland. Soil Science Society of America Journal 71, 720–729.

Palmborg, C., Scherer-Lorenzen, M., Jumpponen, A., Carlsson, G., Huss-Danell, K., Högberg, P., 2005. Inorganic soil nitrogen under grassland plant communities of different species composition and diversity. Oikos 110, 271–282.

Palmer, K., Biasi, C., Horn, M.A., 2012. Contrasting denitrifier communities relate to contrasting N_2O emission patterns from acidic peat soils in arctic tundra. The ISME Journal 6, 1058–77.

Philippot, L., Andert, J., Jones, C.M., Bru, D., Hallin, S., 2011. Importance of denitrifiers lacking the genes encoding the nitrous oxide reductase for N_2O emissions from soil. Global Change Biology 17, 1497–1504.

R Core Team, 2016. R: A language and environment for statistical computing [WWW Document]. Version 3.3.0. URL http://www.r-project.org

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems and Environment 216, 344–355.

Severin, M., Fuß, R., Well, R., Hähndel, R., Van den Weghe, H., 2016. Greenhouse gas emissions after application of digestate: short-term effects of nitrification inhibitor and application technique effects. Archives of Agronomy and Soil Science 62, 1007–1020.

Stan Development Team, 2015. Stan: A C++ library for probability and sampling, Version 2.8.0 [WWW Document]. URL http://mc-stan.org/

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., et al., 2003. The COG database: An updated version includes eukaryotes. BMC Bioinformatics 4, 1–14.

Uchida, Y., Wang, Y., Akiyama, H., Nakajima, Y., Hayatsu, M., 2014. Expression of denitrification genes in response to a waterlogging event in a Fluvisol and its relationship with large nitrous oxide pulses. FEMS Microbiology Ecology 88, 407–23.

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

Wickham, H., 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.

Wieder, W.R., Grandy, A.S., Kallenbach, C.M., Bonan, G.B., 2014. Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS)

model. Biogeosciences 11, 3899–3917.

Yoon, S., Nissen, S., Park, D., Sanford, R.A., Löffler, F.E., 2016. Nitrous oxide reduction kinetics distinguish bacteria harboring Clade I *nosZ* from those harboring Clade II *nosZ.* Applied and Environmental Microbiology 82, 3793–3800.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320.

Table 5.1 Number of environmental data observations and samples for different levels of data completeness

	ARL			KBS		
	Obs.	Samples	Years	Obs.	Samples	Years
Full dataset	6689	416	2009-14	3990	343	2009-14
All field data	2685	207	2010-13	1017	112	2010-12
Field + microbial data	1534	105	2010-12	276	27	2012

A sample is defined as a plot measured over a year. Measurements with full field data have N_2O flux, ammonium/nitrate concentration, water-filled pore space, and soil temperature data and came from a sample with a minimum of five measurements.

Table 5.2 Nitrogen fertilizer application dates

			Year					
Site	**Code**	**System**	2009	2010	2011	2012	2013	2014
ARL	G01	C. Corn	**11-Jun**	**7-Jun**	**28-Jun**	**7-Jun**	**28-Jun**	**16-Jun**
	G02	Rotation		27-May	**28-Jun**	**7-Jun**	**28-Jun**	**16-Jun**
	G03	Rotation	18-May	**7-Jun**		**7-Jun**		**16-Jun**
	G04	Rotation	**11-Jun**		27-May		**28-Jun**	
	G05	Switchgrass		27-May	27-May	11-May	30-May	5-Jun
	G06	Miscanthus		*Replanted*	27-May	11-May	30-May	5-Jun
	G07	Grass mix		27-May	27-May	11-May	30-May	5-Jun
	G08	Poplar		21-Apr				
	G09	Old field	18-May	27-May	27-May	11-May	30-May	5-Jun
	G10	Prairie		27-May	27-May	11-May	30-May	5-Jun
KBS	G01	C. Corn	**22-Jun**	**15-Jun**	**13-Jun**	**11-Jun**	**12-Jun**	**18-Jun**
	G02	Rotation		10-May	**13-Jun**	**11-Jun**	**28-Jun**	**27-Jun**
	G03	Rotation	15-Jun	**15-Jun**		**11-Jun**		**27-Jun**
	G04	Rotation	**22-Jun**		20-May		**28-Jun**	
	G05	Switchgrass	*Replanted*	10-May	18-May	4-May	16-May	23-May
	G06	Miscanthus	22-Jun	10-May	18-May	4-May	16-May	23-May
	G07	Grass mix	*Replanted*	10-May	18-May	4-May	16-May	23-May
	G08	Poplar		1-Jun				
	G09	Old field	15-Jun	10-May	18-May	4-May	16-May	23-May
	G10	Prairie	*Replanted*	10-May	18-May	4-May	16-May	23-May

Bolded values indicate corn phases From2009-2011, G02-G04 were in a corn-canola-soybean rotation. From 2012-2014, G02 was continuous corn with a cover crop, while G03-G04 were a corn-soybean rotation with a cover crop. Soybean phases received no nitrogen fertilizer.

Table 5.3 Description and prior distributions for parameters used in N_2O emissions model recalibration

			Prior distribution	
Process	**Parameter**	**Description**	Min	Max
Denitrification	DK_M	Half saturation constant (nitrate)	5	120
	DQ_{10}	Q_{10} factor	1.0	2.5
	DW_A	Exponential coefficient for WFPS relationship	0.1	10.0
	PDR	Potential denitrification-driven N_2O flux	1	10000
Nitrification	*NKM*	Half saturation constant (ammonium)	1	50
	NQ_{10}	Q_{10} factor	1.9	13
	NW_A	Beta distribution parameters for WFPS	5	15
	NW_B	relationship	1	12
	PNR	Potential nitrification-driven N_2O flux	1	1000

Half saturation constants are in units of ug g soil, potential fluxes are in g ha^{-1} day^{-1}, all other terms are unitless. PDR and PNR are not denitrification potential rates, as they reflect N_2O production ratios for these processes, upper values are based on Hénault et al. (2005). Half saturation constants and NQ_{10} ranges are based on values from Lehuger et al. (2009), other values were given minimally informative priors.

Table 5.4 Clusters of orthologous groups (COG) function categories.

Category	Function
J	Translation, ribosomal structure and biogenesis
A	RNA processing and modification
K	Transcription
L	Replication, recombination and repair
B	Chromatin structure and dynamics
D	Cell cycle control, cell division, chromosome partitioning
Y	Nuclear structure
V	Defense mechanisms
T	Signal transduction mechanisms
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
Z	Cytoskeleton
W	Extracellular structures
U	Intracellular trafficking, secretion, and vesicular transport
O	Posttranslational modification, protein turnover, chaperones
X	Mobilome: prophages, transposons
C	Energy production and conversion
G	Carbohydrate transport and metabolism
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport and catabolism
R	General function prediction only
S	Function unknown

Information copied from
ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data/fun2003-2014.tab

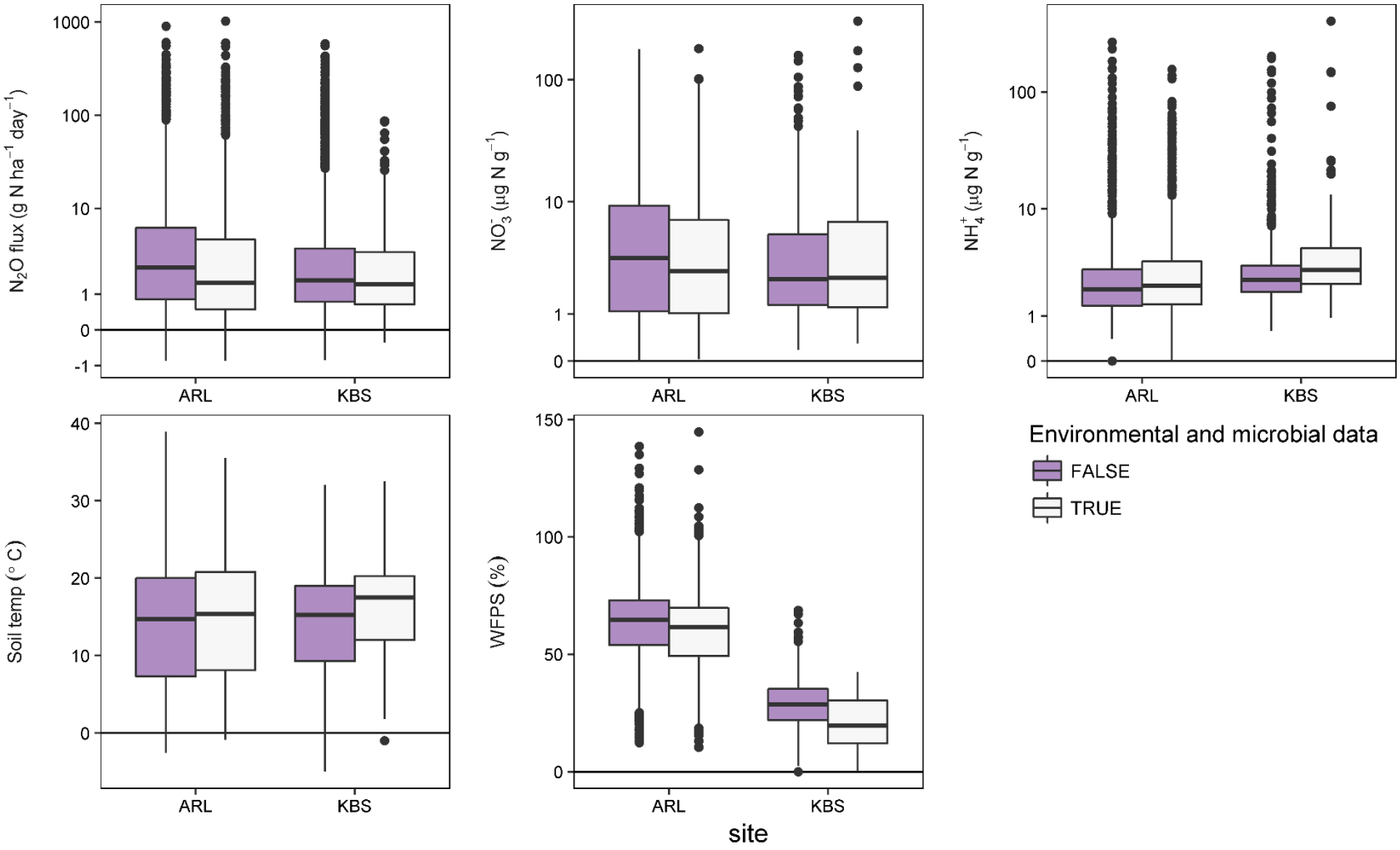

Figure 5.1 Comparison of N_2O flux and environmental measurement distributions between samples with and without full microbial and environmental data. Top panels were transformed by inverse hyperbolic sine (IHS). Boxes indicate 25th, 50th, and 75th percentiles.

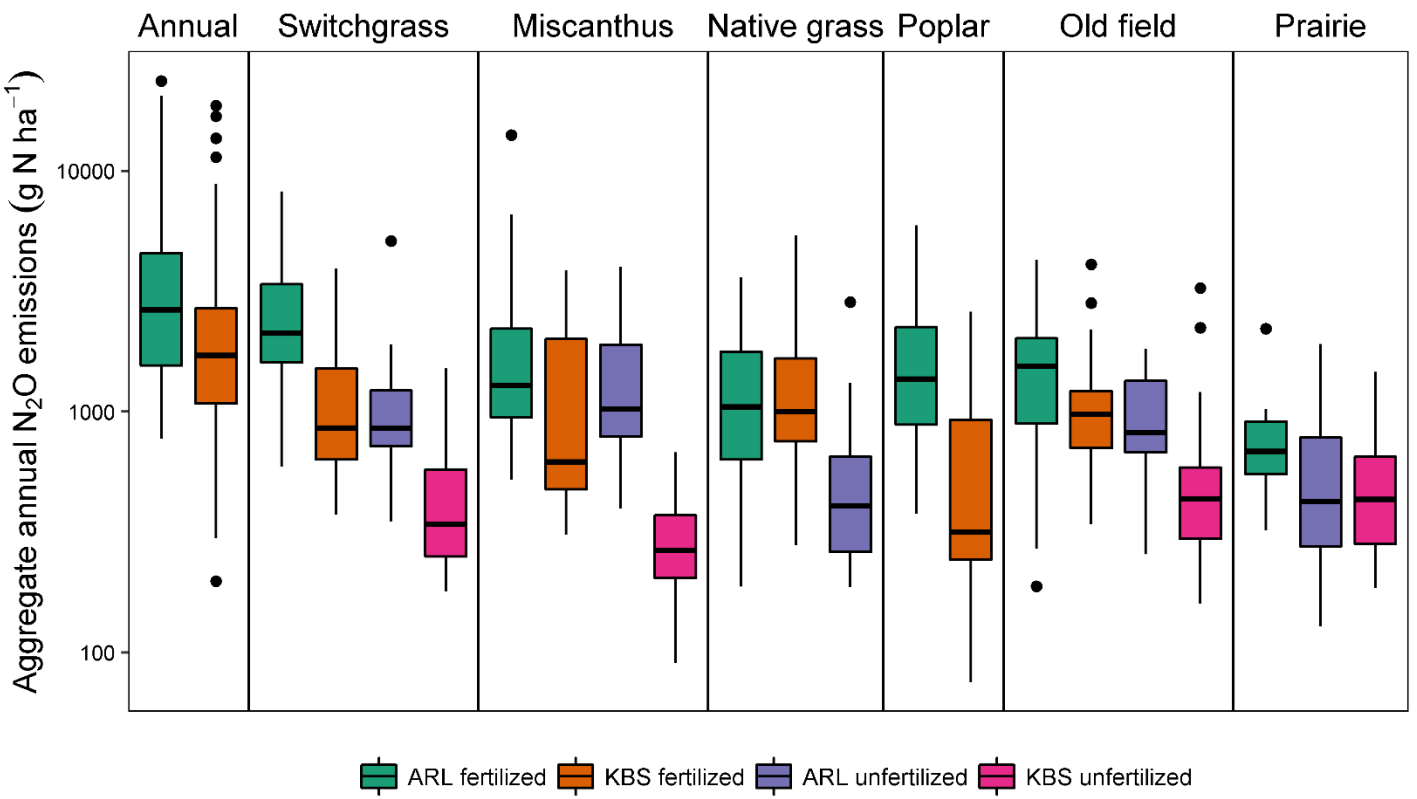

Figure 5.2 Distribution of aggregate annual N_2O emissions among sites and fertilization. Values are plotted on a logarithmic scale. Boxes indicate 25^{th}, 50^{th}, and 75^{th} percentiles. Note that the number of observations and the range of years represented varied by cropping system and site (see **Fig. 5.3**).

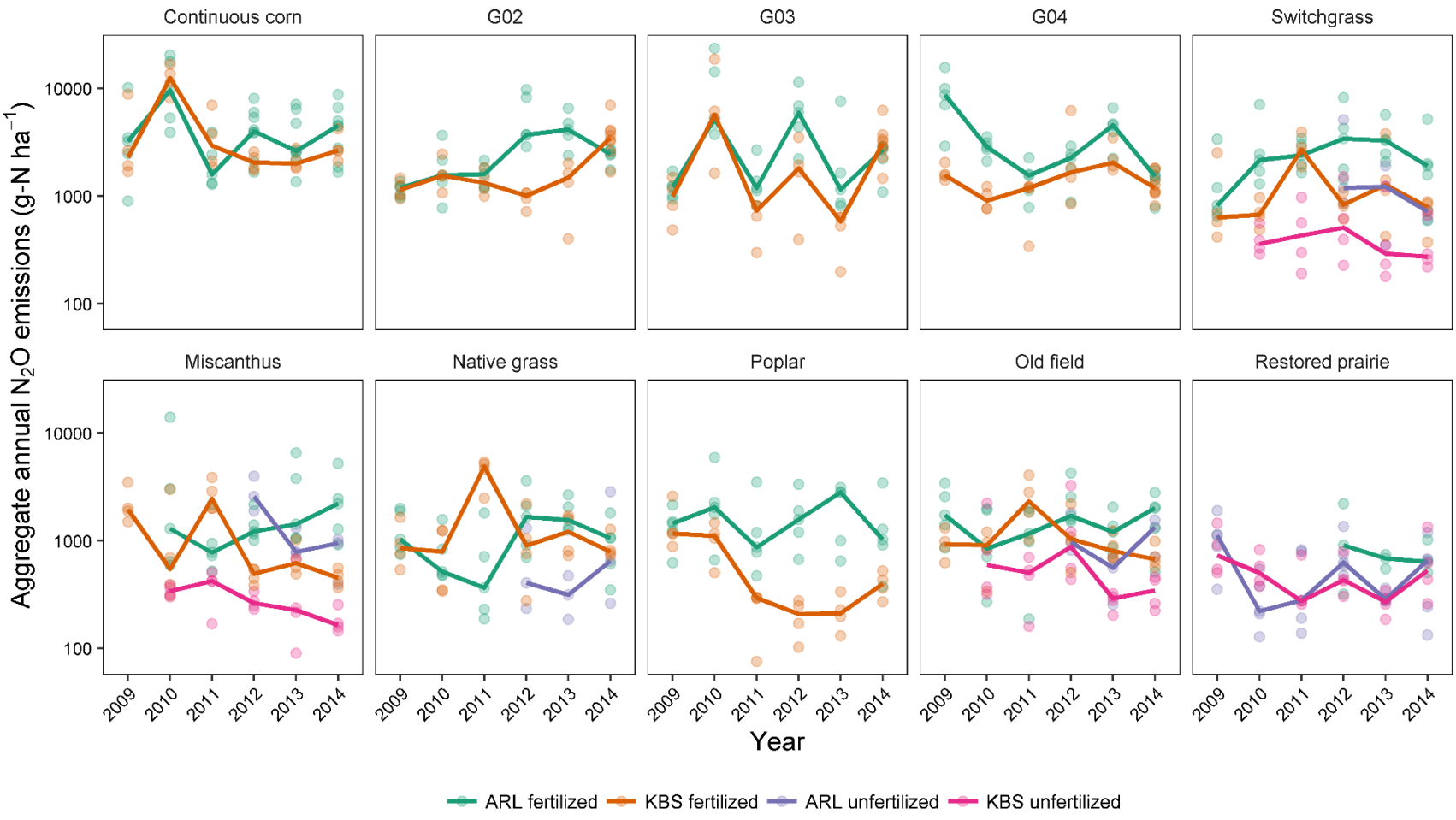

Figure 5.3 Aggregate annual N_2O emissions by year and treatment. Lines indicate median values, n=5 at ARL, n=4 at KBS. G02-G04 are annual crop rotations: corn-soybean-canola from 2009-11, after which G02 was cover-cropped continuous corn and G03-G04 were a corn-soybean rotation grown with cover crops (rotation schedule is given in **Table S4.1**).

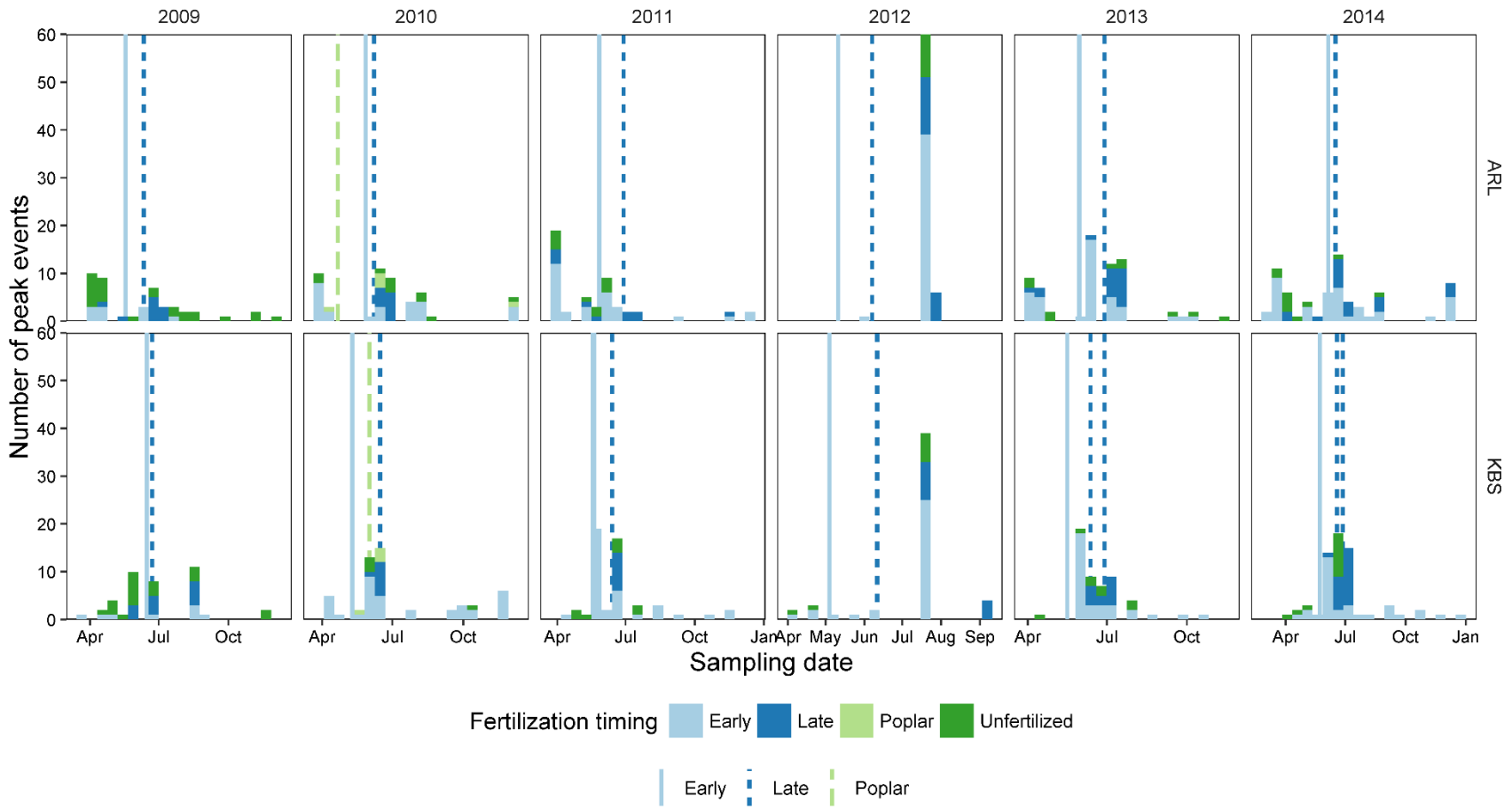

Figure 5.4 Dates on which yearly peak flux events were observed. One peak flux was recorded per sample. Typically, corn systems were fertilized later than all other systems; full fertilization schedule is given in **Table 5.2**.

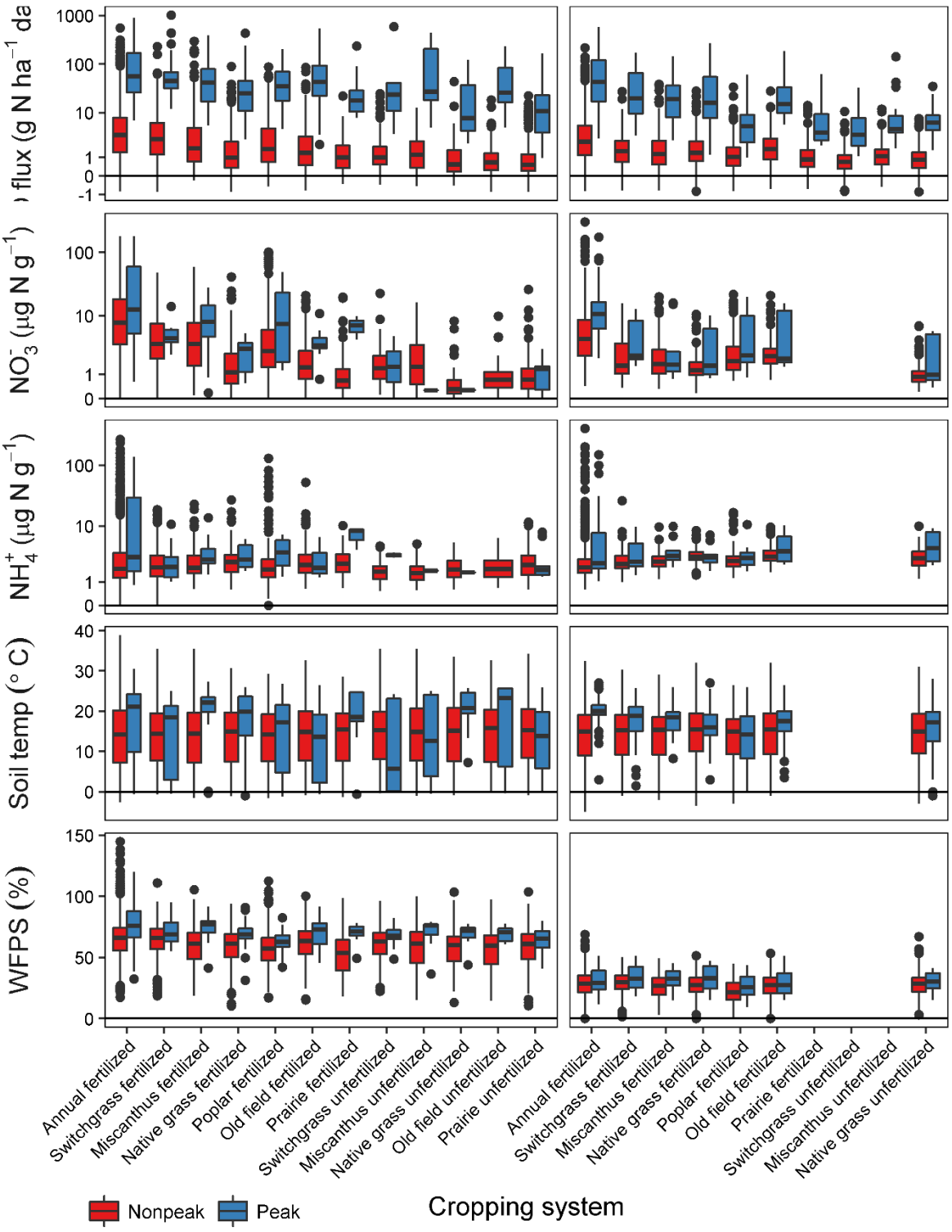

Figure 5.5 Comparison of N_2O fluxes and environmental parameters measured on dates with and without peak flux events. Within a system, all dates on which at least one sample recorded a peak flux were considered peak flux dates. Top three panels are IHS transformed. Boxes indicate 25th, 50th, and 75th percentiles.

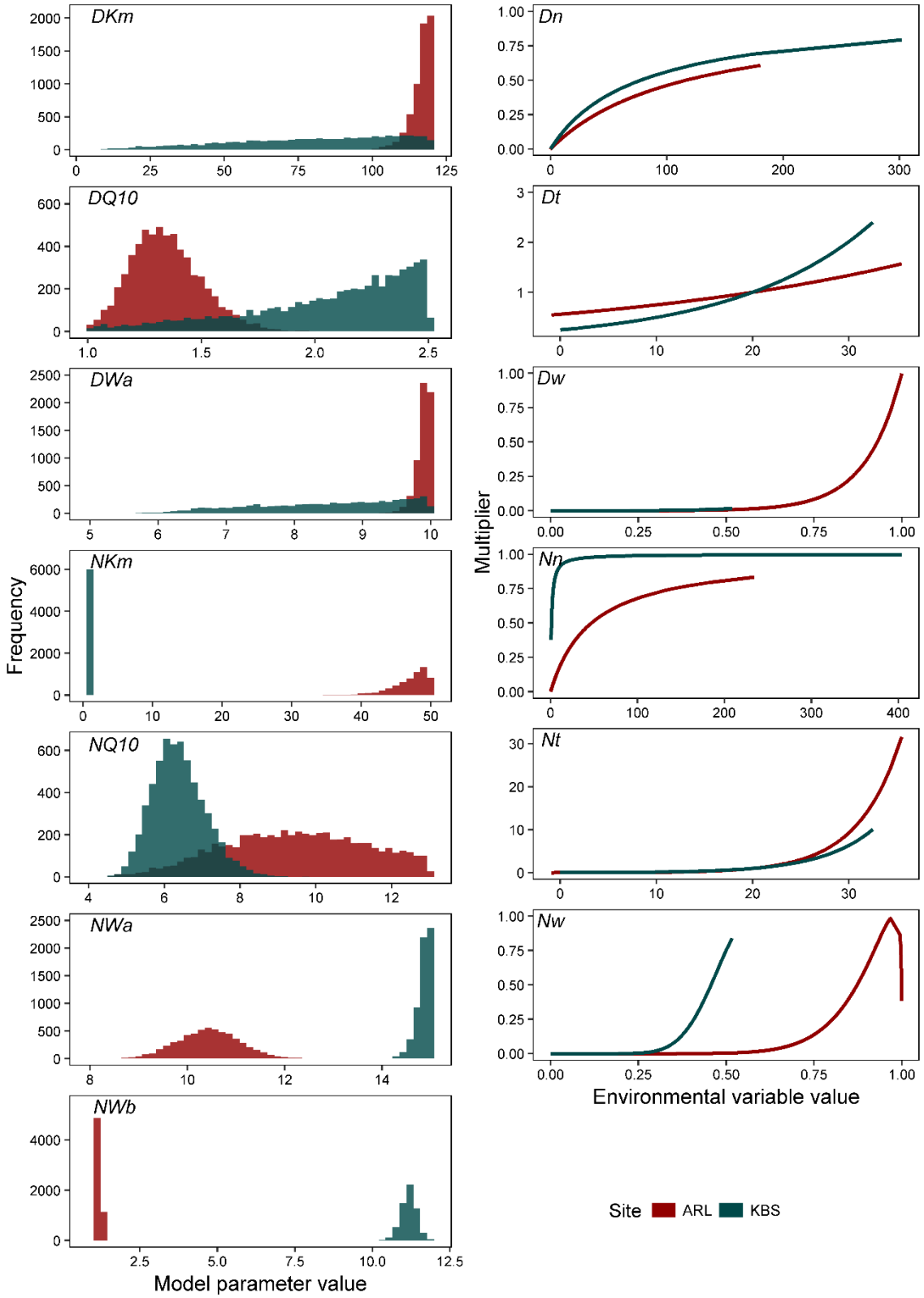

Figure 5.6 Posterior distributions of N_2O emissions model parameters (left), and of mean values for the resulting multiplicative constraints (right). Units and interpretations of parameters are given in **Table 5.3**.

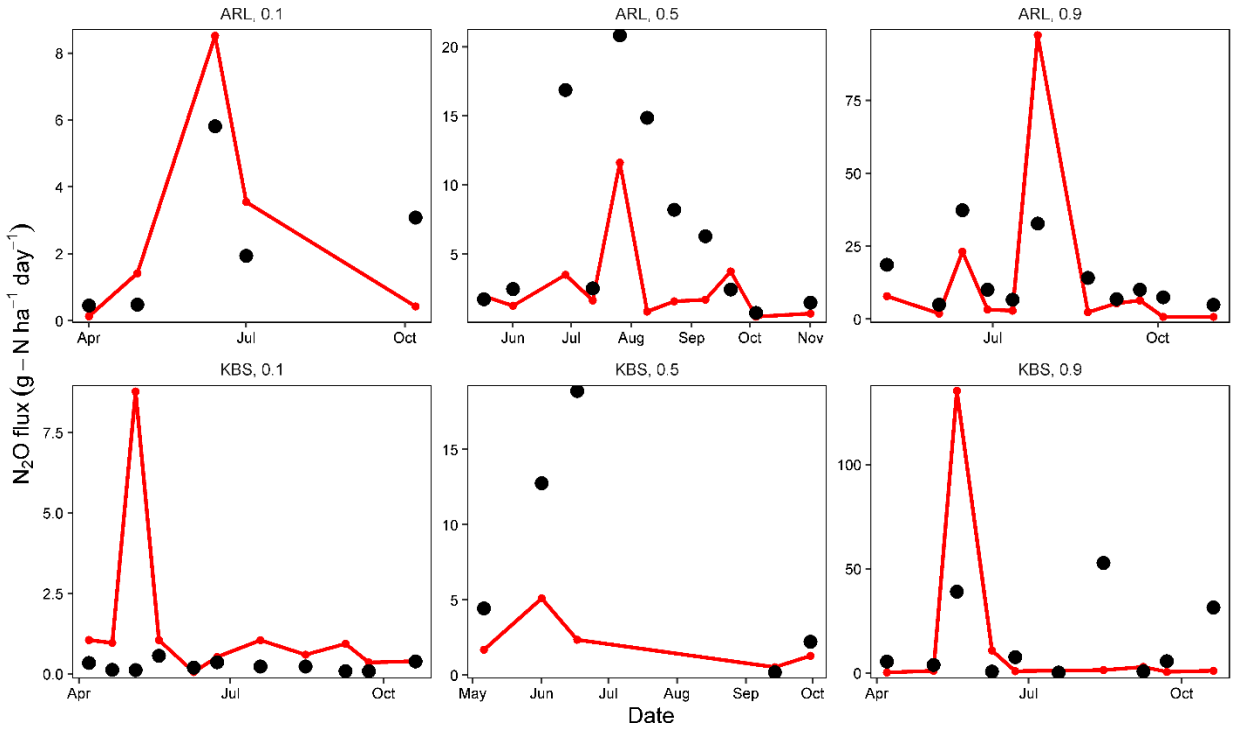

Figure 5.7 Individual plot-year examples of the accuracy of N_2O flux predictions obtained through model recalibration (black points) relative to observed values (red points and lines). Samples selected to represent the 10^{th}, 50^{th}, and 90^{th} percentile of sample-level root mean squared error for all plot-years from a site.

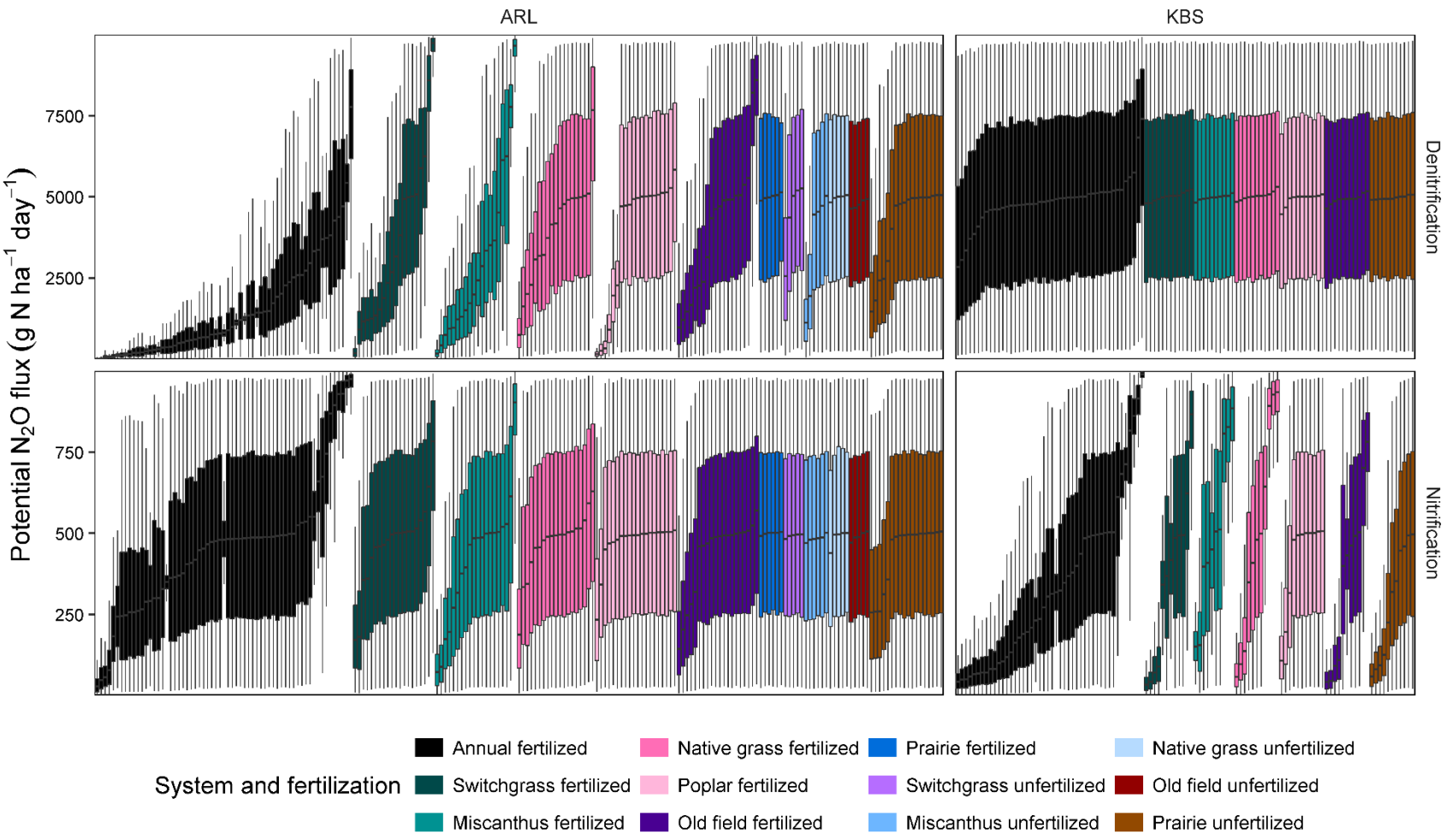

Figure 5.8 Posterior estimates of potential production of N_2O via denitrification and nitrification, based on N_2O emissions model recalibration. Distributions are based on 6000 samples, with boxes indicating 25th, 50th, and 75th percentiles. Within a cropping system and process, samples are sorted by ascending median values to facilitate visualization.

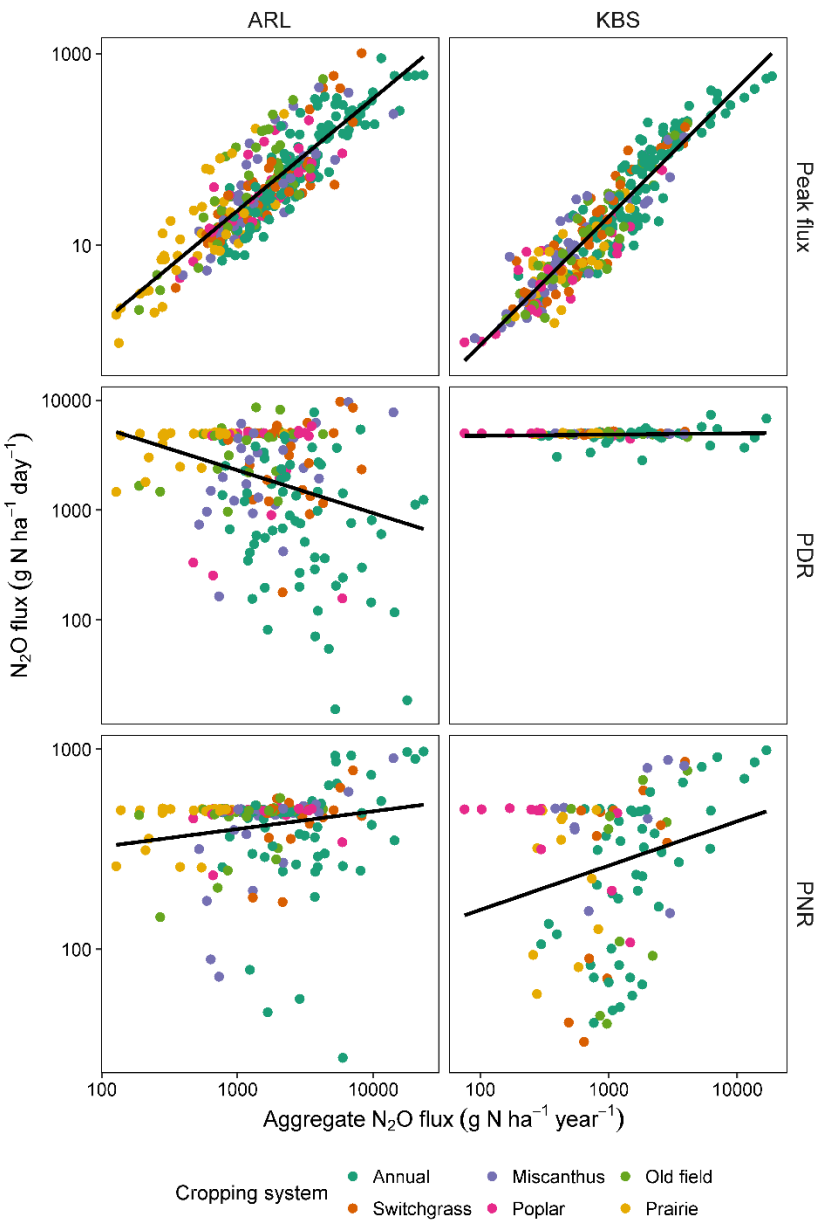

Figure 5.9 Correlation of aggregate annual N_2O emissions to peak fluxes and model-derived potential N_2O production from denitrification (PDR) and nitrification (PNR). Black lines indicate linear fits. All axes are plotted on a logarithmic scale.

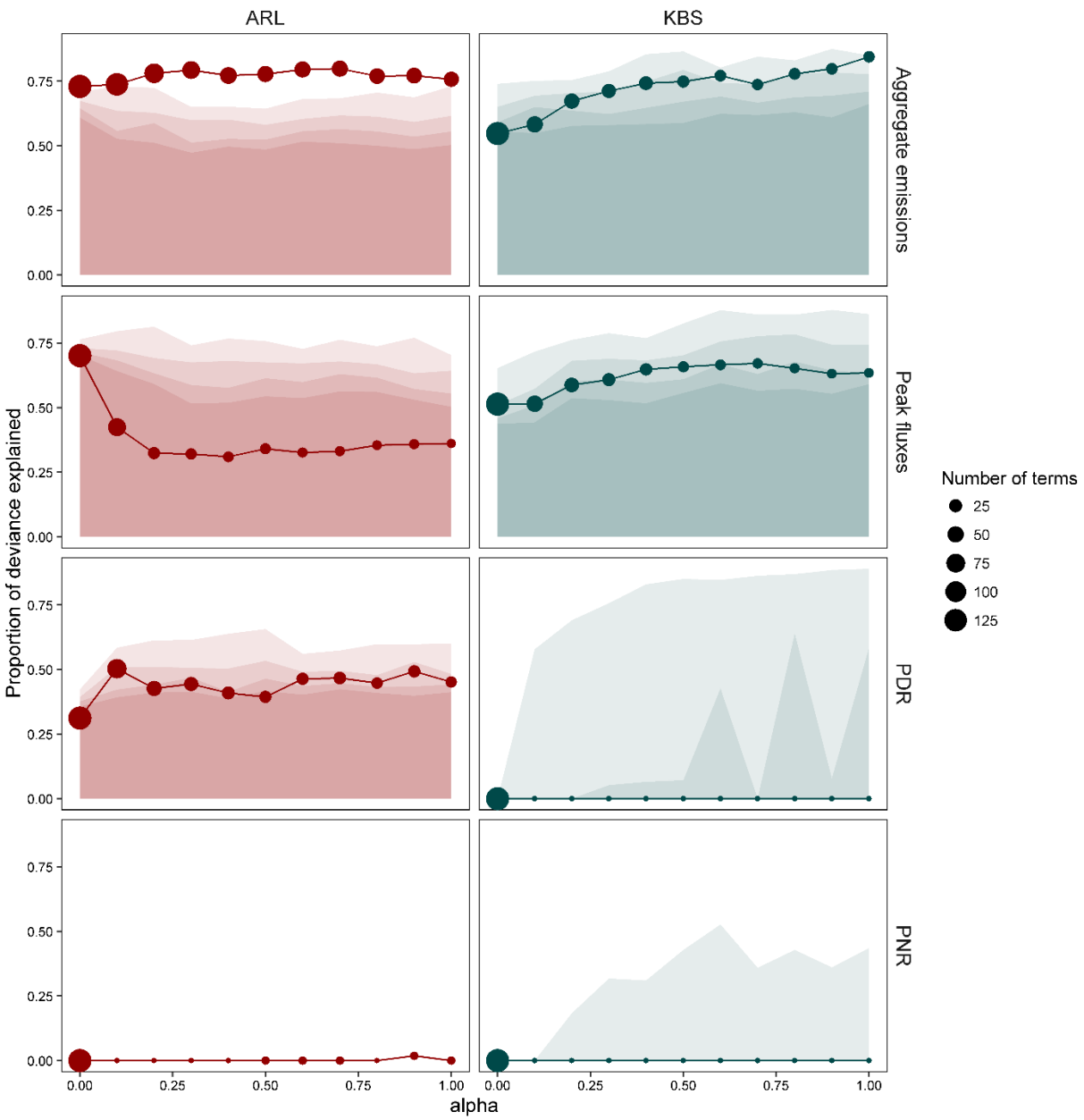

Figure 5.10 Deviance in estimators of N_2O production capacity explained by microbial functional gene profiles by elastic net modeling, as a function of the alpha regularization parameter. Point size was capped at 131 terms for display. Successively lighter shades indicate 80th, 85th, 90th, and 95th percentiles for deviance ratios for 200 permutations of response variables within cropping system and year. Sites were analyzed independently.

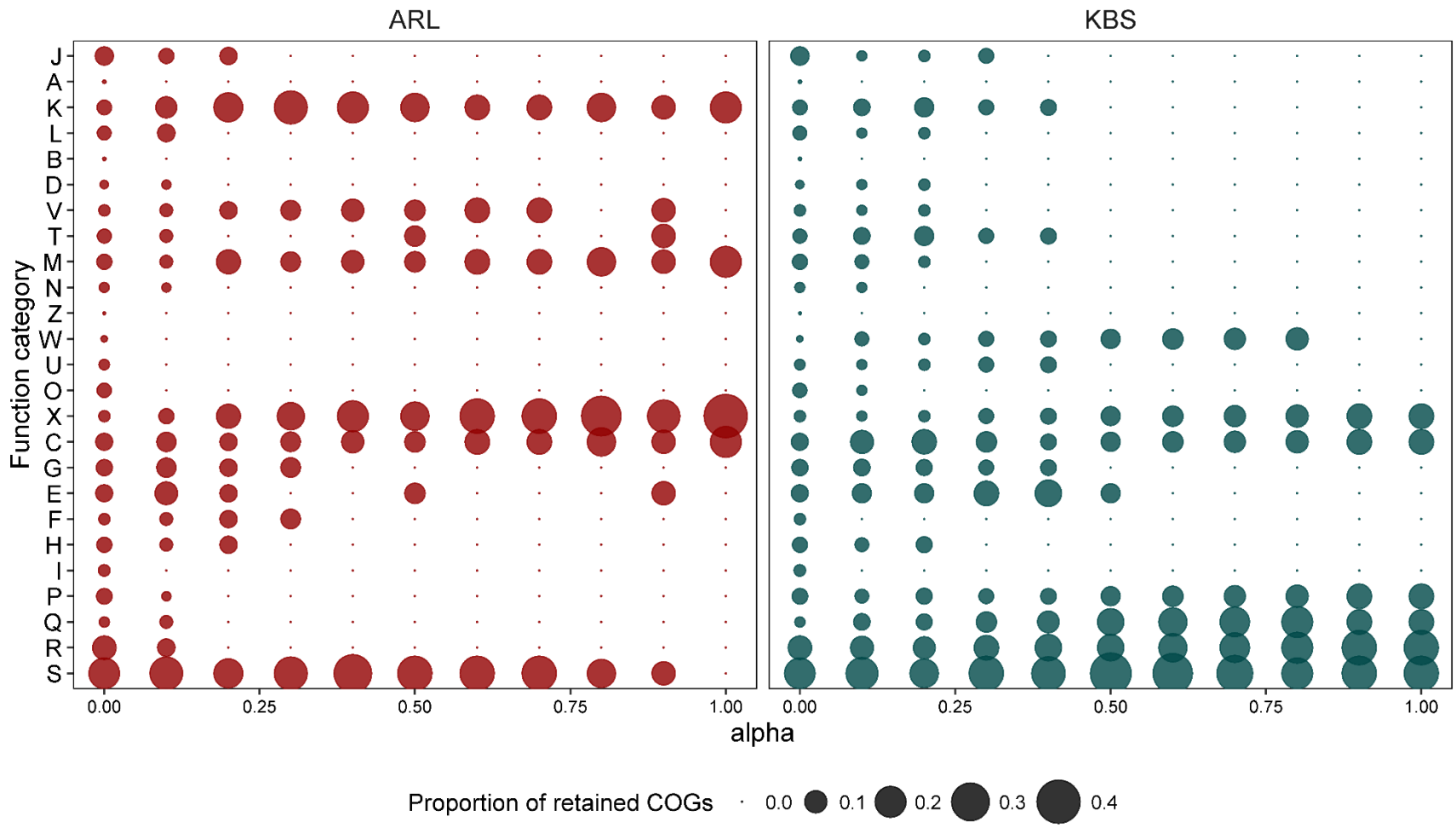

Figure 5.11 Proportional representation of function categories in COGs retained by elastic net models of aggregate N_2O emissions. See **Fig. 5.10** for actual number of terms retained at each level of alpha and **Table 5.4** for interpretation of function categories.

EPILOGUE

Next steps and publication plans

The preceding chapters were organized to try to present an overarching narrative about the effects of cropping systems on soil microbial communities and nitrous oxide (N_2O) production. Some reorganization of these analyses will be necessary to produce standalone narratives for publication.

The analysis of cropping system and environmental effects on N_2O production could combine the environmental constraints analyzed in Chapter 4 with the aggregate emission and peak flux analyses from Chapter 5. This dataset lends itself to two analyses. First, it allows comparison of N_2O emissions from a diverse array of potential biomass feedstock cropping systems. While similar data have been generated before (Roth et al., 2014; Toma et al., 2011), such studies include far fewer systems, frequently with a single example of attributes such as perenniality and polyculture. The greater breadth of systems in this dataset, with multiple instances of key attributes, illustrates the limited consistency of these effects across different cropping systems. Second, this allows for a more robust description of how environmental factors influence effects attributed to cropping systems. These combined analyses would provide data from a variety of systems, which would be useful for subsequent meta-analyses and modeling efforts, as well as concrete exploration of the interactions between cropping systems and their environmental conditions.

The microbial community data, by contrast, will benefit from further analysis prior to publication. As discussed in Chapter 3, batch effects influenced functional gene profiles, quite likely due to differences in annotation pipelines. Reanalyzing the data within a single pipeline is unlikely to drastically alter our findings; in an exercise not presented in this dissertation, we conducted a combined assembly with many of our samples, which produced results that did not differ qualitatively from those reported here. Nonetheless, a uniform annotation process is tractable and could clarify the influence of treatment factors on functional gene profiles by reducing any methodological noise we were unable to correct. Another benefit of this reanalysis is that it would enable us to employ KEGG orthology (KO) functional gene annotations, which is currently impossible for many of our samples. KO annotations are more

extensively curated and updated than the COG annotations we employed, and are contextualized within biochemical reactions (Kanehisa et al., 2016). Moreover, KEGG annotations support a more robust system for copy-number estimation (Manor and Borenstein, 2015), which could further improve the accuracy of our abundance profiling. All of this would allow us to place greater emphasis on individual annotations, leading to a more reliable interpretation of the functions retained through elastic net modeling, as in Chapter 5. Combining the description of cropping system effects on soil microbial communities with their connection to variability in N_2O emissions would produce an exciting story.

The N_2O production model recalibration does not appear worth pursuing further. There are sufficient deficiencies in the underlying model and in our ability to implement it with our extant dataset to make this analysis worth reporting. More generally, the approach we attempted here is of limited utility. N_2O production potential can be measured directly using denitrifying enzyme activity assays (Smith and Tiedje, 1979) or other methods, without the need for numerous field observations or the risk of failing to sample a sufficiently broad range of environmental conditions to accurately model the system. We are presently conducting such an analysis in a separate project, which looks at the effects of fertilization rates on nitrogen cycling in switchgrass.

Broader findings

Cropping system differences are less extreme and less consistent than expected

The underlying premise of the bioenergy cropping systems experiment was that meaningful, management-relevant differences existed among agroecosystems. Properties like perenniality, plant species diversity, or "natural" status have been hypothesized or demonstrated to influence productivity, nutrient cycling, suitability as habitat, and provision other ecosystem services (Daily et al., 1997; Meehan et al., 2011; Tilman et al., 1997; Werling et al., 2014). This reasoning led to an emphasis on perennial, preferably polycultural, systems as the ideal means of providing bioenergy feedstocks (Griffith et al., 2011; Robertson et al., 2008; Tilman et al., 2006). These properties influenced nitrogen cycling, including N_2O emissions, potentially even changing their response to environmental conditions

(Dechow and Freibauer, 2011; Niklaus et al., 2006; Oelmann et al., 2007; Palmborg et al., 2005). Microbial communities similarly responded to agroecosystem differences (Allison et al., 2005; He et al., 2008; Liang et al., 2012). Our initial studies of this system also indicated major cropping system differences in soil microbial community composition (Herzberger et al., 2014) and N_2O dynamics (Oates et al., 2016). In the analyses presented here, however, we consistently encountered weaker and less consistent cropping system effects than we anticipated.

Cropping systems lie at the heart of all five chapters in this dissertation, and in all cases they exerted significant effects on the property we studied, although frequently to a smaller extent than we expected. We anticipated that many cropping system effects would reflect ecological distinctions like perenniality and plant diversity, but this was rarely the case. In Chapter 1, we observed massive differences between the membrane lipid profiles of corn and prairie systems, but much smaller differences in their 16S rRNA and functional gene abundance profiles. In Chapter 3, functional gene abundances often varied as much within a cropping system as they did between different systems and relative differences among systems changed from year to year. In Chapter 4, we saw considerable overlap in the N_2O flux distributions of switchgrass and annual systems, as well as minimal system-level differences in N_2O flux responses to environmental constraints, despite differences in emissions factors among some of these systems (Duran et al., 2016). In Chapter 5, we saw considerable overlap in distributions of aggregate emissions for many systems, and similar overlap in the values of peak N_2O fluxes from those systems.

These findings, though unexpected, are plausible. Many of our expectations for the relative behavior of these cropping systems stems from pre-existing, established examples. In many cases, the decision to establish a particular system reflected soil and other ecosystem properties, which may have driven much of the variability we observed (Liang et al., 2013, 2012). The systems in this study were all grown in an agronomic trial framework that sought to minimize the influence of system-independent conditions (Sanford et al., 2016). Moreover, systems were managed following university extension recommendations, which among other things seeks to minimize undue environmental impacts. If

management had included mechanical disruption via tillage, manure application, or excessive fertilization we might have observed more extreme and consistent differences in N_2O emissions and microbial community composition. It is critical to note that limited cropping system effects did not result in limited variability in our responses. N_2O production and soil nitrogen concentrations varied over orders of magnitude within a system (Chapter 4, Chapter 5). While microbial community variability was less dramatic (Chapter 1, Chapter 3), there were nonetheless substantial differences in the abundances of specific genes and in the overall abundance profile.

The limited and inconsistent cropping system effects are frustrating from a management and policy perspective, as they indicate it will be difficult to predict outcomes and design straightforward policies based primarily on cropping system identity. From a research perspective, particularly an ecological one, this is tremendously exciting, as it implies uncharacterized sources of variation that may ultimately result in a greater dynamic range over which agroecosystem functioning can be influenced.

Cropping system effects manifest rapidly but remain small relative to interannual variability

We anticipated to observe a gradual, progressive emergence of differences among cropping systems. For perennial systems in particular, the conceptual model was of an initial establishment phase followed by a more stable phase, with nutrient cycling, productivity, and other system functions differing between the two (Oates et al., 2016). Soil carbon, changing over decades to centuries, provided a perfect example of the long-term dynamics we expected to observe (Sanford et al., 2012). From a microbial perspective, this expectation was supported by literature reports of long-term land use legacies outweighing current plant species compositions (Jangid et al., 2011) and ongoing directional microbial community shift for years following land use conversion (Allison et al., 2005; Jangid et al., 2010).

Yield data from BCSE cropping systems largely followed the expected patterns. Yields from most perennial systems increased over multiple years before stabilizing (Sanford et al., 2016). In contrast, cropping system differences were immediately visible in both the microbial and N_2O production data we analyzed, but any subsequent increase in that difference was lost amidst interannual variability. Microbial community metrics diverged significantly among systems two years after BCSE establishment, but the

magnitude of that divergence did not in subsequent years (Chapter 1, Chapter 3). Similarly, "establishment phase" N_2O dynamics did not differ from those observed in more "mature" systems (Chapter 4, Chapter 5), with the sole exception of KBS poplar, where we observed the expected progressive decrease in N_2O emissions until the system was disrupted. Cropping system effects manifested quickly and did not increase over the study period. At the same time, we observed considerable temporal variability. Relative differences among cropping systems were often inconsistent across years and measurements taken from individual plots correlated weakly over time; for many analyses, plot-years could be treated as independent entities. Overall, the systems we studied were characterized by high variability that was only partially explained by cropping system and interannual effects.

Microbial community properties reflect plot-level environmental variability

At the outset of this project, we expected our story would reflect gradually increasing cropping system influence on soil microbial communities and N_2O production dynamics. While we did observe cropping system influence, most of the variance we observed resided among plots with ostensibly similar environmental conditions and selective pressures.

This might plausibly have rsulted

It is easy to imagine complete independence between the broad distributions of N_2O emissions and the minimally-structured ordinations of microbial community properties, each generated by their own unmeasured stochastic processes or resulting from particular methodological issues. Instead, we found that these seemingly disorganized functional gene profiles contained indicators that accounted for much of the seemingly random plot-level variability in N_2O emissions. In Chapter 2, our analysis was clouded by the risk of spurious correlation driven by small sample sizes. In Chapter 5, we found a similar strength of relationships with a much larger, more heterogeneous dataset. These models were not built purely from microbial indicators, as many retained year- or system-specific terms, but microbial data were central to their capacity to explain variability within treatment groups. The functional genes identified through this approach did not immediately suggest a mechanism by which this relationship operated, and our analyses

could not have determined whether microbial communities drove the patterns of N_2O emissions we observed or merely served as sensitive, informative indicators of the actual drivers. However, the relationship we observed between the highly variable, seemingly noisy patterns of N_2O emission and microbial community composition suggests a strong link, whether direct or indirect, between the two. This link promises to be more complex and informative than we initially anticipated; it and its implications about the shaping of microbial communities and regulation of environmental processes would be very exciting.

References

Allison, V.J., Miller, R.M., Jastrow, J.D., Matamala, R., Zak, D.R., 2005. Changes in soil microbial community structure in a tallgrass prairie chronosequence. Soil Science Society of America Journal 69, 1412–1421.

Daily, G.C., Alexander, S., Ehrlich, P.R., et al., 1997. Ecosystem services: Benefits provided to human societies by natural ecosystems. Issues in Ecology 1, 2–16.

Dechow, R., Freibauer, A., 2011. Assessment of German nitrous oxide emissions using empirical modelling approaches. Nutrient Cycling in Agroecosystems 91, 235–254.

Duran, B.E.L., Duncan, D.S., Oates, L.G., Kucharik, C.J., Jackson, R.D., 2016. Nitrogen fertilization effects on productivity and nitrogen loss in three grass-based perennial bioenergy cropping systems. PLOS ONE 11, e0151919.

Griffith, A.P., Epplin, F.M., Fuhlendorf, S.D., Gillen, R., 2011. A comparison of perennial polycultures and monocultures for producing biomass for biorefinery feedstock. Agronomy Journal 103, 617–627.

He, X.-Y., Wang, K.-L., Zhang, W., Chen, Z.-H., Zhu, Y.-G., Chen, H.-S., 2008. Positive correlation between soil bacterial metabolic and plant species diversity and bacterial and fungal diversity in a vegetation succession on Karst. Plant and Soil 307, 123–134.

Herzberger, A.J., Duncan, D.S., Jackson, R.D., 2014. Bouncing back: Plant-associated soil microbes respond rapidly to prairie establishment. PLoS ONE 9, e115775.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Blair, J.M., Coleman, D.C., Whitman, W.B., 2010. Development of soil microbial communities during tallgrass prairie restoration. Soil Biology and Biochemistry 42, 302–312.

Jangid, K., Williams, M.A., Franzluebbers, A.J., Schmidt, T.M., Coleman, D.C., Whitman, W.B., 2011. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. Soil Biology & Biochemistry 43, 2184–2193.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research 44, D457–D462.

Liang, C., Duncan, D.S., Balser, T.C., Tiedje, J.M., Jackson, R.D., 2013. Soil microbial residue storage linked to soil legacy under biofuel cropping systems in southern Wisconsin, USA. Soil Biology & Biochemistry 57, 939–942.

Liang, C., Jesus, E. da C., Duncan, D.S., Jackson, R.D., Tiedje, J.M., Balser, T.C., 2012. Soil microbial communities under model biofuel cropping systems in southern Wisconsin, USA: Impact of crop species and soil properties. Applied Soil Ecology 54, 24–31.

Manor, O., Borenstein, E., 2015. MUSiCC: a marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. Genome Biology 16, 1–20.

Meehan, T.D., Werling, B.P., Landis, D. a, Gratton, C., 2011. Agricultural landscape simplification and insecticide use in the Midwestern United States. Proceedings of the National Academy of Sciences of the United States of America 108, 11500–11505.

Niklaus, P.A., Wardle, D.A., Tate, K.R., 2006. Effects of plant species diversity and composition on nitrogen cycling and the trace gas balance of soils. Plant and Soil 282, 83–98.

Oates, L.G., Duncan, D.S., Gelfand, I., Millar, N., Robertson, G.P., Jackson, R.D., 2016. Nitrous oxide emissions during establishment of eight alternative cellulosic bioenergy cropping systems in the North Central United States. GCB Bioenergy 8, 539–549.

Oelmann, Y., Wilcke, W., Temperton, V.M., et al., 2007. Soil and plant nitrogen pools as related to plant

diversity in an experimental grassland. Soil Science Society of America Journal 71, 720–729.

Palmborg, C., Scherer-Lorenzen, M., Jumpponen, A., Carlsson, G., Huss-Danell, K., Högberg, P., 2005. Inorganic soil nitrogen under grassland plant communities of different species composition and diversity. Oikos 110, 271–282.

Robertson, G.P., Dale, V.H., Doering, O.C., et al., 2008. Sustainable biofuels redux. Science 322, 49–50.

Roth, B., Finnan, J.M., Jones, M.B., Burke, J.I., Williams, M.L., 2014. Are the benefits of yield responses to nitrogen fertilizer application in the bioenergy crop *Miscanthus* × *giganteus* offset by increased soil emissions of nitrous oxide? GCB Bioenergy 7, 145–152.

Sanford, G.R., Oates, L.G., Jasrotia, P., Thelen, K.D., Robertson, G.P., Jackson, R.D., 2016. Comparative productivity of alternative cellulosic bioenergy cropping systems in the North Central USA. Agriculture, Ecosystems and Environment 216, 344–355.

Sanford, G.R., Posner, J.L., Jackson, R.D., Kucharik, C.J., Hedtcke, J.L., Lin, T.-L., 2012. Soil carbon lost from Mollisols of the North Central U.S.A. with 20 years of agricultural best management practices. Agriculture, Ecosystems & Environment 162, 68–76.

Smith, M.S., Tiedje, J.M., 1979. Phases of denitrification following oxygen depletion in soil. Soil Biology and Biochemistry 11, 261–267.

Tilman, D., Hill, J., Lehman, C.L., 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. Science 314, 1598–1600.

Tilman, D., Knops, J.M.H., Wedin, D., Reich, P.B., Siemann, E., 1997. The influence of functional diversity and composition on ecosystem processes. Science 277, 1300–1302.

Toma, Y., Fernández, F.G., Sato, S., et al., 2011. Carbon budget and methane and nitrous oxide emissions over the growing season in a Miscanthus sinensis grassland in Tomakomai, Hokkaido, Japan. GCB Bioenergy 3, 116–134.

Werling, B.P., Dickson, T.L., Isaacs, R., et al., 2014. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. Proceedings of the National Academy of Sciences of the United States of America 111, 1652–1657.