

2019

Relationships between soil microbial physiology, community structure and carbon and nitrogen cycling in temperate forest ecosystems

<https://hdl.handle.net/2144/34934>

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS & SCIENCES

Dissertation

**RELATIONSHIPS BETWEEN SOIL MICROBIAL PHYSIOLOGY,
COMMUNITY STRUCTURE AND CARBON AND NITROGEN CYCLING
IN TEMPERATE FOREST ECOSYSTEMS**

by

MUSTAFA SAIFUDDIN

B.A., University of Texas at Austin, 2013

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2019

Approved by

First Reader

Adrien C. Finzi, Ph.D.
Professor of Biology

Second Reader

Jennifer M. Bhatnagar, Ph.D.
Assistant Professor of Biology

ACKNOWLEDGMENTS

I am extremely grateful for the opportunities, guidance and support offered by my advisor, Adrien Finzi. I entered graduate school with limited experience, and Adrien was very patient in guiding me towards a dissertation topic at the intersection of our interests. I am especially thankful for the many ways in which he encouraged me to reach for opportunities I would have otherwise shied away from, and for providing me with confidence in learning a variety of new methods.

I have been extremely fortunate to have the advice of a talented committee. Jennifer Bhatnagar, my second reader, essentially provided me a second home in her lab, where I was able to learn new molecular techniques. I will be forever thankful for the times when I left Jenny's office with newfound enthusiasm and motivation for my dissertation. Pamela Templer, my committee chair, has been incredibly supportive throughout my graduate experience. Many of the ideas for my work on N cycling were developed while in her Forest Ecology class. Working with Michael Dietze has transformed the way I understand models and data, while making a number of once-intimidating tools more accessible to me. Daniel Segre has provided me with unique perspectives on microbial metabolism, and has been an exciting link to scientists interested in diverse types of microbiomes.

I am also indebted to a number of helpful people in the Finzi Lab. None of the lab assays I performed would have been possible without the skilled assistance of Marc-Andre Giasson, who seemed to always have the solutions to any problems I encountered. I am inspired by former fellow graduate students, Rose Abramoff, Patrick Sorensen, and Allison Gill, who made a wonderful, welcoming environment for me when I started as a new

student in the lab, and who remain valuable friends and mentors. I am thankful to I-Fang Hsieh, Ryan Quinn and Stephen Gougherty for their more recent helpfulness, friendship, and community.

I enjoyed working with a number of lab technicians and undergraduate students who made my time in the field and lab exciting, including Michael Dare, Alex Kerr, Rieka Yu, Ralina Karagenova, Sarah Goldsmith, Grace Bowden, Zachary Finzi, and Nina Finkhouse, among others. A portion of my field work was conducted during a brief but meaningful stay at the Phillips Lab, and I am indebted to a large number of extremely helpful people who guided me through the beautiful forests of central Indiana. Finally, I am grateful for the opportunity to be a part of several, lively communities at BU and beyond, including the EBE department, Biogeosciences program, and Harvard Forest.

**RELATIONSHIPS BETWEEN SOIL MICROBIAL PHYSIOLOGY,
COMMUNITY STRUCTURE AND CARBON AND NITROGEN CYCLING IN
TEMPERATE FOREST ECOSYSTEMS**

MUSTAFA SAIFUDDIN

Boston University Graduate School of Arts & Sciences, 2019

Major Professor: Adrien C. Finzi, Professor of Biology

ABSTRACT

Soil bacteria and fungi play a central role in the biogeochemical cycling of both carbon (C) and nitrogen (N) through terrestrial ecosystems. In the C cycle, soil microbial groups regulate the depolymerization of large stocks of soil organic matter and contribute 35-69 Pg C to the atmosphere annually through heterotrophic respiration. Soil microbial groups also mediate several important transformations of N, including making limiting nutrients available for uptake by plants through N-fixation, converting N between inorganic forms through nitrification, and returning N to the atmosphere through denitrification. While each of these functions is performed by soil microbes, scaling microbial physiology and community structure to biogeochemical cycling remains a significant research challenge. This dissertation integrates three distinct approaches to characterizing relationships between microbial physiology, microbial community structure and biogeochemical cycling. First, I explore the role of microbial physiology in C cycling by developing a novel method to predict bacterial carbon use efficiency (CUE) from genomes using metabolic modeling. I find that bacterial CUE is phylogenetically

structured, with the class and order levels explaining the greatest proportion of variance in CUE, and I identify particular bacterial traits that most strongly predict CUE. These findings highlight the importance of accounting for microbial physiology when modeling soil C cycling. Second, I explore how differences in the abundance and activity of microbial functional groups and their interactions with mycorrhizal fungi impact temperate forest N cycling. I find that N availability and rates of N-fixation, nitrification and denitrification are structured in relation to mycorrhizal fungal types, but that the abundances of bacterial functional groups are not correlated with biogeochemical fluxes. Finally, I use a soil biogeochemical model to identify sources of uncertainty and data needs in advancing our understanding of microbially-mediated soil biogeochemical cycling. I isolate specific microbial physiological and enzyme kinetic parameters that have disproportionately large impacts on projections of coupled C and N cycling, and I quantify the potential for particular types of data to help reduce uncertainties. Overall, this dissertation advances our understanding of how microbial processes impact the biogeochemical cycling of C and N in terrestrial ecosystems.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT.....	vi
TABLE OF CONTENTS.....	viii
LIST OF TABLES.....	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS.....	xiii
CHAPTER ONE: INTRODUCTION – BRIDGING FROM GENOMES TO GREENHOUSE GASES	1
CHAPTER TWO: BACTERIAL CARBON USE EFFICIENCY PREDICTED FROM GENOME SCALE METABOLIC MODELS SHOWS PHYLOGENETIC VARIATION WITH SIGNIFICANT BIOGEOCHEMICAL CONSEQUENCES	8
Abstract.....	8
Introduction.....	10
Methods.....	14
Results.....	23
Discussion	26
Tables and Figures	34

CHAPTER THREE: ECTOMYCORRHIZAL FUNGI SUPPRESS THE ACTIVITY BUT NOT THE ABUNDANCE OF SOIL BACTERIAL DIAZOTROPHS, NITRIFIERS AND DENITRIFIERS IN TEMPERATE FORESTS	46
Abstract	46
Introduction	48
Methods	51
Results	56
Discussion	59
Tables and Figures	66
CHAPTER FOUR: IDENTIFYING DATA NEEDS AND ASSESSING UNCERTAINTY IN A COUPLED MICROBIAL SOIL C AND N DECOMPOSITION MODEL	74
Abstract	74
Introduction	76
Methods	79
Results	85
Discussion	89
Tables and Figures	97
CHAPTER FIVE: CONCLUSIONS	111
BIBLIOGRAPHY	117
CURRICULUM VITAE	143

LIST OF TABLES

TABLE 2.1 — CUE FOR BIGG MODELS ACROSS SUBSTRATE TYPES.....	34
TABLE 2.2 — CONSTRAINED CUE SUMMARY.....	35
TABLE 2.3 — CONTRIBUTION INDICES FOR VARIATION IN POTENTIAL CUE BY TAXONOMIC LEVEL.....	37
TABLE 2.4 — REGRESSION SUMMARY FOR PREDICTORS OF POTENTIAL CUE.....	38
TABLE 3.1 — N POOLS AND FLUXES SUMMARY.....	72
TABLE 3.2 — N POOLS AND FLUXES GLM REGRESSION SUMMARY	73
TABLE 4.1 — DAMM-MCNIP PARAMETERS, UNITS, DEFAULT VALUES AND DEFINITIONS.....	98
TABLE 4.2 — PARAMETER ESTIMATES GIVEN FIELD OBSERVATIONS OF HETEROTROPHIC RESPIRATION.....	99

LIST OF FIGURES

FIGURE 2.1 — METABOLIC MODEL REACTION CLASSIFICATION	
FRAMEWORK	39
FIGURE 2.2 — HISTOGRAM OF POTENTIAL CUE ESTIMATES AND EMPIRICAL	
OBSERVATIONS ACROSS TAXA	40
FIGURE 2.3 — PHYLOGENETIC HEATMAP OF POTENTIAL CUE VALUES.	41
FIGURE 2.4 — BOXPLOTS OF CONSTRAINED CUE ACROSS SUBSTRATE	
TYPES	42
FIGURE 2.5 — REGRESSION OF MEAN CUE AGAINST GC CONTENT.....	43
FIGURE 2.6 — REGRESSION OF POTENTIAL CUE AGAINST GENOME SIZE ...	44
FIGURE 2.7 — MODELED C CYCLING VARIATION ACROSS CUE RANGE	45
FIGURE 3.1. — FRAMEWORK OF N CYCLING IN ECM AND AM STANDS.....	66
FIGURE 3.2 — BOXPLOTS OF SOIL N POOLS BY PLOT TYPE.	67
FIGURE 3.3 — SOIL N POOLS ACROSS GRADIENTS.	68
FIGURE 3.4 — BOXPLOTS OF N CYCLE FLUXES BY PLOT TYPE.....	69
FIGURE 3.5 — REGRESSIONS OF N CYCLE FLUXES AGAINST PROPORTION	
ECM.....	70
FIGURE 3.6 — BOXPLOTS OF GENE ABUNDANCES BY PLOT TYPE.....	71
FIGURE 4.1 — MAJOR POOLS, FLUXES AND EQUATIONS IN DAMM-MCNIP.	
.....	100
FIGURE 4.2 — HEATMAP SHOWING SENSITIVITY OF MODEL OUTPUTS TO	
PARAMETERS.	101

FIGURE 4.3 — MARGINAL PARAMETER UNCERTAINTIES FOLLOWING INCORPORATION OF OBSERVED RESPIRATION DATA. PRIOR DISTRIBUTIONS ARE UNIFORM SPANNING THE FULL RANGE OF X- AXES.	102
FIGURE 4.4 — MODEL PREDICTIONS COMPARED TO HETEROTROPHIC RESPIRATION DATA.	103
FIGURE 4.5 — UNCERTAINTY REDUCTIONS AND ACCURACIES FOR PARAMETERS ESTIMATED PROVIDED DIFFERENT TYPES OF DATA. ..	106
FIGURE 4.6 — MARGINAL PARAMETER UNCERTAINTIES FOLLOWING INCORPORATION OF SIMULATED DOC DATA.	107
FIGURE 4.7. – MODELED OUTPUTS USING MAP PARAMETER ESTIMATES.	108
FIGURE 4.8 – MODELED C OUTPUTS WITH PRIOR AND POSTERIOR ESTIMATES OF EA_{DEP}	109
FIGURE 4.9 – MODELED N OUTPUTS WITH PRIOR AND POSTERIOR ESTIMATES OF EA_{DEP}	110

LIST OF ABBREVIATIONS

AM	arbuscular mycorrhizal fungi
ANOVA	analysis of variance
C	carbon
CI	contribution index
COBRA	constraint-based reconstruction analysis
CO ₂	carbon dioxide
COG	clusters of orthologous groups
CTP	cytosine triphosphate
CUE	carbon use efficiency
DAMM	dual Arrhenius Michaelis Menten
DBH	diameter at breast height
DOC	dissolved organic carbon
DOCN	dissolved organic carbon and nitrogen
DOE kBase	Department of Energy knowledgebase
DON	dissolved organic nitrogen
ECM	ectomycorrhizal fungi
FBA	flux balance analysis
GC	guanine-cytosine
GLM	generalized linear model
grDW	grams dry weight
GTP	guanine triphosphate

GW	Griffy Woods
HF	Harvard Forest
IN	Indiana
LD	Lilly-Dickey Woods
MA	Massachusetts
MAP	maximum a posteriori
Mbp	mega base pairs
MCNiP	microbial carbon and nitrogen physiology
N.....	nitrogen
NH ₄	ammonium
NO ₃	nitrate
PGLS	phylogenetic generalized least squares
PH	Prospect Hill tract
qPCR.....	quantitative polymerase chain reaction
RAST	rapid annotation using subsystem technology
R _H	heterotrophic respiration
RMSE.....	root mean square error
SE.....	standard error
SM.....	simes tract
SOC.....	soil organic carbon
SOCN.....	soil organic carbon and nitrogen
SON	soil organic nitrogen

CHAPTER ONE:

INTRODUCTION – BRIDGING FROM GENOMES TO GREENHOUSE GASES

Soil bacteria and fungi play a central role in the biogeochemical cycling of both carbon (C) and nitrogen (N) through terrestrial ecosystems. Heterotrophic respiration, a direct product of microbial metabolism, is one of the largest fluxes of C in terrestrial environments, contributing 35-69 Pg C to the atmosphere annually as CO₂ (Tian et al. 2015). Additionally, the microbial production of extracellular enzymes to depolymerize soil organic matter is responsible for regulating large stocks of C in soil (Sinsabaugh et al. 2014). Microbial necromass and microbially-derived compounds can also be large components of soil organic matter themselves (Schmidt et al. 2011, Kögel-Knabner 2017). Thus, through the combined processes of respiration, decomposition, and cell death, microbes have the potential to regulate rates of soil C storage and release. Factors regulating the decomposition of these large stocks of C are critical to understand, particularly in the context of global change, which may alter the global distribution of C (Schmidt et al. 2011).

Microbes also mediate several important transformations of N. An estimated 97% of new N inputs in unmanaged terrestrial ecosystems are attributed to biological N₂-fixation by either symbiotic or free-living diazotrophs (Vitousek et al. 2002, Galloway et al. 2004, Reed et al. 2011). Soil microbial groups are involved in the conversion of fixed N between inorganic forms through the process of nitrification, which has important implications for the availability of N for uptake by plants and for N losses from soil (Barnard et al. 2005). Finally, soil microbial groups are also involved in returning N to the

atmosphere through denitrification, sometimes in the form of greenhouse gasses (Groffman and Tiedje 1989, Robertson and Groffman 2015).

While each of these functions is performed directly by soil microbes, scaling microbial physiology and community structure up to biogeochemical cycling remains a significant research challenge (Wieder et al. 2013, 2018, Li et al. 2014, Rocca et al. 2015). It is not well understood how variation among microbial taxa in C metabolism and N cycling activities impact estimates of the biogeochemical cycling of these elements. Soil biogeochemical models tend to collapse microbial metabolic and functional diversity into homogenous and functionally-redundant microbial pools, but these simplifying assumptions can have significant consequences for model predictions of biogeochemical cycling (McGuire and Treseder 2010, Li et al. 2014). With global change, it is particularly important to explore these relationships because microbial metabolism and the composition of soil microbial communities can shift under scenarios of warming, elevated CO₂, or altered nutrient availability, while the biogeochemical consequences of these shifts remain unclear (Allison and Martiny 2008, Shade et al. 2012, Berthrong et al. 2014, DeAngelis et al. 2015). This dissertation explores three distinct approaches to characterizing relationships between microbial physiology, microbial community structure and biogeochemical cycling.

Microbial physiology and C cycling

In Chapter 2, I explore the implications of potential physiological variation across bacterial taxa in the metabolism of C. Specifically, I focus on variation in carbon use

efficiency (CUE), which quantifies the proportion of consumed C retained in biomass, with the remainder potentially lost through respiration. CUE is used in several microbial biogeochemical models, and understanding variation in CUE is critical to improving our ability to model CO₂ emissions (Geyer et al. 2016). In particular, characterizing how CUE varies between bacterial taxa and across substrate types is important to accurately represent the interplay between microbial community structure and access to substrates in soil.

In this chapter, I implement a pioneering, *in silico* approach to predict bacterial CUE for over 200 bacterial species using a combination of literature synthesis and genome-specific constraint-based metabolic modeling. I show that intrinsic differences between taxa alone resulted in >300% variation in CUE, with anywhere between 20 to 90% of consumed C remaining in biomass. I find that bacterial CUE is phylogenetically structured, with the class and order levels explaining the greatest proportion of variance in CUE, and I identify particular bacterial traits that most strongly predict CUE. I integrate the range of variation in CUE observed across taxa into the soil C and N cycling model DAMM-MCNiP, and I find that phylogenetic variation in CUE can lead to large differences in projected estimates of C cycling. These findings highlight the importance of accounting for differences in physiology between bacterial taxa when modeling soil C cycling, and suggest that compositional shifts in microbial communities could have consequences for ecosystem-level biogeochemistry if they shift the physiological profile of the microbial community in favor of higher or lower CUEs.

Microbial communities and N cycling

While respiration is a ubiquitous metabolic pathway across bacterial taxa, specific microbial groups can show distinctive functions within the N cycle. N availability can be a primary constraint to terrestrial productivity, but quantifying N cycling rates remains challenging in temperate forests (Bormann 1993, Bormann et al. 2002, Yanai et al. 2013). Microbial processes account for major inputs, transformations and exports of N in soil; however, consistent relationships between microbial community structure, microbial activity and N cycle fluxes have not emerged from current research (Rocca et al. 2015). The presence of clearly delineated microbial functional groups in N cycling provides a unique framework for exploring linkages between biogeochemical function and community structure (Levy-Booth et al. 2014).

The availability of N in soil is dependent on both abiotic and biotic factors, such as competition for N between plant roots, mycorrhizal fungi, and free-living bacteria involved in N-fixation, nitrification and denitrification. In Chapter 3, I explore how differences in the abundance and activity of microbial functional groups impact temperate forest N cycling. I characterize N-fixer, nitrifier, and denitrifier bacterial functional group abundances and activity in relation to biotic interactions with mycorrhizae and plant roots at four temperate forest sites in Massachusetts and Indiana. I find that mycorrhizal type plays a significant role in structuring the availability of N as well as microbial N-cycling activity. Specifically, I find evidence for suppression of N-fixation, nitrification and denitrification in forest stands dominated by ectomycorrhizal fungi. However, I do not find structuring of bacterial functional groups across gradients of mycorrhizal abundance,

suggesting a disconnect between the presence and activity of these groups. These findings indicate that the biogeochemical cycling of N in temperate forests is subject to different controls depending on mycorrhizal type, but that these may not have consistent impacts on the functional group composition of bacterial groups. I also suggest potential mechanisms by which community structure might be decoupled from function, and I consider improvements to sampling design and intensity that would be helpful for detecting these relationships in the future.

Microbial processes in coupled C and N cycling

The cycling of C and N are closely coupled in terrestrial ecosystems due to the incorporation of N into biomass constituents including certain enzymes, proteins and nucleic acids. Additionally, N is commonly a limiting nutrient for primary productivity, thereby regulating potential rates of C sequestration (LeBauer and Treseder 2008). Thus, it is important to consider the cycling of these elemental cycles in a coupled framework. Microbial processes represent a key linkage across these processes, as soil bacteria and fungi produce extracellular enzymes that depolymerize C and N from soil organic matter and stoichiometrically integrate C and N into biomass (Sinsabaugh et al. 2014, 2015). As extracellular enzymes produced by microbes directly impact rates of soil C sequestration, it is important to accurately represent extracellular enzyme kinetics.

Incorporating microbial parameters may substantially reduce uncertainties in global climate models (Todd-Brown et al. 2012, Wieder et al. 2013), which currently often collapse microbial community composition and physiology into a “black box” (McGuire

and Treseder 2010). Despite these simplifying assumptions, microbial-based biogeochemical models still have numerous parameters, and the empirical basis or confidence in these parameter selections is often unclear (Manzoni and Porporato 2009). Differences in model structure as well as parameter selections can lead biogeochemical models to make wildly different predictions in response to global change, even when they agree in the short term (Allison et al. 2010, Friedlingstein et al. 2014, Li et al. 2014, Sulman et al. 2018, Wieder et al. 2018), thus it is necessary to explicitly quantify uncertainty associated with specific parameters.

In Chapter 4, I continue to use the model DAMM-MCNP as a tool to explore how microbial processes impact coupled C and N cycling. In this chapter, I focus on exploring how parameter uncertainty effects our ability to use biogeochemical models as tools for predicting fluxes and identify which types of data would be most useful targets for collection to reduce uncertainty in model predictions. I find that particular microbial parameters (e.g. CUE) can be fairly tightly constrained from respiration data alone, while other microbial physiological and enzyme parameters require additional data on microbial biomass or dissolved organic C and N pool sizes to be constrained. I demonstrate how small shifts in a few key parameters representing microbial physiology and extracellular enzyme activities can have disproportionately large impacts on estimates of coupled C and N cycling.

Collectively, the work in this dissertation advances our understanding of how metabolic processes at the level of individual microbes and microbial interactions at the community level impact biogeochemical cycling. Continuing research on soil microbial

biogeochemistry will require integrating approaches across each of these scales, from genome-scale metabolic modeling, to field-based research on microbial community structure and function, to ecosystem modeling of coupled C and N cycling.

CHAPTER TWO: BACTERIAL CARBON USE EFFICIENCY PREDICTED FROM GENOME SCALE METABOLIC MODELS SHOWS PHYLOGENETIC VARIATION WITH SIGNIFICANT BIOGEOCHEMICAL CONSEQUENCES

Abstract

Respiration by soil bacteria and fungi is one of the largest fluxes of carbon (C) from the land surface, producing 35-69 Pg C/yr in the form of CO₂. Although this flux is a direct product of microbial metabolism, controls over metabolism and their responses to global change are a major uncertainty in the global C cycle. Here, we implement a pioneering, *in silico* approach to predict bacterial C-use efficiency (CUE) for over 200 species using genome-specific constraint-based metabolic modeling. We show that intrinsic differences between taxa alone resulted in >300% variation in CUE (0.2-0.9). CUE has a significant phylogenetic signal, with variation among taxa structured primarily at the class and order levels. CUE is negatively correlated with GC content and with genome size, reflecting a tradeoff between efficiency and access to diverse substrate types. Taxa with larger genomes access a wide range of C sources at the expense of efficiency compared to taxa with smaller genomes, which may represent more specialist trophic strategies. We also find strong evidence that substrate chemistry and supply rate alter apparent microbial CUE. Under C-replete conditions, mean potential CUE is 0.62 ± 0.17 whereas substrate-limiting scenarios decrease CUE to 0.29 ± 0.19 . We incorporate the full range of CUE values into a next-generation model of soil biogeochemistry and show that differences among taxa alter estimates of soil C storage by 200% and microbial biomass by 400% over 100 year

projections. When substrate availability and diversity is altered under global change, specialist bacterial taxa with small, C-use efficient genomes may be replaced by generalist taxa with larger genomes whose CO₂ emissions enhance anthropogenic feedbacks to climate change.

Introduction

Soil respiration is one of the largest exchanges of carbon (C) from the land surface to the atmosphere, releasing an estimated 98 ± 12 Pg C/yr from soil as CO₂ (Bond-Lamberty and Thomson 2010, Hashimoto et al. 2015). Heterotrophic respiration (R_H) by soil bacteria and fungi can account for a large proportion of total global soil respiration (35-69 Pg C/yr; Bond-Lamberty et al. 2004, Tian et al. 2015). Despite their central importance, estimates of global R_H vary widely with large uncertainty (Tian et al. 2015). Although this flux is a direct product of microbial metabolism, controls over physiology and their responses to global change are a major uncertainty in the global C cycle (Allison et al. 2010, Li et al. 2014).

Soil respiration rates are determined in part by microbial physiology, which controls the partitioning of C consumed by soil bacteria and fungi between respiration, biomass production and other processes. Following assimilation into microbial biomass, C can subsequently be sequestered in soil as necromass and soil organic matter, while C lost through respiration directly alters greenhouse gas concentrations. Quantifying how specific bacterial taxa allocate C between each of these pathways is critical to understanding relationships between physiology, community composition and soil C cycling.

Carbon use efficiency (CUE) measures the partitioning of C between microbial biomass and respiration (Geyer et al. 2016). Empirical estimates of microbial CUE range from near zero to over 0.8 (Keiblinger et al. 2010, Manzoni et al. 2012), while most biogeochemical models use a fixed value selected between 0.15 and 0.6 (Manzoni et al. 2012). Some of the variation observed in CUE may be attributed to the sensitivity of CUE

to abiotic factors such as temperature and pH (Manzoni et al. 2012, Sinsabaugh et al. 2016). However, an additional and often neglected source of variation in CUE may be due to physiological differences between soil microbial groups and their differential capacities for accessing particular substrate types (Bölscher et al. 2016, Geyer et al. 2016, Roller et al. 2016). This single parameter in microbial biogeochemistry models has direct impacts on estimates of greenhouse gas emissions and terrestrial C storage (Allison et al. 2010, Li et al. 2014), making it necessary to survey how CUE varies both among taxa and across substrate types.

Characterizing microbial C metabolism is particularly important in the context of global change, which may alter the structure and activity of microbial communities and their access to substrates (Leff et al. 2015). Previous work on functional and physiological variability suggests that defining bacterial taxa along a spectrum of copiotrophy (fast-growing, adapted to high substrate availability) to oligotrophy (slow-growing, adapted to limiting resource concentrations) may be one useful approach for understanding how groups respond to changes in temperature and resource availability (Fierer et al. 2007, Sistla et al. 2014). Classification schemes based on trophic strategy may be useful from a biogeochemical perspective if differences in growth strategies correspond to variation in CUE. For example, copiotrophs are hypothesized to show lower CUE than oligotrophs (Roller and Schmidt 2015, Roller et al. 2016), and this could potentially alter the CUE of microbial communities observed to shift toward a greater proportion of copiotrophic bacteria in response to global change manipulations that increase substrate availability, such as soil warming (Frey et al. 2008, Fierer et al. 2012).

Observations from global change experiments and genome-based estimates of minimal generation times support the classification of particular phyla, such as Acidobacteria as oligotrophic and Actinobacteria as copiotrophic (Leff et al. 2015). However, these phylum-level classifications are not consistent across studies (Bernard et al. 2007, Bastian et al. 2009). An improved understanding of the phylogenetic structure of microbial biogeochemically-relevant traits is needed to identify how microbial community structure impacts C cycling (Morrissey et al. 2016, 2017). Many functional genes show strong conservation within prokaryotes, leading to the possibility for strong phylogenetic structure in functional traits, particularly those that emerge from the coordinated activity of multiple genes (Martiny et al. 2013). Certain quantitative bacterial traits such as growth rates in the presence of labile C show strong phylogenetic signals; however, other traits such as responses to priming show shallow phylogenetic signals (Morrissey et al. 2016, 2017). Thus, the level of phylogenetic resolution required to characterize variation in CUE across bacterial taxa remains unclear.

In addition to understanding the phylogenetic structure of variation in CUE, it may be useful to explore whether particular genomic traits predict CUE. For example, copy numbers of ribosomal RNA operons are inversely related to growth efficiency in bacteria, providing a method for predicting growth efficiencies from genomes (Roller et al. 2016). Similarly, genomic traits have been useful for predicting microbial trophic strategies and biogeography, with bacterial taxa with larger genomes occupying a wider range of habitat types (Barberán et al. 2014) and dominating communities where resources are available in diverse forms but limiting concentrations (Konstantinidis and Tiedje 2004). Comparable

efforts for predicting CUE from genomic traits are necessary to help overcome challenges with measuring taxa-specific CUE for highly-diverse soil bacterial communities.

In the environment, microbial taxa are exposed to variation in substrate chemistry and availability, which can impact rates of C uptake and growth. These abiotic factors are likely to interact with intrinsic differences in physiology between individual taxa to ultimately determine CUE. For example, observations from bacterial cultures show that CUE increases with limiting resource concentration and with the free energy content of available resources (Roller and Schmidt 2015). These patterns are overlaid with differences between taxa, with potentially oligotrophic groups showing less responsivity to limiting resource availability than copiotrophic taxa (Roller and Schmidt 2015). Thus, estimates of CUE must consider both biotic and abiotic sources of variability, including bacterial physiology, substrate availability and substrate chemistry.

Prior work on estimating CUE is limited to a small set of individual microbial taxa, or involves mixed, whole communities (Roller and Schmidt 2015, Bölscher et al. 2016, Sinsabaugh et al. 2016). Direct measurements of CUE have been made using a wide range of methods including calorimetry (Bölscher et al. 2016) and stable isotope approaches (Dijkstra et al. 2015, Morrissey et al. 2017). CUE has also been estimated indirectly for whole communities based on environmental variables such as resource stoichiometry (Sinsabaugh et al. 2013, 2016). These methods can lead to CUE estimates that vary by a factor of two or more, making direct inter-comparisons challenging (Sinsabaugh et al. 2016).

Using a consistent methodology to measure CUE across a broad range of microbial taxa is necessary to determine how physiological variation in resource use between taxa impacts CUE. Metabolic models of bacterial physiology can be generated from annotated genomes (Thiele and Palsson 2010) and can be used to estimate taxa-specific biological fluxes, including biomass growth and C uptake (O'Brien et al. 2015). Here, we introduce a pioneering, *in silico* approach to generate theoretical predictions of CUE for over 200 taxa using genome-scale constraint-based metabolic modeling. We use these draft predictions to explore phylogenetic variation in bacterial CUE, identify particular genomic traits correlated with CUE, and explore the implications of phylogenetic variation in CUE on ecosystem-level estimates of C cycling.

Methods

Metabolic Modeling

Genome-scale metabolic modeling (also known as stoichiometric modeling, or constraint-based modeling) can be used to quantitatively analyze the complete set of metabolic reactions in an organism. This approach has been successfully used to represent bacterial metabolism and growth patterns (Henry et al. 2010, Sridhara et al. 2014), uptake and secretion (Varma and Palsson 1994, Edwards et al. 2001) and complex community interactions (Harcombe et al. 2014) *in silico*. The metabolic model for a given organism can be generated by extracting the list of all biochemical reactions available to an organism from its annotated genome. In addition to intracellular reactions, the model includes exchange reactions, which involve uptake or secretion of metabolites, either through

genome-encoded transporters, or expected free diffusion through the membrane. For convenience of subsequent mathematical analysis, this list is converted into a stoichiometric matrix, $\mathbf{S}$, whose element S_{ij} corresponds to the stoichiometric coefficients of metabolite i in reaction j . Due to incomplete genome annotations, gapfilling is often required to supplement models with additional reactions before models are capable of producing a nonzero biomass flux.

Genome-scale metabolic models were selected for analysis from two separate databases. Thirteen microbial models were selected from the Biochemically, Genetically and Genomically structured knowledgebase of metabolic reconstructions (BiGG), which contains a small set of manually-curated metabolic models from diverse environments (King et al. 2016). We utilized the following thirteen taxa in this analysis: *Clostridium ljungdahlii* DSM 13528, *Staphylococcus aureus* subsp. *aureus* N315, *Saccharomyces cerevisiae* S288c, *Methanosarcina barkeri* str. *Fusaro*, *Bacillus subtilis* subsp. *subtilis* str. 168, *Thermotoga maritima* MSB8, *Synechocystis* sp. PCC 6803, *Escherichia coli* str. K-12 substr. MG1655, *Shigella boydii* Sb227, *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* str. LT2, *Klebsiella pneumoniae* subsp. *pneumoniae* MGH 78578, *Geobacter metallireducens* GS-15, and *Mycobacterium tuberculosis* H37Rv. Mean CUE was calculated from CUE on growth on each of the following C-sources individually: D-Glucose, Fumarate, Acetate, Acetaldehyde, 2-Oxoglutarate, Ethanol, Formate, D-Fructose, L-Glutamine, L-Glutamate, D-lactate, L-Malate, Pyruvate, Succinate. Due to the limited number of manually-curated microbial metabolic models, we expanded our approach to

include models generated using automated pipelines (*described below*) for over 200 bacterial taxa from phyla commonly observed in soil environments.

Genome selection

The BiGG database primarily includes microbial models associated with the human microbiome, limiting our capacity to extrapolate our findings from these well-curated metabolic models to environmental microbial communities. We expanded our approach to target bacterial genomes belonging to phyla commonly observed in soil environments (Fierer et al. 2012, DeAngelis et al. 2015), which are of particular interest due to their major contributions to global respiration. We queried the Department of Energy's kBase for over 200 taxa and used automated pipelines to construct a large set of draft metabolic models.

The Department of Energy systems biology knowledgebase (kBase; Arkin et al. 2018) was searched in March 2016 for bacterial genomes belonging to phyla that have been observed to dominate forest soil bacterial community composition based on 16S ribosomal RNA and DNA sequencing (Fierer et al. 2007, DeAngelis et al. 2015). A total of 23,530 genomes belonging to the six selected phyla were identified in kBase, corresponding to 1,064 unique genera. For each phylum, at least 25 genomes were selected for analysis. For phyla with more than 50 available genomes, the full list of unique genera was scanned to target genera that have been observed in soil environments when possible.

The Build Metabolic Model tool was used in kBase to generate metabolic models from 231 selected genomes (Henry et al. 2010). Model construction in kBase involves functional annotation of the genome to identify metabolic genes and their associated biochemical

reactions using the *Rapid Annotation* of microbial genomes using *Subsystems Technology* (RAST) genome annotation pipeline and the model SEED framework (Henry et al. 2010, Overbeek et al. 2014). Draft metabolic models were gapfilled using the Gapfill Metabolic Model tool in kBase to add the minimal set of reactions required to produce biomass on complete media, which contains all possible metabolites available for uptake (Latendresse 2014). Gapfilling on complete media results in conservative gapfilling by assuming that metabolites necessary for growth but not produced intracellularly based on genome annotation are available in the environment.

Flux balance analysis

Flux balance analysis (FBA) allows for an estimation of metabolic fluxes, such as rates of C uptake and utilization, through a metabolic model based on linear optimization of a specified objective function, such as biomass production. FBA makes a steady-state assumption, circumventing the need for knowledge of kinetic parameters, and uses the stoichiometry of metabolic reactions to determine the feasible space of all possible combinations of reaction rates. By prescribing an optimization scheme, it is possible to identify specific points in this feasible space, resulting in putative predictions of all metabolic reaction rates in the organism, including uptake and secretion fluxes and growth. This approach requires specification of (1) a flux or set of fluxes to maximize (or minimize) and (2) upper and lower bounds for all reactions within the metabolic model. Upon specification of these inputs, FBA is able to estimate the particular combination of fluxes through all reactions in the model that satisfy the given conditions. FBA was performed in

MATLAB R2014a using the *optimizeCbModel* command in the *COntstraint-Based Reconstruction and Analysis* (COBRA) Toolbox (Schellenberger et al. 2011). All FBA analyses were set to maximize bacterial biomass production in this analysis, in accordance with standard FBA assumptions (Harcombe et al. 2014).

C use efficiency

C use efficiency (CUE) is calculated as the proportion of C retained in biomass relative to total C uptake (Eq. 1). For a metabolic model with n exchange reactions, and where C is equal to the number of C atoms taken up or secreted in a given reaction:

$$\text{CUE} = \frac{\sum_1^n \text{Uptake Flux}_i * C_i - \sum_1^n \text{Secretion Flux}_i * C_i}{\sum_1^n \text{Uptake Flux}_i * C_i} \quad (\text{Equation 2.1})$$

For the set of manually-curated models, the availability of one of 14 individual C sources was manipulated, and CUE was calculated under exclusive uptake of each metabolite separately. For the larger set of models from kBase, CUE was explored under two scenarios. (1) potential CUE was calculated by allowing a model to utilize all exchange reactions present, and (2) constrained CUE was calculated by limiting the availability of a single C-containing metabolite relative to the availability of all other metabolites. Potential CUE was calculated to explore intrinsic metabolic variation in CUE, and these values were most useful for comparisons between taxa and for identifying relationships between genome traits and CUE. All uptake reactions present in a model were made available for uptake by allowing for a default maximum flux of 1000 mmol/(grDW·hour), where grDW indicates the cellular biomass dry weight in grams. As CUE was calculated as a ratio of fluxes, values were not sensitive to the order of magnitude of maximum flux bounds as

long as these were consistent across reactions. Certain models produced a respiration flux of 0 mmol/(grDW·hour) and were excluded from subsequent analyses of CUE.

To calculate CUE under conditions of limited substrate availability, reactions in each metabolic model were first classified according to the following hierarchy: (1) *exchange*, (2) *C-containing*, (3) *utilized* when potential, (4) *essential* to biomass production, and (5) *constraining* to biomass production (Figure 2.1). For a given model, all C-containing exchange reactions with a nonzero flux under potential conditions were classified as *utilized*. The maximum uptake flux for each individual *utilized* reaction was then set to 0 mmol/(grDW·hour) and FBA was performed again to identify reactions that were *essential* for biomass production. Finally, maximum uptake for all *essential* reactions was individually set to 5% of the maximum uptake flux for all other metabolites (50 mmol/(grDW·hour)), and FBA was performed again to detect the impact of constraining particular *essential* reactions. Reactions that resulted in a reduction of the biomass flux by at least 5% were classified as *constraining*, meaning that the biomass production flux showed a direct response to the availability of metabolites dictated by these reactions.

Uptake fluxes for the most commonly occurring *constraining* reactions across all models were analyzed to determine the response of biomass production relative to availability for each metabolite (*biomass/uptake*). For 18 of the most commonly *constraining* reactions, the uptake flux corresponding to a 75% reduction in the biomass flux was identified for each model. For all models containing a given *constraining* reaction, FBA was performed after setting the maximum uptake flux for the *constraining* reaction to this reduced value while leaving all other exchange reaction fluxes potential.

Constrained CUE was then calculated according to equation 2.1. Constrained CUE was compared to potential CUE for all models with a given *constraining* reaction using paired T-tests and Cohen's D calculated from the *lsr* package (Navarro 2015) in R Studio (R Core Team 2017).

Model evaluation and empirical comparisons

To test the sensitivity of our results to the method of gapfilling, two parallel sets of models were constructed for each taxon. One set of models were gapfilled to achieve a minimum biomass flux of 0.1 mmol/(grDW·hour), while a second set was more heavily gapfilled to achieve a (default) minimum biomass flux of 1000 mmol/(grDW·hour). A total of 246 exchange reactions, including 211 C-containing exchange reactions, were observed across 231 models gapfilled to the lower biomass threshold. A total of 318 exchange reactions, including 279 C-containing exchange reactions, were observed across 231 models gapfilled to the higher biomass threshold. On average, models gapfilled to the higher biomass threshold had only 8 additional C-containing exchange reactions. Gapfilling intensity had a significant impact on subsequent calculations including CUE, but inter-model comparisons and rank order were not strongly affected by gapfilling intensity. Potential CUE values calculated from models gapfilled at the two intensities were strongly correlated (Pearson's rank correlation coefficient = 0.7).

To compare our predictions with empirical observations, we assembled a database of 130 published observations of CUE, growth yield, or percent assimilation for over 40 microbial taxa grown on a variety of substrate types. Several of these observations come

from a synthesis of growth yields (Payne 1970). To convert from growth yield to CUE, we assumed 50% biomass C content and calculated C uptake based on the limiting substrate reported.

Phylogenetic analyses

The Build Phylogenetic Tree tool was used in the DOE kBase to generate a phylogenetic tree for 220 of the 231 genomes analyzed based on 49 highly-conserved clusters of orthologous group (COG) families and the FastTree maximum likelihood method (Price et al. 2010). Branch lengths were computed according to the Grafen method (Grafen 1989) using the *compute.brlen* command in the *ape* package (Paradis et al. 2004) in R Studio. To test for phylogenetic signals, Blomberg's K statistic (Blomberg et al. 2003) was calculated using the *multiPhylosignal* function in the *Picante* package (Kembel et al. 2010) in R Studio. Mean differences in potential CUE between phyla were compared using phylogenetic ANOVA with the *phylANOVA* function in the *phytools* package (Revell 2012) in R Studio.

In order to determine the taxonomic level which best describes variation in potential CUE, we used Phylocom to calculate the contribution index (CI) for each of the 191 nodes in the bacterial phylogeny (Webb et al. 2008). The CI indicates how much a particular node on the phylogeny accounts for the total variation in potential CUE (Moles et al. 2005). After calculating the CI for all 191 nodes in our analysis, we classified the subset of nodes where collective contributions accounted 90% of the variation in potential CUE based on the taxonomic level at which descendent species diverged.

The relationship between potential CUE and (1) the number of exchange reactions, (2) the number of C-containing exchange reactions, (3) genome size, (4) guanine-cytosine (GC) content, and (5) number of genes was assessed using phylogenetic generalized least-squares regression with the *nlme* package (Pinheiro et al. 2007) in R Studio. The proportion of variance explained by each predictor was estimated using a pseudo R^2 value designed for nonlinear regression (Burnham and Anderson 2002) using the *r.squaredLR* function in the *MuMIn* package (Bartón 2018), which estimates the improvement of the fit model relative to a null model based on a likelihood ratio test. The *stepAIC* function in the *MASS* package (Venables and Ripley 2002) was additionally used to determine the simplest regression model with multiple predictors.

Ecosystem modeling

The Dual-Arrhenius Michaelis-Menten Microbial C and Nitrogen Physiology (DAMM-MCNI_P) model was used to estimate potential impacts of the observed variation in CUE on ecosystem-level C fluxes. DAMM-MCNI_P models the effects of soil moisture and temperature on coupled C and nitrogen fluxes through soil pools and microbial biomass (Abramoff et al. 2017). Specifically, the model uses Michaelis-Menten kinetics to describe the depolymerization of soil organic C and soil organic Nitrogen by microbial extracellular enzymes to produce dissolved organic C (DOC) and dissolved organic N (DON). Maximum reaction velocities are governed by temperature-sensitive Arrhenius functions. Uptake of DOC and DON by microbial biomass is governed by a second series of Arrhenius and Michaelis-Menten kinetic equations, which are sensitive to moisture-

mediated O₂ availability. Following uptake, the parameter CUE is used to determine the partitioning of C between microbial biomass and soil respiration. DAMM-MCNI_P has been parameterized to describe seasonal patterns in heterotrophic soil respiration at a temperate forest site, and is able to capture 56% of variation in empirical observations of seasonal heterotrophic respiration at an hourly scale (RMSE = 0.25, R² adjusted = 56; Abramoff et al. 2017). The published model uses a default CUE value of 0.3 as used in several other ecosystem models. We modified the parameterization of CUE in this model (while retaining all other parameter settings as described in detail in Abramoff et al. 2017) to reflect the range of variation observed in CUE across taxa. We then quantified the impact of this variation on model estimates over 100 repeated cycles of annual variation in temperature and moisture to assess long-term impacts.

Results

Carbon use efficiency predicted from manually-curated metabolic models depends on both microbial identity and substrate type

We first calculated CUE using the set of 13 manually-curated, published metabolic models from diverse environments found in the BiGG database (King et al. 2016). Flux balance analysis was performed for each metabolic model with C supplied exclusively through one of 14 individual C-containing metabolites, and CUE was calculated as the proportion of C assimilated into biomass relative to C uptake. We observe a mean CUE of 0.53 ± 0.25 across taxa and substrate types, suggesting that nearly half of consumed C is lost via respiration on average (Table 2.1). However, these models also indicate wide

variation in mean CUE between individual taxa (0.14 ± 0.07 to 0.84 ± 0.17 , Supplementary Table 2.1) and equally large variation in mean CUE across substrate types (0.26 ± 0.24 to 0.66 ± 0.20 , Table 2.1).

Predicted potential carbon use efficiency from metabolic models for soil phyla varies widely across taxa

Potential CUE values represent intrinsic variation in CUE based on genomic differences between taxa, and these values were most useful for comparisons between taxa and for identifying relationships between genome traits and CUE. Potential CUE ranges from 0.22 to 0.98 across all taxa, with a mean of 0.62 ± 0.17 (Figure 2.2). The range of potential CUE values from this analysis corresponds to the high end of parameter settings currently used in microbial models of the C cycle (0.15 – 0.6; Manzoni et al. 2012). We also compared these predictions with a synthesis of published, taxa-specific measurements of CUE and found that empirical measurements of CUE extend well above our mean predicted potential CUE, with individual taxa showing CUE values as high as 0.82 (Figure 2.2).

We were able to generate corresponding metabolic models for a subset of these observations to compare potential CUE predictions with empirical observations. While substrate conditions differ across the empirical observations and are not identical to those used in the potential CUE modeling, we observe strong rank correlation in CUE across taxa (Pearson's $R=0.63$). In both our predicted potential CUE estimates and the empirical dataset, we see wide variability across mean CUE due to differences between taxa.

Predicted carbon use efficiency responds to the availability of particular metabolites

To assess the impact of substrate chemistry on CUE, we calculated the dependence of biomass production on all transport reactions associated with C uptake and secretion. We then identified the set of C-containing metabolites that most commonly limited biomass production across the full set of taxa in our analysis, and calculated CUE after reducing the availability of each of these constraining metabolites individually. The most common constraining reactions were related to amino acid and dipeptide uptake (Table 2.2, Figure 2.4). When uptake of individual constraining metabolites was set to reduce biomass production by 75%, mean CUE across all 18 constraining metabolites was 0.29 ± 0.19 . This corresponds to an average decline in CUE of 0.33, or a 53% reduction in CUE, compared to the potential CUE scenario.

Potential CUE is phylogenetically structured and correlated with genome traits

Potential CUE shows a significant phylogenetic signal ($K = 0.99$, $p < 0.01$, Figure 2.3), indicating a Brownian pattern of trait evolution, with closely-related taxa showing similarity in potential CUE values. The class ($CI = 0.02 \pm 0.019$, Table 2.3) and order ($CI = 0.016 \pm 0.020$, Table 2.3) levels explained the greatest level of variation in CUE. Therefore, these phylogenetic levels may be more appropriate than the phylum level for considering relationships between C cycling and bacterial community composition.

Consistent with our observations based on the BIGG models (Figure 2.5), we found a negative correlation between potential CUE and GC content in the larger set of metabolic models from kBase (Pseudo $R^2 = 0.20$, Table 2.4). Additionally, potential CUE is

significantly negatively correlated with genome size (Pseudo $R^2=0.36$, Table 2.4, Figure 2.6), the number of genes coded for within a genome (Pseudo $R^2=0.34$, Table 2.4) and the number of transport reactions associated with C uptake and secretion (Pseudo $R^2=0.50$, Table 2.4).

Phylogenetic variation in CUE has quantitatively-important implications for ecosystem-level estimates of C cycling

Under a scenario in which the microbial community exhibited high efficiency (CUE = 0.9), soil organic C pool sizes were nearly twice as large following 100 years of simulation compared to the low efficiency scenario (CUE = 0.2, Figure 2.7). This was driven, in part, by large sustained increases in microbial biomass, with the highly efficient microbial communities producing nearly four times greater microbial biomass than low efficiency communities over the same time span. Despite this large increase in microbial biomass, rates of respiration were reduced by 25% compared to the low efficiency communities (Figure 2.7).

Discussion

Genome-specific metabolic models have typically been used to explore variation in growth and microbial community interactions for small sets of microbial taxa (Henry et al. 2010, Harcombe et al. 2014). To date, this approach has not been applied to better understand microbial CUE, a key parameter in emerging microbial models of the C cycle. Here we show large, phylogenetically-structured variation in potential CUE attributed to differences in physiology among >200 individual bacterial taxa. We observed that CUE

was sensitive to both substrate chemistry and variation in microbial physiology between taxa. The intrinsic variation in CUE we observed among taxa is as large as that previously attributed to abiotic factors such as temperature and substrate chemistry (Manzoni et al. 2012, Sinsabaugh et al. 2016). For example, the temperature sensitivity of CUE for whole communities has been modeled as declining 0.4 units over a range of 25 C (Allison et al. 2010), while we observed over 0.6 units of variation in potential CUE between individual bacterial taxa. We detected a significant phylogenetic signal in potential CUE corresponding to clustering at sub-phylum levels, and we found that potential CUE was negatively correlated with particular genome traits, including genome size and GC content. Additionally, we identified a particular set of amino acids, dipeptides, fatty acids and carbohydrates that resulted in large reductions in CUE when their availability was constrained. Finally, we found that the range of variation we observed in CUE across taxa could have major implications for estimates of respiration and C storage at the ecosystem level.

Potential CUE

Overall, we observed a mean potential CUE of 0.62 ± 0.17 (Figure 2.2), which may represent a mean maximum CUE for bacteria in the absence of resource limitation. Potential CUE values represent intrinsic variation in CUE based on genomic differences between taxa, and these values were most useful for comparisons between taxa and for identifying relationships between genome traits and CUE. The range of observed potential CUE values from this analysis corresponds to the high end of parameter settings currently

used in microbial models of the C cycle (0.15 – 0.6, Manzoni et al. 2012). However, empirical measurements of CUE extend well above this mean in the absence of resource limitation (Keiblinger et al. 2010; Figure 2.2). When the availability of metabolites was constrained to reflect more reasonable expectations of resource limitation in the environment, we observed consistent declines of approximately 53% with CUE averaging 0.29 ± 0.19 (Table 2.2).

Potential CUE varied across bacterial lineages, although not at the phylum-level. The significant phylogenetic signal in potential CUE indicates a Brownian pattern of trait evolution, with closely-related taxa showing similarity in potential CUE values. However, we did not observe significant differences in potential CUE between bacterial phyla, and the greatest level of variation was structured at finer phylogenetic resolutions, including the class and order levels (Table 2.3). Similar conclusions warning against broad phylum-level generalizations regarding carbon use traits have emerged from recent work using stable-isotope approaches (Morrissey et al. 2016).

Certain genomic traits, such as GC content and genome size, can be useful predictors of bacterial niche preferences and the response of bacterial communities to environmental changes (Lauro et al. 2009, Barberán et al. 2014, Roller et al. 2016). Bacteria with larger genomes must allocate greater resources towards maintenance, while smaller genomes can exhibit greater efficiency (Lynch 2006). In soil, taxa with larger genomes tend to dominate communities where substrates are available in diverse forms but limiting concentrations (Konstantinidis and Tiedje 2004), and bacterial taxa with larger genomes tend to occupy a wider range of habitat types (Barberán et al. 2014). In our analysis, potential CUE declined

by 0.04 units per additional Mbp in a genome (Figure 2.6). However, there was a tradeoff between efficiency and access to substrates, as taxa with larger genomes were able to access a larger breadth of C sources at the cost of reductions in potential CUE. Thus, taxa with the highest CUEs may be less adaptable to changes in substrate chemistry, representing a more specialized trophic strategy. Prior studies also observe copiotrophic taxa having large numbers of genes associated with transport proteins, which would correspond to large numbers of transport reactions associated with C uptake and secretion (Lauro et al. 2009).

Nutrient limitation can lead to shifts in community composition that favor GC-poor genomes, potentially due to the greater energetic cost of producing GTP and CTP bases (Mann and Chen 2010). Consistent with these findings, we observed a strong negative correlation between CUE and GC content ($R^2=0.522$; Figure 2.5). Thus, environmental changes that favor GC-poor genomes may also have ramifications for C cycling through correlated increases in CUE and corresponding reductions in CO₂ emissions.

Substrate limitation

In the environment, microbial taxa are exposed to variation in substrate chemistry and availability, which can impact rates of C uptake and growth (Mooshammer et al. 2014). In prior studies, CUE shows sensitivity to substrate availability and stoichiometry at both the organismal (Keiblinger et al. 2010, Roller and Schmidt 2015) and community levels (Manzoni et al. 2012, Sinsabaugh et al. 2013). In our analysis, we identified several specific amino acids and dipeptides whose availabilities limited CUE. These findings comport with

patterns of amino acid uptake by bacteria in the environment (e.g. Farrell et al. 2013, Broughton et al. 2015) and the incorporation of amino acids, such as alanine, directly into cell wall components (Lam et al. 2009). Amino acids represent the largest input of N in soil (Jones et al. 2009), and rapid uptake by microbes results in short residence time of these compounds in soil (Finzi and Berthrong 2005, Wilkinson et al. 2014). It is hypothesized that microbes consume amino acids primarily as a C source, which may support the large impact of constraining amino acid availability on CUE we observed. Similarly, dipeptides contain higher C:N ratios than their component amino acid monomers, and their uptake is greater than that of amino acids (Farrell et al. 2013).

Structured variability in soil organic matter chemistry in soil could favor particular bacterial taxa over others based on their capacity to consume available C sources. In our analysis, taxa with the ability to consume a wide range of metabolites showed the lowest potential CUE values because of increased uptake of non-essential C-containing metabolites (Table 2.4). In contrast, taxa with fewer exchange reactions were able to maintain higher CUE in the potential environment through reduced C uptake. These differences may be related to differences between copiotrophs and oligotrophs in terms of resource specialization, with less-specialized copiotrophic taxa showing lower CUE. Prior studies also observe copiotrophs having large numbers of genes associated with transport proteins, which would correspond to large numbers of C-containing exchange reactions in this analysis (Lauro et al. 2009).

Ecosystem scaling

Accounting for the variation in CUE we observed across taxa can have significant consequences to ecosystem-level estimates of C pool sizes and respiration rates (Figure 2.7). Soil organic C pool sizes were reduced by almost half when a community shift towards low efficiency bacteria (CUE=0.2, Figure 2.7) was modeled compared to a community comprised of high efficiency bacteria (CUE=0.9). These values represent the extremes of our potential CUE observations and therefore represent the widest range of expected outcomes. The change in soil organic C pool sizes we observed was driven, in part, by large sustained increases in microbial biomass, with the highly efficient microbial communities producing nearly four times greater microbial biomass than low efficiency communities over the same time span. Despite this large increase in microbial biomass, rates of respiration were reduced by 25% compared to the low efficiency communities (Figure 2.7). Thus, accounting for variation in CUE among taxa alone can have significant consequences to ecosystem-level estimates of C storage and respiration rates, and these differences can persist even at decadal timescales. While DAMM-MCNP was parameterized and validated for a specific temperate forest ecosystem, the general model structure and its dependence on CUE are representative of soil C models used across several ecosystem types (Manzoni and Porporato 2009), suggesting that variation in CUE across taxa is likely to have important implications for soil C cycling more broadly.

Limitations of genome-scale metabolic modeling

Direct comparisons between the values observed here and those in other studies are challenging as potential CUE may not have exact parallels to empirical observations in which CUE has been measured for a small number of individual taxa or complex soil communities. We interpret potential CUE values as intrinsic variation based on genetic differences between taxa that may be most useful in terms of exploring comparisons between taxa and for identifying relationships between genome traits and CUE. Despite the challenges with measuring CUE, developing empirical approaches (e.g. Bolscher et al. 2016, Morrissey et al. 2017) to directly estimate taxa-specific CUE will be necessary and useful for validating these observations.

It is important to note that the limited capacity of ascribing functions to genes through annotation pipelines, the poor knowledge of taxon-specific microbial biomass composition, and the need to implement a gap-filling algorithm to compensate for missing reactions in genome scale reconstructions can each impact flux estimates generated through FBA models. Despite these limitations, models generated through the pipeline used for our draft predictions have been demonstrated to closely match empirical phenotype data (Henry et al. 2010), have been used to explore complex community interactions (Magnúsdóttir et al. 2017), and have been shown to successfully predict community structure and environmental metabolomics (Garza et al. 2018). We expect that future advancements in genome annotation and metabolic model construction paired with taxa-specific empirical observations of CUE could use the same conceptual framework proposed here to provide predictions with improved precision and fidelity. We interpret

potential CUE values as useful theoretical predictions for exploring comparisons between taxa and for identifying relationships between genome traits and CUE. Collecting empirical measurements of taxa-specific bacterial CUE across a range of substrate types using a consistent methodology is recommended for further validation of these findings.

Conclusions

The range of potential CUE values we observed between taxa is comparable to that observed in other studies in which wide ranges of CUE are attributed to differences between ecosystems or due to abiotic factors (Sinsabaugh et al. 2016). Soil microbial communities undergo shifts in composition under global change (DeAngelis et al. 2015), and these changes may alter the overall soil microbial CUE if particular taxa with uniquely high or low CUE values are favored based on growth strategy or substrate preference. Failing to account for relationships between CUE and microbial community composition may cause ecosystem models to miss important biotic feedbacks that can impact respiration fluxes and soil-C balance (Li et al. 2014). This analysis introduces a novel method for generating draft predictions of taxa-specific CUE from metabolic models and identifies genome size and GC content as traits that may link genomic variation with C utilization strategies. We show that large phylogenetic variation in CUE between individual taxa make microbial physiology and community composition important factors to consider when estimating microbial contributions to C cycling.

Tables and Figures

Table 2.1 — CUE for BiGG models across substrate types.

Genus	Species	Genome Size (Mbp)	GC Content	D-Glucose	Fumarate	Acetate	Acetaldehyde	2-Oxoglutarate	Ethanol	Formate	D-Fructose	L-Glutamine	L-Glutamate	D-lactate	L-Malate	Pyruvate	Succinate
Bacillus	<i>subtilis subsp. subtilis str. 168</i>	4.22	43.50	0.94	0.70	NG	NA	0.64	NG	NG	0.94	0.64	0.64	0.94	0.70	0.78	0.70
Clostridium	<i>ljungdahlii DSM 13528</i>	4.63	31.10	0.95	NA	NA	NA	NA	NA	0.48	0.94	0.88	0.87	0.97	NA	0.80	NA
Escherichia	<i>coli str. K-12 substr. MG1655</i>	4.64	50.79	0.62	0.39	0.37	0.60	0.45	0.70	NG	0.62	0.48	0.51	0.50	0.39	0.41	0.42
Geobacter	<i>metallireducens GS-15</i>	4.01	59.49	NA	NA	0.11	NA	NA	0.28	0.06	NA	NA	NA	NA	NA	0.23	NA
Klebsiella	<i>pneumoniae subsp. pneumoniae e MGH 78578</i>	5.69	57.15	0.63	NG	0.46	0.68	NG	0.80	NG	0.63	0.53	0.53	0.50	0.43	0.45	NG
Methanosarcina	<i>barkeri str. Fusaro</i>	4.87	39.23	NA	NA	0.55	NA	NA	NA	NA	NA	NA	NG	NA	NA	0.56	NA
Mycobacterium	<i>tuberculosis H37Rv</i>	4.41	65.61	0.43	0.25	NA	NA	NA	NA	NA	NA	NA	0.28	0.31	0.25	0.27	0.29
Saccharomyces	<i>cerevisiae S288c</i>	12.2	38.38	0.70	0.49	0.90	0.98	0.82	0.98	0.54	0.70	0.95	0.95	0.93	0.49	0.53	0.85
Salmonella	<i>enterica subsp. enterica serovar Typhimurium str. LT2</i>	4.81	52.23	0.64	0.44	0.44	0.66	0.48	0.78	0.02	0.64	0.55	0.54	0.50	0.44	0.45	0.48
Shigella	<i>boydii Sb227</i>	4.65	51.11	0.27	0.08	NG	0.14	0.15	0.14	NG	0.27	0.15	0.15	0.09	0.08	0.09	0.08
Staphylococcus	<i>aureus subsp. aureus N315</i>	2.84	32.81	0.60	NA	NA	NA	0.77	0.74	0.22	0.60	NA	0.56	0.54	0.94	NA	NA
Synechocystis	<i>sp. PCC 6803</i>	3.95	47.36	0.50	0.35	0.31	NA	0.36	NA	NA	0.50	0.81	0.62	NA	0.35	0.37	0.39
Thermotoga	<i>maritima MSB8</i>	1.86	46.25	0.72	NA	NG	NA	NA	NA	NA	0.72	NA	NA	0.70	NA	NA	NA

CUE from manually-curated metabolic models from BiGG database.

NG = no growth, NA = exchange reaction absent from model.

Table 2.2 — Constrained CUE summary

Constraining metabolite	Class	Mean $\pm$ SD Constrained CUE	Biomass/ Uptake Slope	Cohen's D	Paired T-test P-value
L-Lysine	Amino Acid	0.27 ± 0.17	3.25 ± 0.36	2.9	***
Myristic Acid	Fatty Acid	0.28 ± 0.17	6.48 ± 0.34	2.9	***
Gly-Phe	Dipeptide	0.27 ± 0.15	15.93 ± 15.53	2.8	***
Trehalose (trhl)	Carbohydrate	0.30 ± 0.19	4.81 ± 4.4	2.7	***
Gly-Tyr	Dipeptide	0.28 ± 0.15	8.52 ± 1.48	2.7	***
Glyceraldehyde 3 Phosphate (g3p)	Carbohydrate	0.35 ± 0.06	14.88 ± 12.33	2.7	**
Gly-Asn	Dipeptide	0.28 ± 0.18	5.25 ± 1.09	2.5	***
Stearic Acid (ocdca)	Fatty Acid	0.25 ± 0.20	8.38 ± 4.13	2.4	***
Lauric Acid (ddca)	Fatty Acid	0.28 ± 0.16	8.68 ± 4.66	2.4	***
Ala-His	Dipeptide	0.29 ± 0.20	12.79 ± 1.55	2.3	***
L-Valine	Amino Acid	0.30 ± 0.21	2.92 ± 0.28	2.2	***
L-Tyrosine	Amino Acid	0.37 ± 0.24	8.42 ± 0.28	1.9	***
L-Arginine	Amino Acid	0.32 ± 0.22	3.99 ± 0.76	1.9	***
L-Phenylalanine	Amino Acid	0.38 ± 0.24	6.43 ± 0.21	1.8	***
L-Isoleucine	Amino Acid	0.36 ± 0.23	3.21 ± 0.55	1.8	***
D-Arabinose	Carbohydrate	0.40 ± 0.22	13.33 ± 0	1.7	***
L-Proline	Amino Acid	0.41 ± 0.30	5.2 ± 1.63	1.6	**
L-Histidine	Amino Acid	0.55 ± 0.25	15.52 ± 6.49	1.3	*

Effect of constraining the availability of particular metabolites on CUE. Maximum uptake of constraining metabolite was set to reduce biomass to 25% of maximum flux, based on observed linear relationships between uptake and biomass production. Biomass/uptake slope indicates mean $\pm$ standard deviation of biomass flux per unit uptake flux for all models with a particular constraining metabolite. Cohen's D value compares potential CUE to constrained CUE for all models that have an exchange reaction for the given metabolite. P-values are from paired t-tests comparing constrained and potential CUE for all models that have an exchange reaction for the given metabolite (***P<0.001, ** P<0.01, * P<0.05).

Table 2.3 — Contribution indices for variation in potential CUE by taxonomic level.

	Average CI	SD	Total CI
Class	0.020	0.019	0.299
Family	0.012	0.007	0.106
Genus	0.009	0.006	0.132
Order	0.016	0.020	0.286
Phylum	0.016	0.009	0.047
Species	0.010	NA	0.010
Strain	0.011	0.005	0.021

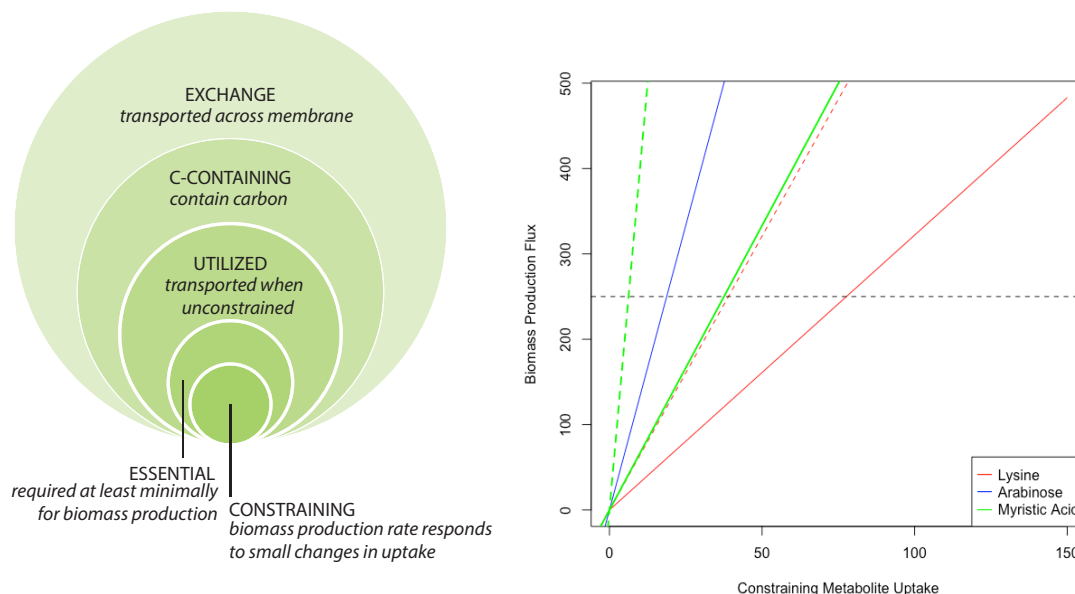
Contribution Index (CI) values for nodes accounting for 90% of variation in potential CUE.

Table 2.4 — Regression summary for predictors of potential CUE.

	GLS intercept	GLS Slope	GLS P-Val	Pseudo R²
Exchange reactions	1.187	-5.91×10^{-3}	<0.001	0.500
C-containing exchange reactions	1.066	-6.37×10^{-3}	<0.001	0.496
GC content	0.909	-4.86×10^{-3}	<0.01	0.201
Genes	0.822	-4.06×10^{-5}	<0.001	0.341
Genome size	0.819	-3.61×10^{-8}	<0.001	0.356

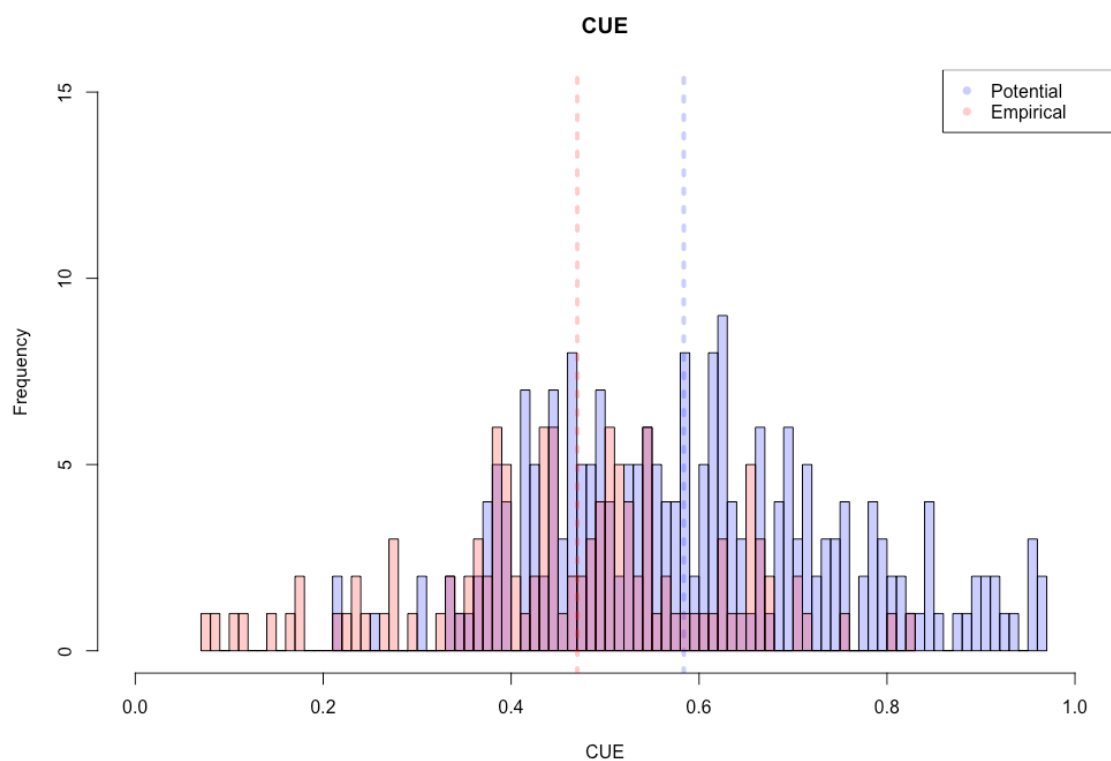
PGLS regression results for potential CUE regressed against individual predictors listed in rows. For multi-regression models, lowest AIC models included genome size, number of C-containing exchange reactions, and number of genes.

Figure 2.1 — Metabolic model reaction classification framework



(A) Framework for classifying reactions to identify metabolites impacting CUE. (B) Uptake fluxes for all common constraining metabolites were analyzed. Fluxes for three metabolites and selected taxa are shown as an example, with lines colored by metabolite. Solid red line shows biomass response to L-lysine uptake seen for 186 models, including *Terriglobus saanensis* SP1PR4 and *Starkeya novella* DSM 506. Dashed red line shows biomass response to L-lysine uptake for *Verrucomicrobia bacterium* SCGC AAA164-I21. Blue line shows biomass response to D-Arabinose uptake seen for 23 models, including *Verrucomicrobia bacterium* SCGC AAA164-I21. Solid green line shows biomass response to myristic acid uptake seen for 106 models including *Terriglobus saanensis* SP1PR4. Dashed green line shows biomass response to myristic acid uptake seen for 38 models including *Starkeya novella* DSM 506. Uptake values corresponding to 25% maximum biomass production (dashed black line) were calculated for each of the most common constraining metabolites for each taxa. This uptake value was then used as the maximum uptake flux for constrained CUE calculations.

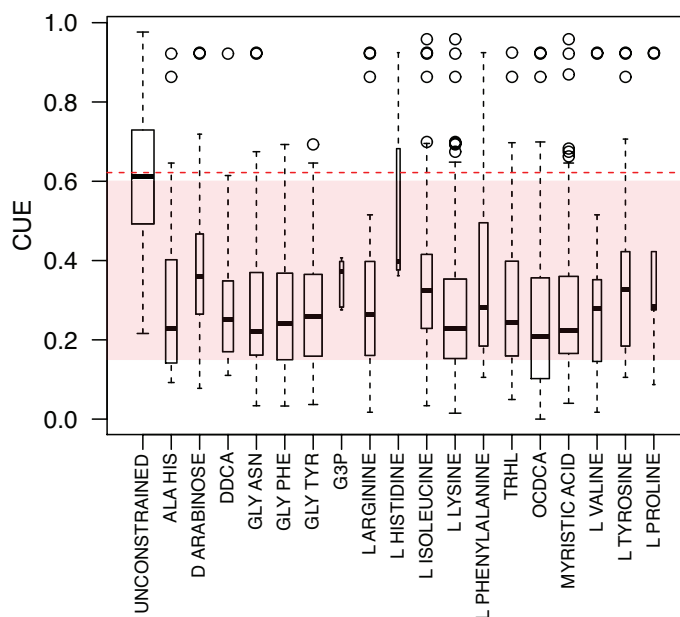
Figure 2.2 — Histogram of potential cue estimates and empirical observations across taxa



Histogram of potential CUE (purple) from kBase models across taxa and empirical measurements of CUE (pink) across taxa and substrate type.

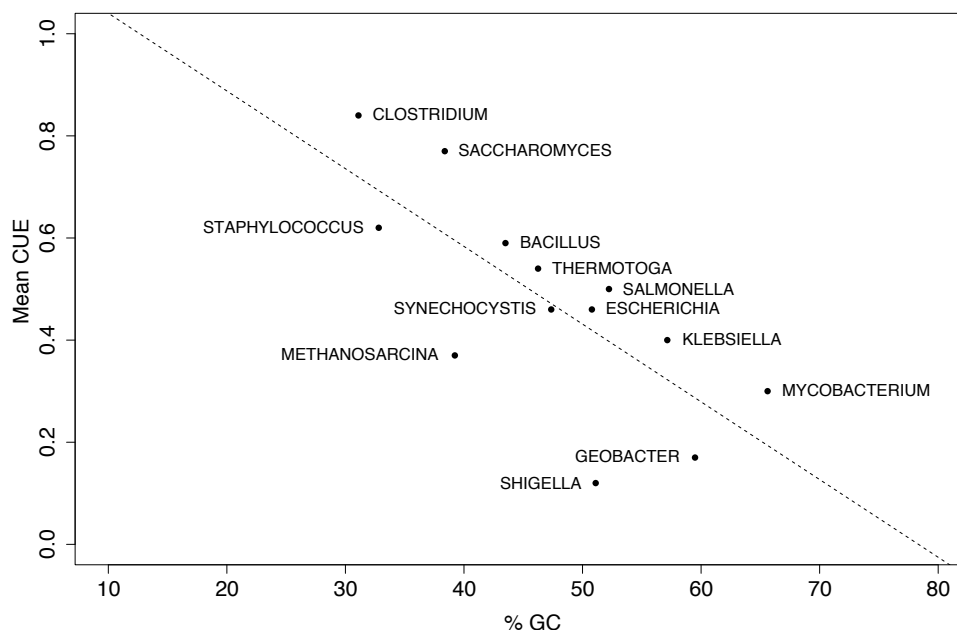
Phylogenetic heatmap of potential CUE values for kBase models. External rings alternate to delineate phyla or Proteobacterial classes. Labels on tips correspond to kBase accession numbers.

Figure 2.4 — Boxplots of constrained cue across substrate types



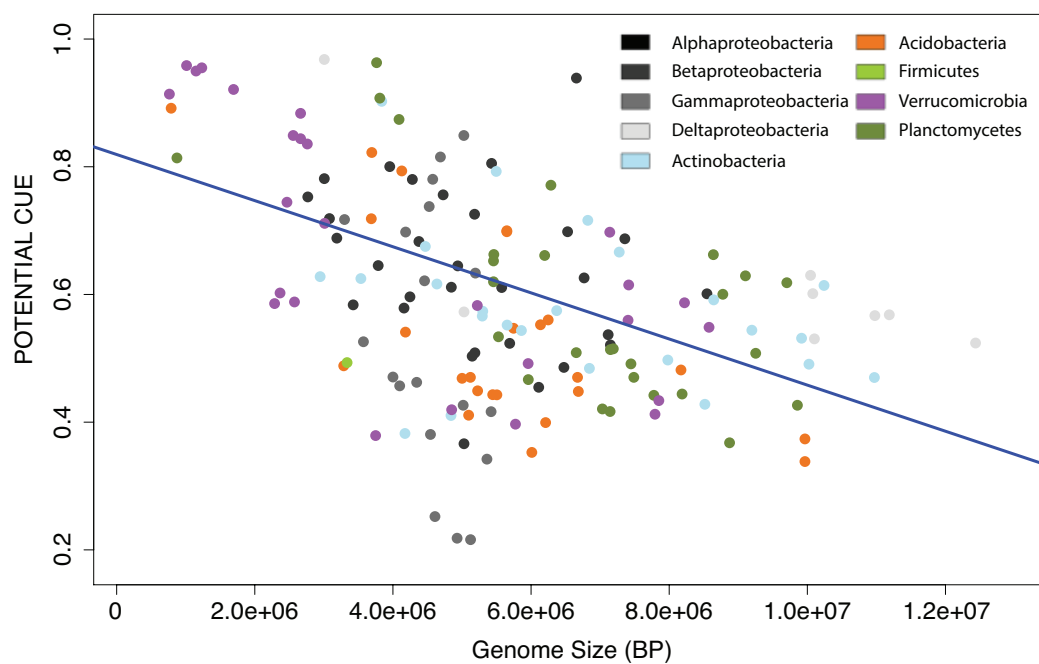
Boxplot of average CUE values across all taxa under potential and constrained scenarios. Boxplot width is proportional to number of models with a given constraining reaction. Dashed red line shows average for potential CUE. Shaded region shows range of values typically used in biogeochemical models. Solid lines within boxplots show median. Bottom and top edges of boxes represent 25th and 75th percentiles, respectively. Whiskers demarcate minimum and maximum datapoints within 1.5x of the interquartile range.

Figure 2.5 — Regression of mean cue against GC content



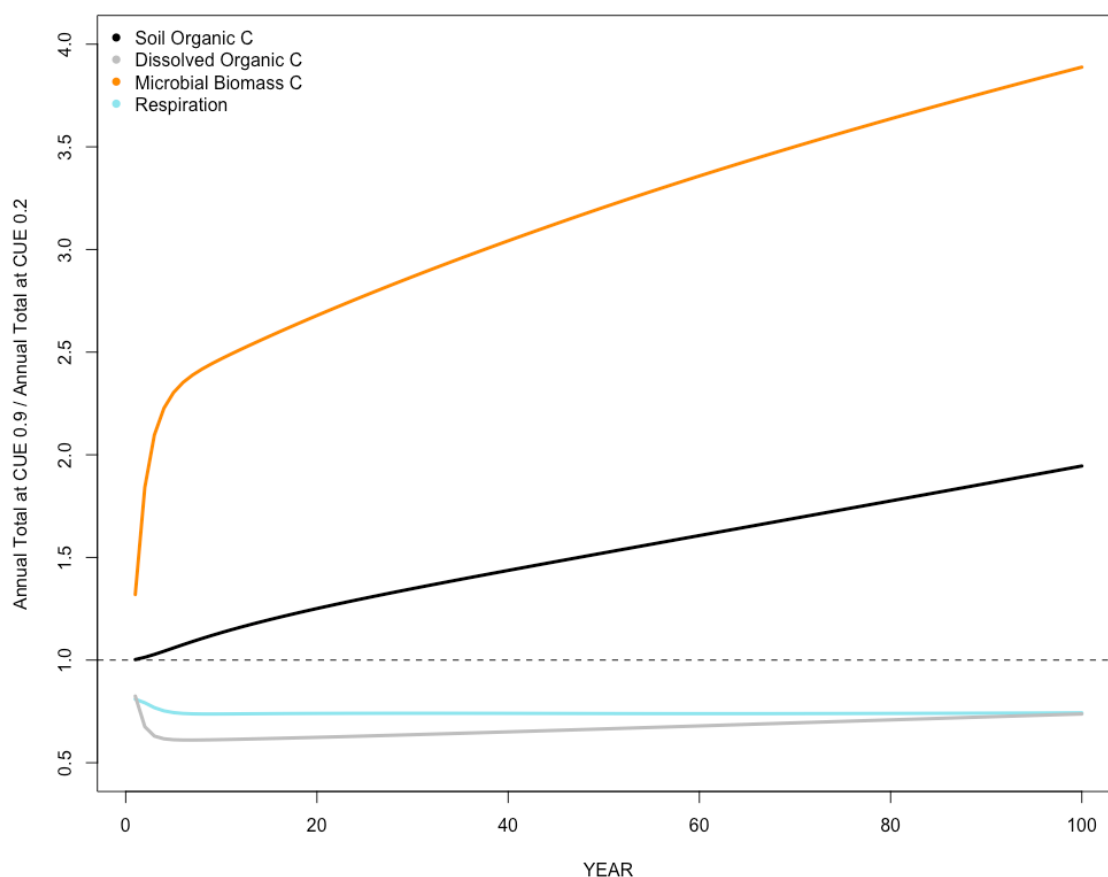
Mean CUE versus GC content for manually-curated metabolic models. Species in order of increasing GC content are *Clostridium ljungdahlii* DSM 13528, *Staphylococcus aureus* subsp. *aureus* N315, *Saccharomyces cerevisiae* S288c, *Methanosarcina barkeri* str. *Fusaro*, *Bacillus subtilis* subsp. *subtilis* str. 168, *Thermotoga maritima* MSB8, *Synechocystis* sp. PCC 6803, *Escherichia coli* str. K-12 substr. MG1655, *Shigella boydii* Sb227, *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* str. LT2, *Klebsiella pneumoniae* subsp. *pneumoniae* MGH 78578, *Geobacter metallireducens* GS-15, *Mycobacterium tuberculosis* H37Rv. Mean CUE was calculated from CUE on growth on each of the following C-sources individually: D-Glucose, Fumarate, Acetate, Acetaldehyde, 2-Oxoglutarate, Ethanol, Formate, D-Fructose, L-Glutamine, L-Glutamate, D-lactate, L-Malate, Pyruvate, Succinate.

Figure 2.6 — Regression of potential CUE against genome size



Potential CUE regressed against genome size (bp). Blue lines show GLS fit. Points are colored by phylum.

Figure 2.7 — Modeled C cycling variation across cue range



Annual totals for C cycle pools and respiration rates for models for high efficiency taxa (CUE=0.9) relative to low efficiency taxa (CUE=0.2) across 100 years. Dashed line represents no difference in model estimates at the two CUE values.

**CHAPTER THREE: ECTOMYCORRHIZAL FUNGI SUPPRESS THE
ACTIVITY BUT NOT THE ABUNDANCE OF SOIL BACTERIAL
DIAZOTROPHS, NITRIFIERS AND DENITRIFIERS IN TEMPERATE
FORESTS**

Abstract

Microbial processes represent some of the greatest sources of uncertainty in N cycling budgets in temperate forests, where the availability of N can be a primary constraint to productivity. While microbial processes account for major inputs, transformations and exports of N in soil, relationships between microbial community structure and N cycle fluxes are not well understood. The availability of N in soil is dependent on both abiotic and biotic factors, such as competition for N between plant roots, mycorrhizal fungi, and free-living bacteria involved in N-fixation, nitrification and denitrification. In particular, competition between free-living soil bacteria and ectomycorrhizal (ECM) fungi has been shown to have important ramifications for biogeochemical cycling in temperate forests, but it is not known how these interactions impact the structure and activity of N cycling bacterial groups. We explored how rates of these N cycle fluxes vary across gradients of mycorrhizal abundance at four temperate forest sites in Massachusetts and Indiana, USA. We paired these measurements of N-fixation, net nitrification and denitrification rates with gene abundance data for specific bacterial functional groups associated with each process. We find that the availability of NO_3^- and rates of N-fixation, net nitrification and denitrification all decline with increasing abundances of plants associated with ECM fungi.

Despite this strong structuring of activity in relation to mycorrhizal abundance, we did not find consistent patterns in the abundances of functional groups across the gradients, and gene abundances were not correlated with process rates. Overall, we find evidence for suppressed rates of N-fixation, net nitrification and denitrification in the presence of ECM fungi. Future work will benefit from continuing to couple measurements of N cycling microbial abundances and associated fluxes.

Introduction

Nitrogen (N) availability can be a primary constraint on productivity in temperate forests (Vitousek and Howarth 1991, LeBauer and Treseder 2008, Xia and Wan 2008), yet quantifying N cycling rates in these ecosystems remains a significant challenge. Long-term datasets are regularly unable to identify processes responsible for sizable accumulations and losses of N in temperate forests, and much of this is attributed to uncertainties in soil microbial transformations of N, including N-fixation, nitrification and denitrification (Bormann 1993, Bormann et al. 2002, Yanai et al. 2012, 2013). Each of these N cycle fluxes is carried out by distinct microbial functional groups, and characterizing controls on the abundances and activities of these microbial groups is critical to an improved understanding N cycling (Wallenstein and Vitousek 2005, Hsu and Buckley 2009, Berthrong et al. 2014, Enanga et al. 2017, Lennon and Houlton 2017). In soil, heterotrophic bacteria are subject to complex interactions with plant roots, mycorrhizal fungi, and other soil organisms, but few studies have explored how these interactions impact the activity and abundance of N cycling functional groups.

In temperate forests, nearly all trees associate with either arbuscular mycorrhizal (AM) or ectomycorrhizal (ECM) fungi, and differences between these two groups of mycorrhizal fungi have key implications for biogeochemical cycling (Phillips et al. 2013, Averill et al. 2014, Craig et al. 2018). ECM fungi are able to produce a wide range of enzymes including those targeting plant substrates such as cutin, lipids, waxes, pectin, cellulose, cellobiose, hemicellulose, polyphenols and lignin (Read and Perez-Moreno 2003). In contrast, AM fungi have a much narrower profile of enzyme capacities, typically

lack saprotrophic capabilities, and are primarily associated with the uptake of inorganic N and phosphorous (Read and Perez-Moreno 2003).

Due to their broad enzyme capacities, ECM fungi are able to compete directly with free-living saprotrophs for access to organic substrates (Gadgil and Gadgil 1975, Cornelissen et al. 2001, Phillips et al. 2013, Sterkenburg et al. 2018). Competition for N between free-living soil microbes and ectomycorrhizal fungi results in suppressed rates of heterotrophic respiration (Averill and Hawkes 2016), and this interaction leads to greater soil C storage in surface soils of ECM-associated forest stands at a global scale (Averill et al. 2014, 2018, Craig et al. 2018). The suppression of free-living bacterial activity in ECM-associated forest stands also results in slowed rates of N mineralization from organic forms, with higher ratios of organic N to inorganic N relative to AM-associated stands (Finzi and Berthrong 2005, Phillips et al. 2013). On the other hand, forest stands dominated by trees associated with AM fungi are characterized by rapid rates of decomposition and high rates of N mineralization, allowing AM-associated trees to acquire N primarily in inorganic forms (Lovett et al. 2004, Gallet-Budynek et al. 2009, Midgley and Phillips 2014).

Competition between ECM fungi and free-living microbial groups is thought to be the primary mechanism explaining mycorrhizal-associated differences in soil C storage (Orwin et al. 2011, Averill et al. 2014). However, relatively little work has been done to characterize and compare the functional profiles of free-living bacterial groups in relation to mycorrhizal fungal type. In particular, it is not well understood how biotic interactions between bacteria, fungi and plant roots affect the abundances and activities of N cycling bacterial functional groups. The genetic basis for N-fixation, nitrification and

denitrification has been well-characterized (Levy-Booth et al. 2014), providing a unique system in which functional group abundances can be linked with biogeochemical flux measurements (Wallenstein and Vitgalys 2005, Hsu and Buckley 2009, Reed et al. 2010).

Microbial community structure can be described using taxonomic and phylogenetic approaches to characterize the composition of the community, or alternatively, by categorizing microbial groups according to function (Rocca et al. 2015). Measures of gene abundances are often used to describe the functional potential of microbial communities and broadly characterize their structure as related to biogeochemical cycling (Wallenstein and Vitgalys 2005, Hsu and Buckley 2009, Berthrong et al. 2014, Levy-Booth et al. 2014, Bier et al. 2015, Rocca et al. 2015). Identifying relationships between microbial community structure and biogeochemical function has been a major goal in microbial biogeochemistry, especially in the context of understanding how shifts in community structure might impact ecosystem processes (Allison and Martiny 2008, McGuire and Treseder 2010). However, characterizing these relationships has proven difficult due to the high diversity of soil microbial communities, potential functional redundancy across taxa, dormancy, and abiotic controls on activity (Bier et al. 2015). Relatively few studies in temperate forest ecosystems pair targeted measurements of gene abundances with their associated biogeochemical flux or pools measurements (Rocca et al. 2015), despite the presence of clearly-delineated bacterial functional groups in the N cycle.

In this study, we paired measurements of microbial N cycling with abundance measurements for specific bacterial functional groups along gradients of mycorrhizal type to explore two hypotheses (Figure 3.1). First, we hypothesize that competitive interactions

between ECM-fungi and free-living N cycling bacterial groups will promote N-fixation as an alternate pathway for access to N, only accessible by diazotrophs. In contrast, we hypothesize that these competitive interactions lead to reduced rates of nitrification and denitrification as the proportion of ECM-associated tree species increases at the plot level. Second, we hypothesize that differences in flux rates are underwritten by differences in community structure, with a greater abundance of diazotrophs and fewer nitrifiers, and denitrifiers in stands dominated by ECM-associated tree species.

Methods

Site Descriptions

This research was conducted at 48 plots (15m x 15m) along four mycorrhizal gradients established in temperate hardwood forests. Two gradients, each consisting of nine plots (15m x 15m), were located in southern Indiana, USA at the Griffy Woods (GW, 39°11'N, 86°30'W) and Lilly-Dickey Woods (LD, 39°14'N, 86°13'W). These plots are described in further detail in Cheeke et al. 2016. Soils in these forests are unglaciated, silty-loams derived from sandstone, shale, and limestone (United States Department of Agriculture Soil Survey). Mean annual precipitation is 120 cm and mean annual temperature is 11.6°C (Phillips et al. 2013, Cheeke et al. 2017). Two gradients, each consisting of 15 plots (15m x 15m), were located at the Prospect Hill (PH) and Simes (SM) tracts of the Harvard Forest Long Term Ecological Research Site (42.58 N, 72.188 W) in Petersham, Massachusetts, USA. Soils at these sites are inceptisols classified as Typic Dystrochrepts derived from glacial till overlying granite-schist-gneiss bedrock (USDA

Natural Resources Conservation Service, Web Soil Survey). Mean annual precipitation is 110 cm and mean annual temperature is 8°C.

At each plot, all trees over 5 cm diameter at breast height (DBH) were measured, identified to species-level, and classified by mycorrhizal type to determine the proportion of tree basal area associated with AM and ECM fungi (Phillips et al. 2013). The proportion of AM-associated trees at the plot-level ranges from all AM-associated to all ECM-associated. At plots located in Harvard Forest, dominant tree species include *Quercus Rubra* (ECM), *Acer Rubrum* (AM), *Acer Saccharum* (AM), *Fraxinus Americana* (AM), and *Betula Lenta* (ECM). At plots located in southern Indiana, dominant tree species include *Acer Saccharum* (AM), *Quercus Rubra* (ECM), *Quercus alba* (ECM), *Liriodendron tulipifera* (AM), and *Fraxinus Americana* (AM).

Soil Sampling and Soil Properties

Four 0-15 cm mineral horizon soil cores were collected from each plot and transported on ice to Boston University, MA, USA. Replicate cores were composited at the plot level. Due to variation in the depth of the organic horizon across the gradients, we chose to focus our analysis on the mineral horizon across all plots to allow for comparisons across functionally comparable soil layers. Gravimetric soil moisture was measured based on mass loss following evaporation. Soil pH was measured with a pH probe placed in soil pastes prepared at a 1:10 ratio of soil to deionized water. Total soil C and N content were measured via flash combustion on a Thermoquest NC 2500 elemental autoanalyzer.

Inorganic N pools and nitrification rates

Total inorganic N extractions were completed using 20 g field-moist soil and 100 mL 2M KCl. Concentrations of NO_3 and NH_4 in extractant were quantified using colorimetric microplate assays followed by spectrophotometry with a Versamax microplate reader (Molecular Devices; Sims et al. 1995, Doane and Horváth 2003). Rates of potential net nitrification were measured by calculating the change in concentrations of NO_3 over 28 days of incubation at room temperature.

Potential N-fixation

Three 5 g subsamples of bulk soil from each plot (n=30) were incubated in gas-tight vials with a mixture of 80% N_2 and 20% O_2 gas for nine days following the date of sampling. Controls received natural abundance N_2 , while treatments received 98 atom% enriched $^{15}\text{N}_2$ gas under anoxic conditions with glucose amendments to measure potential rates in the absence of C-limitation. Following the incubation period, samples were dried, ground and underwent isotopic analyses using mass spectrometry. Net potential N-fixation was calculated based on the difference in total ^{15}N in treatments compared to control samples.

Denitrification

20 g of soil were placed in a mason jar and incubated with 2mL of solution containing 16 μg of 98 atom% enriched $\text{K}^{15}\text{NO}_3\text{-N}$. Gas samples of 10mL were collected through rubber septa attached to the mason jars immediately following labeling and after 24 hours. Gas samples were transferred to evacuated exetainers and concentrations of $^{15}\text{N}_2$

and $^{15}\text{N}_2\text{O}$ gas were analyzed at the UC Davis Stable Isotope Facility using a ThermoScientific GasBench and Precon gas concentration system paired with a ThermoScientific Delta V Plus isotope-ratio mass spectrometer (Bremen, Germany).

Molecular analyses

For all samples from the Harvard Forest, rhizosphere soil was manually separated from bulk soil on the date of sampling by agitating plant roots and isolating soil remaining adhered to roots. This fractionation was not performed for samples from Indiana due to a longer transportation time on ice. DNA was extracted in triplicate for 0.25g subsamples of soils using MoBio PowerSoil DNA kits. Following extraction and quality-checking using a NanoDrop spectrophotometer, replicate DNA extractions were pooled. Quantitative PCR (qPCR) was performed using primer sets and annealing temperatures published in Levy-Booth et al. 2014 to quantify the abundances of *nifH*, *amoA* and *nosZ*. The gene *nifH* codes for the enzyme nitrogenase, which catalyzes N-fixation. The gene *amoA* codes for the alpha-subunit of the ammonium oxidase enzyme, which is involved in nitrification. The gene *nosZ* codes for the catalytic subunit of nitrous oxide reductase, which is involved in denitrification. Standard curves were calculated based on serial dilution of gBlocks containing amplicon regions for each respective gene synthesized by integrated DNA technologies (Coralville, IA). All reactions were performed in triplicate in 384-well microplates using ABsolute qPCR Master Mix, containing SYBR Green and ROX, on an ABI 7900ht qPCR machine.

Statistical analyses

All statistical analyses were conducted using Rstudio v. 1.1.383 (R Core Team 2017). To explore how N cycle rates and gene abundances vary in relation to mycorrhizal abundance, generalized linear models (GLMs) with gamma-distributions and inverse link functions were analyzed using the proportion of ECM-associated basal area at the plot level as predictor variables and rates of N-fixation, nitrification, denitrification, or gene abundances of *nifH*, *amoA*, or *nosZ* as response variables. To explore relationships between gene abundances and N pool sizes, we analyzed GLMs with gene abundances as response variables and inorganic N pool concentrations associated with each respective gene as predictors. Similarly, we used a series of GLMs to explore gene-flux relationships with *nifH* regressed against N-fixation rates, *amoA* regressed against nitrification rates, and *nosZ* regressed against denitrification rates. With 48 samples, the lowest correlation detectable with a power of 80% is 0.39.

We used stepwise AIC model selection to explore more complex models of N-fixation, nitrification and denitrification rates. The *stepAIC* function in the *MASS* package (Venables and Ripley 2002) was used to compare models with proportion ECM basal area, total basal area, *Acer saccharum* basal area, soil pH, and inorganic N concentrations as potential predictors of N-fixation and net nitrification rates. Similarly, we used a model with proportion ECM basal area, total basal area, *Acer Saccharum* basal area and soil pH as potential predictors of denitrification rates. We chose to include *Acer Saccharum* basal area in these models due to interest in the potential for this tree species to drive N cycling patterns (Finzi et al. 1998, Templer et al. 2005, Phillips et al. 2013).

We also classified plots as AM-dominated (>70% AM), ECM-dominated (>70% ECM), or mixed (30-70% ECM), and used these classifications as categorical variables in analysis of variance (ANOVA) tests. We used the Tukey test of honest significant differences to compare flux rates, N pools and gene abundances across these classifications. Gene abundances and rates of net nitrification were log-transformed in these analyses. Rates of denitrification were averaged across 2016 and 2017.

Results

N availability

The concentration of extractable NO_3 ranged from 0 to $6.15 \mu\text{g NO}_3\text{-N/g soil}$ across all plots, with a mean of $0.96 \pm 0.199 \mu\text{g NO}_3\text{-N/g soil (SE)}$. Nitrate availability consistently declined with increasing ECM-associated basal area ($p < 0.01$, Table 3.2, Figure 3.3B), with mean nitrate availability in stands dominated exclusively by ECM-associated trees ($0.242 \pm 0.065 \mu\text{g NO}_3\text{-N/g soil}$) significantly reduced compared to mean nitrate availability in stands dominated by AM-associated trees ($2.13 \pm 0.461 \mu\text{g NO}_3\text{-N/g soil}$, Figure 3.3B, Table 3.1).

In contrast, ammonium availability was not correlated with ECM-associated basal area ($p = 0.98$, Table 3.2, Figure 3.3A) and was not significantly different between AM-dominated and ECM-dominated stands ($p = 0.99$, Table 3.1). Ammonium availability ranged from 0 to $7.327 \mu\text{g NO}_3\text{-N/g soil}$ with a mean of $2.84 \pm 0.330 \mu\text{g NH}_4\text{-N/g soil}$ (Figure 3.2A). On average, NH_4 availability exceeded NO_3 availability. However, we

observed differing patterns between tracts, with plots located in Indiana showing lower mean $\text{NO}_3\text{-N}$ availability (0.55 ± 0.363) and higher mean $\text{NH}_4\text{-N}$ availability (4.869 ± 0.401) compared to plots in Massachusetts ($1.21 \pm 0.23 \mu\text{g NO}_3\text{-N/g soil}$ and $1.6 \pm 0.299 \mu\text{g NH}_4\text{-N/g soil}$).

The ratio of NO_3 to NH_4 across plots was 0.98 ± 2.1 , indicating comparable pool sizes on average. However, this ratio tended to be much higher at Massachusetts sites (1.4 ± 2.5), compared to at the Indiana sites (0.46 ± 1.2), where NH_4 pool sizes were larger than NO_3 pool sizes on average. Across all plots, this ratio was negatively correlated with ECM abundance, indicating greater availability of NO_3 in AM-dominated stands ($P < 0.05$, Table 3.2).

Soil N ranged from 0.26 to 1.23% by mass, with a mean of $0.26 \pm 0.028\%$. Soil N declined with increasing ECM-associated basal area ($p < 0.01$, Figure 3.3C, Table 3.2). Similarly, the ratio of soil C:N increased with increasing ECM-associated basal area ($p < 0.001$, Table 3.2, Figure 3.3D), indicating reductions in the availability of N relative to C.

Nitrogen cycling rates

Rates of potential N-fixation, net nitrification and denitrification all showed negative relationships with ECM-associated basal area across the gradients. Potential N-fixation rates ranged from 0 to $139.9 \mu\text{g N/kg soil/day}$, with a mean of $32.4 \pm 4.7 \mu\text{g N/kg soil/day}$. We observed consistent declines of potential N-fixation rates with increasing ECM-associated basal area ($P < 0.05$, Figure 3.5A, Table 3.2). Potential N-fixation rates in ECM-dominated stands were approximately 45% those of rates in AM-dominated stands

on average, with mean potential N-fixation rates of $21.9 \pm 5.8 \mu\text{g N/kg soil/day}$ in ECM-dominated stands compared to mean potential N-fixation rates of $10.24 \pm 0.456 \mu\text{g N/kg soil/day}$ in AM-dominated stands (Table 3.1, Figure 3.4A). Ammonium availability was a significant predictor of potential N fixation rates, with greater availability of NH_4 in plots with high rates of potential N fixation.

Net nitrification rates ranged from 0 to $0.97 \mu\text{g NO}_3\text{-N/g soil/day}$, with a mean rate of $0.18 \pm 0.034 \mu\text{g NO}_3\text{-N/g soil/day}$. Net nitrification rates declined with increasing ECM-associated basal area ($p < 0.001$, Figure 3.5B, Table 3.2), and mean net nitrification rates were several times greater in AM-dominated stands ($0.34 \pm 0.077 \mu\text{g NO}_3\text{-N/g soil/day}$) compared to ECM-dominated stands ($0.04 \pm 0.008 \mu\text{g NO}_3\text{-N/g soil/day}$, Figure 3.4B, Table 3.1). Nitrate availability was significantly correlated with net nitrification rates.

Denitrification rates ranged from 0 to $22.1 \mu\text{g N/kg soil/day}$, with a mean rate of $4.77 \pm 1.34 \mu\text{g N/kg soil/day}$. Denitrification rates declined with increasing ECM-associated basal area ($p = 0.06$; Figure 3.5C, Table 3.2) and were significantly reduced in ECM-dominated stands ($2.37 \pm 1.28 \mu\text{g N/kg soil/day}$) compared to AM-dominated stands ($7.07 \pm 2.66 \mu\text{g N/kg soil/day}$, Figure 3.4C, Table 3.1). Ammonium availability was a significant predictor of denitrification rates, with greater availability of NH_4 in plots with high rates of denitrification ($P < 0.001$).

Stepwise AIC model selection did not support inclusion of total basal area, *Acer* *Saccharum* basal area, or pH as predictors of N-fixation. The model including proportion ECM basal area and NH_4 and NO_3 concentrations was supported (AIC=155.11) in favor of the model including all predictors (AIC=162.67). Stepwise AIC model selection did

not support inclusion of total basal area, *Acer Saccharum* basal area, pH, or inorganic N concentrations as predictors of net nitrification. The model with ECM basal area as the sole predictor (AIC=-152) was favored over the full model (AIC=-143.49). Stepwise AIC model selection did not support inclusion of total basal area, *Acer Saccharum* basal area, or pH as predictors of denitrification. The model with ECM basal area as the sole predictor (AIC=91.71) was favored over the full model (AIC=97.21). Thus, for all three N cycle fluxes measured, proportion of ECM basal area was supported as a predictor of flux rates outside of the effects of *Acer Saccharum* basal area, total basal area or soil pH.

Gene abundances

Gene abundances for *nifH* were not related to ECM-associated basal area (Figure 3.6A), rates of potential N-fixation or NH₄ pool sizes. Gene abundances for *amoA* were not differentially associated with ECM-associated basal area (Figure 3.6B), rates of nitrification, or pool sizes of NO₃ and NH₄. Gene abundances for *nosZ* were not differentially associated with ECM-associated basal area (Figure 3.6C) or with rates of denitrification. Soil fraction (bulk versus rhizosphere) was a significant predictor of gene abundance for *nosZ* only.

Discussion

Interactions between mycorrhizal fungi and free-living microbes have significant consequences for biogeochemical cycling, but it not known whether these are mediated by differences in the activities or abundances of N cycling bacterial functional groups. In this

study, we paired measurements of rates of potential N-fixation, nitrification and denitrification rates with measurements of the abundances of corresponding bacterial functional groups across gradients of mycorrhizal abundance. These data allowed us to first test whether mycorrhizal abundance was associated with differences in N cycling bacterial activity. Consistent with this hypothesis, we observed decreased rates of net nitrification, denitrification and inorganic N availability across multiple temperate forest sites in the presence of ECM fungi. However, we also observed reduced rates of N-fixation in ECM-dominated stands. Second, we tested for relationships between gene abundances and corresponding flux rates across the mycorrhizal gradients. In contrast to our hypothesis, we did not find evidence for structuring of N-cycling bacterial functional groups along AM-ECM gradients, indicating a potential decoupling of flux rates from functional group abundances.

Rates of potential N-fixation, net nitrification and denitrification each showed negative correlations with ECM-associated basal area across the gradients, indicating that mycorrhizal abundance may play a strong control on N cycling bacterial activity. N-fixation represents the only biotic pathway by which new N enters ecosystems, making it essential to characterize in N-limited ecosystems (Reed et al. 2011). The stands we explored in this study lack tree species that form symbiotic associations with N-fixers, making free-living diazotrophs in soil the only source of new N inputs, outside of N deposition. N-fixation is an energetically costly strategy, and free-living diazotrophs may switch between particular N acquisition pathways depending on favorability (Norman and Friesen 2017). While rates of N-fixation are relatively small compared to other N-cycle

fluxes, they play an important role over long-terms by allowing N to accumulate (Reed et al. 2011). We observed N-fixation rates in AM-dominated stands nearly twice those in ECM-dominated stands, suggesting that ECM-associated suppression of bacterial activity may have substantial implications for N accumulation over time. Although multiple processes are responsible for the production of ammonium in soil, including both mineralization of organic N during decomposition and N-fixation, potential N-fixation rates were correlated with $\text{NH}_4\text{-N}$ pool sizes.

Nitrification represents a pathway for potential loss of inorganic N from soil as NO_3 in streamwater or through subsequent release in gaseous forms via denitrification. This flux requires the activities of ammonia-oxidizers and nitrite-oxidizers, which are comprised of both bacteria and archaea (Levy-Booth et al. 2014). Consistent with other studies, we found the highest rates of net nitrification in AM-associated stands, while ECM-associated stands tended to show near-zero rates of net nitrification (Phillips et al. 2013). Likewise, the availability of inorganic N in the form of NO_3 was highest in AM-associated stands and declined with increasing ECM-associated basal area.

Following nitrification, inorganic N becomes available for gaseous loss as N_2O or N_2 from soil through denitrification. This microbial process accounts for approximately one-third of N outputs from terrestrial ecosystems globally, and can contribute to N-limitation (Houlton and Bai 2009). Over the 24-hour incubation time that we conducted denitrification measurements, we only observed production of N_2O based on consumption of isotopically-labeled NO_3 . Rates of $\text{N}_2\text{O-N}$ production declined with increasing ECM-

associated basal area, with denitrification rates in AM-associated stands 3.5 times greater than those in ECM-associated stands on average.

Consistent with global patterns of increased C storage per unit N in ECM-dominated systems (Averill et al. 2014), soil C:N increased with ECM-associated basal area in the gradients explored in this study. Increased C sequestration relative to N availability in ECM-dominated systems is primarily attributed to slowed decomposition due to competition between free-living heterotrophs and ECM fungi; however, ECM associated plants may also produce more recalcitrant litter, which could potentially also contribute to wider soil C:N ratios in these systems (Midgley and Phillips 2014). The results presented here are limited to the upper 15 cm of the mineral soil horizon. One recent study suggests that temperate AM-associated forest stands store more C at depth because fast decomposing leaf litter is correlated with the accumulation of microbial residues in these sites, suggesting it may be important to explore these patterns across a broader depth profile and understand factors contributing to soil C storage (Craig et al. 2018).

Although we observed strong structuring of N cycling bacterial activity in relation to mycorrhizal abundance, we did not find strong patterns in bacterial functional group abundance. Prior work, including sites overlapping with the present study, indicate that free-living microbial communities in ECM-associated stands are composed of fewer bacteria relative to fungi when compared with communities in AM-associated stands (Cheeke et al. 2017). Similarly, field-exclusion of ECM fungi results in increases in bacterial abundance (Averill and Hawkes 2016). While fungal:bacterial ratios may differ across gradients of mycorrhizal abundance, our findings suggest that the functional profile

of the bacterial population can be highly variable. Furthermore, the lack of correlation between functional gene abundances and corresponding flux rates indicates that the presence of particular functional groups is not itself a sufficient indicator of a process occurring.

Identifying relationships between gene abundances and ecosystem function has remained a challenge in microbial biogeochemistry (Bier et al. 2015, Rocca et al. 2015, Trivedi et al. 2016). Although commonly assumed otherwise, few forest soil studies find significant gene-function correlations for C and N cycle fluxes (Rocca et al. 2015). Several potential mechanisms have been cited to explain frequent observations of decoupling between flux rates and gene abundances. Gene abundances are thought to better predict process rates than transcript abundances due to the rapid degradation of RNA in the environment; however, gene abundances may also represent dormant and inactive communities. This is a particularly important consideration for processes like N-fixation which are known to be tightly regulated based on abiotic factors, such as resource demand (Reed et al. 2011).

An additional challenge with exploring gene-function relationships involves consideration of the scales at which gene abundances vary compared to flux rates. The abundances of diazotrophs, nitrifiers and denitrifiers have been shown to vary widely over short timespans and across small distances, while flux rates tend to remain more stable within systems over time (Regan et al. 2017). Determining whether frequent observations of nonsignificant relationships between community structure and associated processes are due to a lack of adequate sampling or due to actual biological decoupling will be necessary

to fully understand these patterns. Finally, it is important to note that we only characterized a single gene associated with each flux, but there are additional gene targets that could potentially track flux rates more closely, including archaeal contributions to nitrification and other subunits of the nitrous oxide reductase enzyme (Levy-Booth et al. 2014).

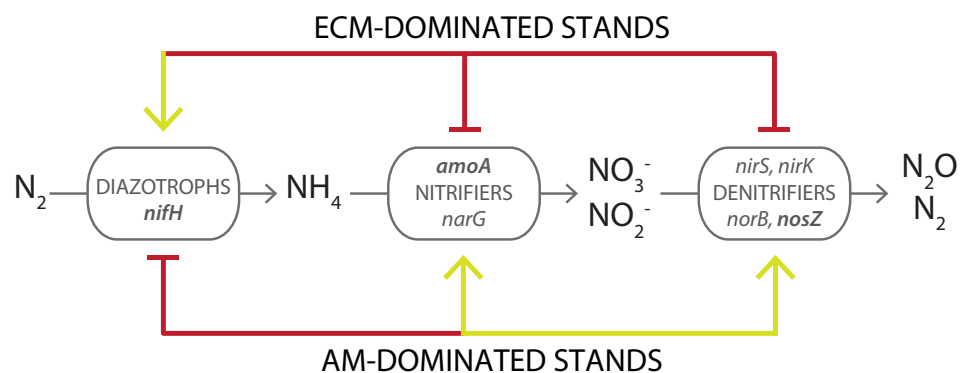
Overall, we find that mycorrhizal type is associated with major differences in N cycling rates and N availability. Forest stands dominated by ECM-associated fungi show suppressed rates of N-fixation, suggesting that the accumulation of new N into ECM-associated ecosystems may be limited due to competitive interactions between ECM fungi and free-living heterotrophic diazotrophs. We additionally observed suppressed rates of nitrification and denitrification in the presence of ECM fungi, indicating reduced rates of N losses from these systems. We found that these differences in fluxes may not underlain by differences in bacterial functional group abundances. As gene abundances are likely to vary across smaller spatial and temporal scales than flux rates, characterizing functional group abundances with greater resolution may help better elucidate the coupling of structure and flux measurements.

Identifying the factors that control microbial N cycling is critical to improving our capacity to quantify inputs and exports of N from temperate forests (Yanai et al. 2012). This analysis indicates that stand-level mycorrhizal types are associated with significant differences in N cycling activity, and these are likely to have substantial quantitative impacts on N cycling budgets. However, future studies pairing biogeochemical flux measurements with comprehensive microbial community composition data will be necessary to better characterize relationships between community structure and N cycling.

In particular, it will be important to identify factors accounting for potential decoupling of N cycling bacterial functional groups from associated flux rates in temperate forests.

Tables and Figures

Figure 3.1. — Framework of N cycling in ECM and AM stands.



Green arrows indicate an expectation of an increased flux and functional group abundance in a particular stand type (ECM above, AM below). Red arrows indicate an expectation of a decreased flux and functional group abundance in a particular stand type. Genes associated with each flux are shown in italics. Genes in bold are ones measured in this study.

Figure 3.2 — Boxplots of soil N pools by plot type.

(A) NH_4 ($\mu\text{g NH}_4\text{-N/g soil}$), (B) NO_3 ($\mu\text{g NO}_3\text{-N/g soil}$), (C) soil percent N, and (D) soil C:N classified by plot proportion of basal area: ECM ($\geq 70\%$ ECM-associated), AM ($\leq 30\%$ ECM-associated), MIX, (30-70% ECM). Letters indicated significantly different groups ($p < 0.05$).

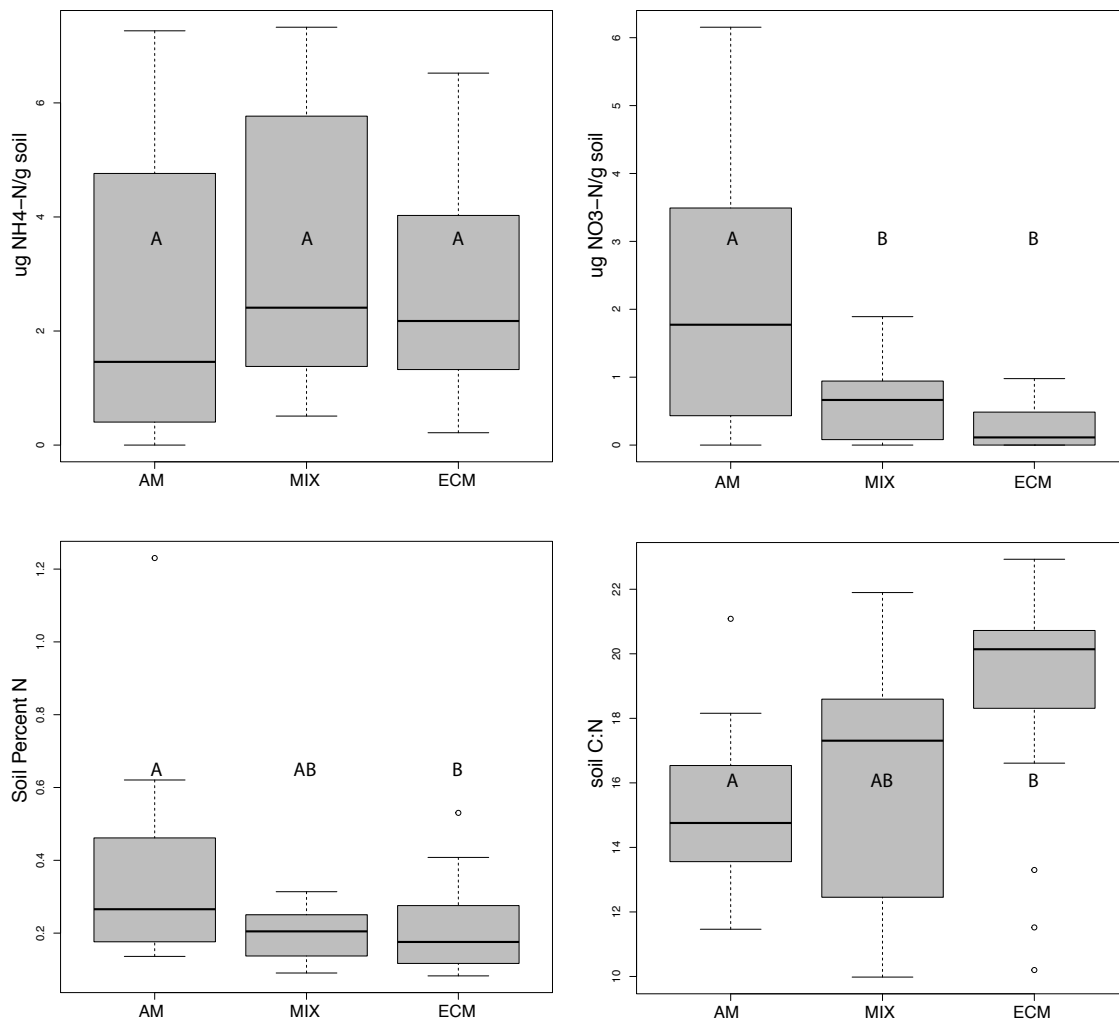


Figure 3.3 — Soil N pools across gradients.

(A) NH_4 ($\mu\text{g NH}_4\text{-N/g soil}$), (B) NO_3 ($\mu\text{g NO}_3\text{-N/g soil}$), (C) soil percent N, and (D) soil C:N versus proportion of plot area associated with ECM.

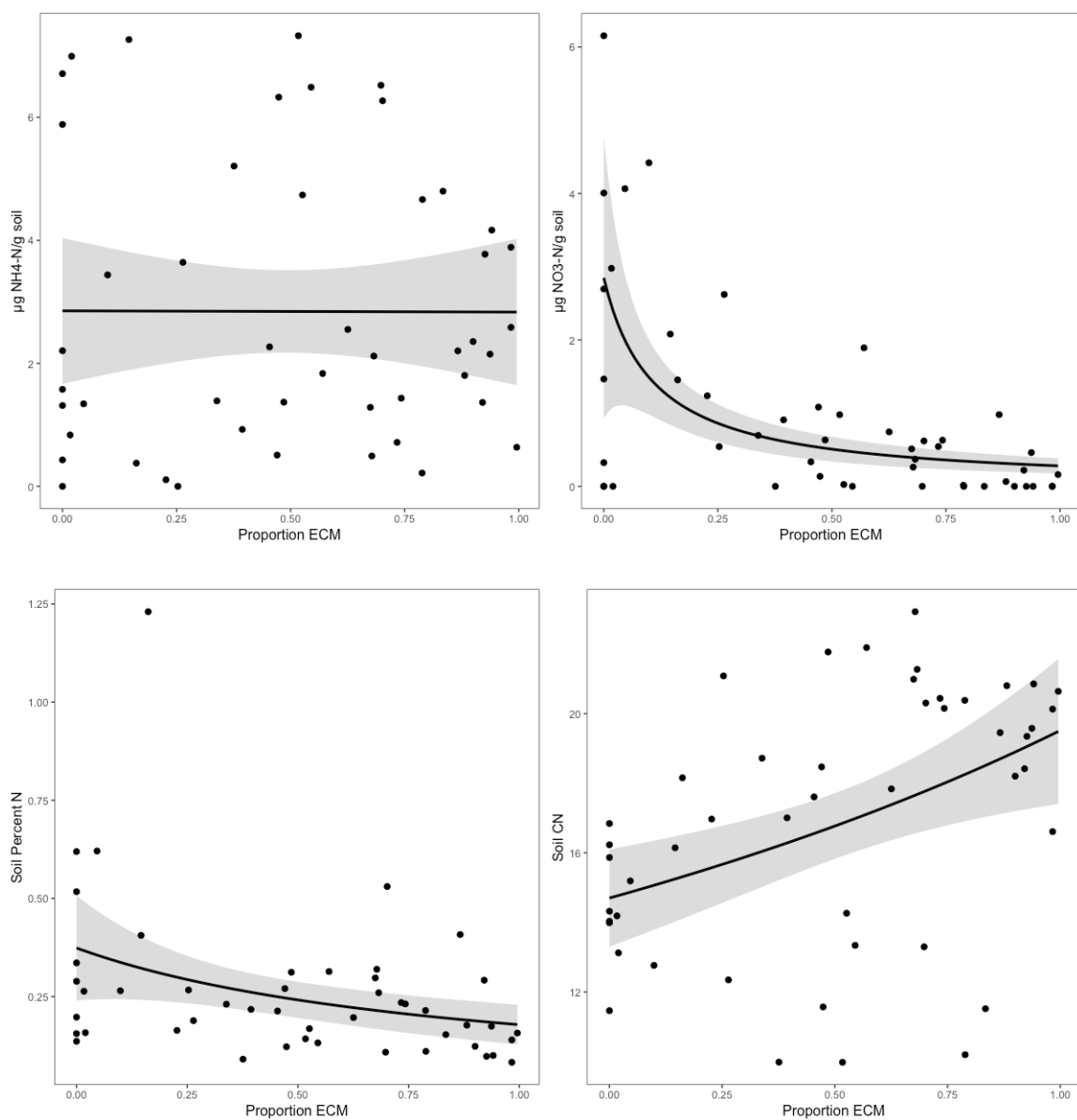


Figure 3.4 — Boxplots of N cycle fluxes by plot type.

Flux rates classified by plot proportion of basal area: ECM ($\geq 70\%$ ECM-associated), AM ($\leq 30\%$ ECM-associated), MIX, (30-70% ECM) for (A) Fixation ($\mu\text{g N/kg soil/day}$), (B) Net Nitrification ($\mu\text{g NO}_3\text{-N/g soil/day}$) and (C) Denitrification ($\mu\text{g N}_2\text{O-N/kg soil/day}$). Letters indicated significantly different groups ($p < 0.05$).

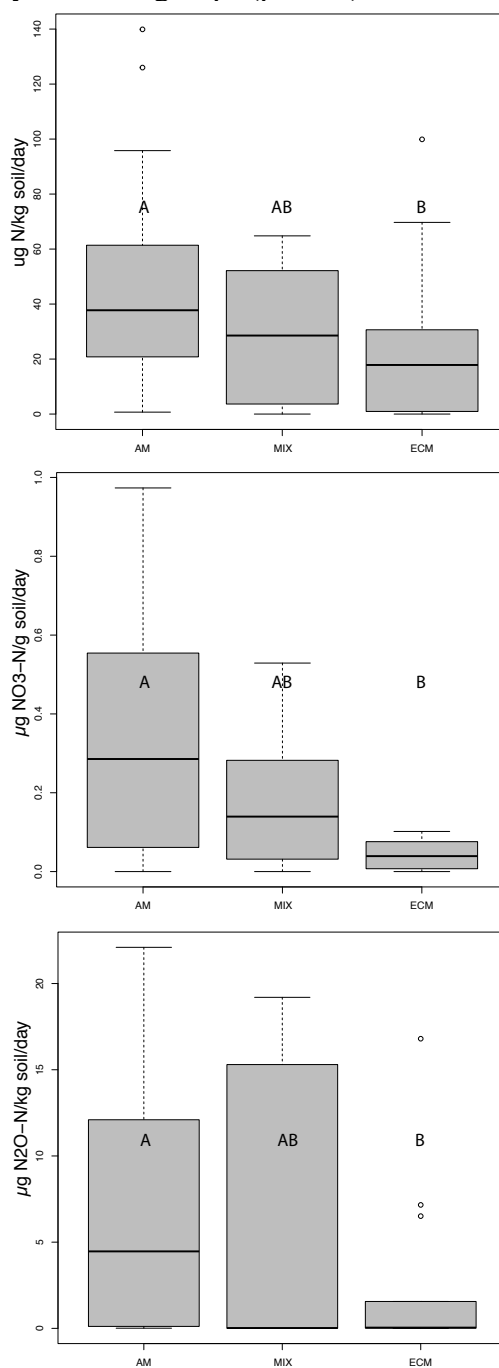


Figure 3.5 — Regressions of N cycle fluxes against proportion ECM.

Flux rates versus proportion ECM-associated basal area. (A) Fixation ($\mu\text{g N/kg soil/day}$), (B) Net Nitrification ($\mu\text{g NO}_3\text{-N/g soil/day}$), (C) Denitrification ($\mu\text{g N}_2\text{O-N/kg soil/day}$).

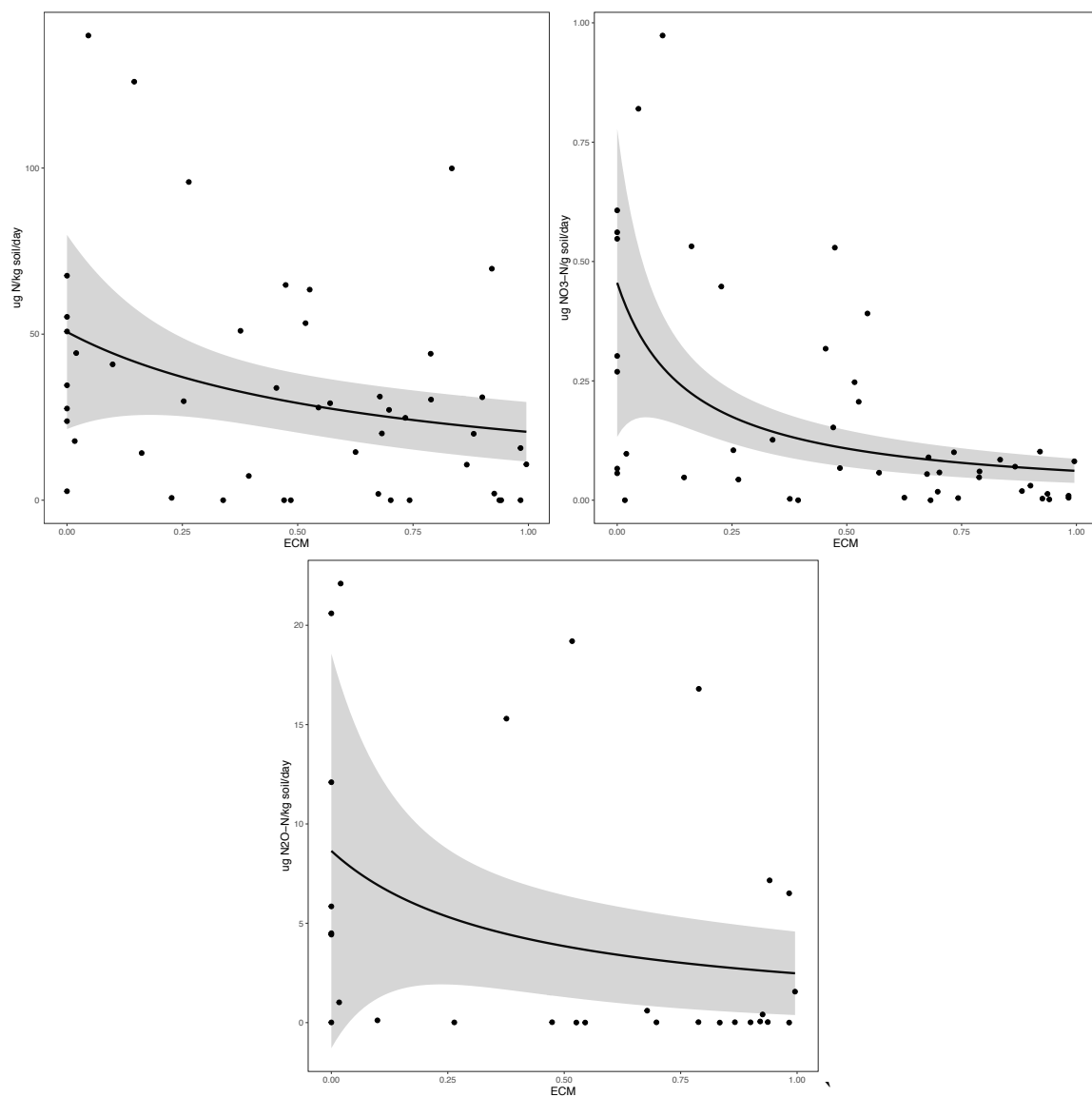
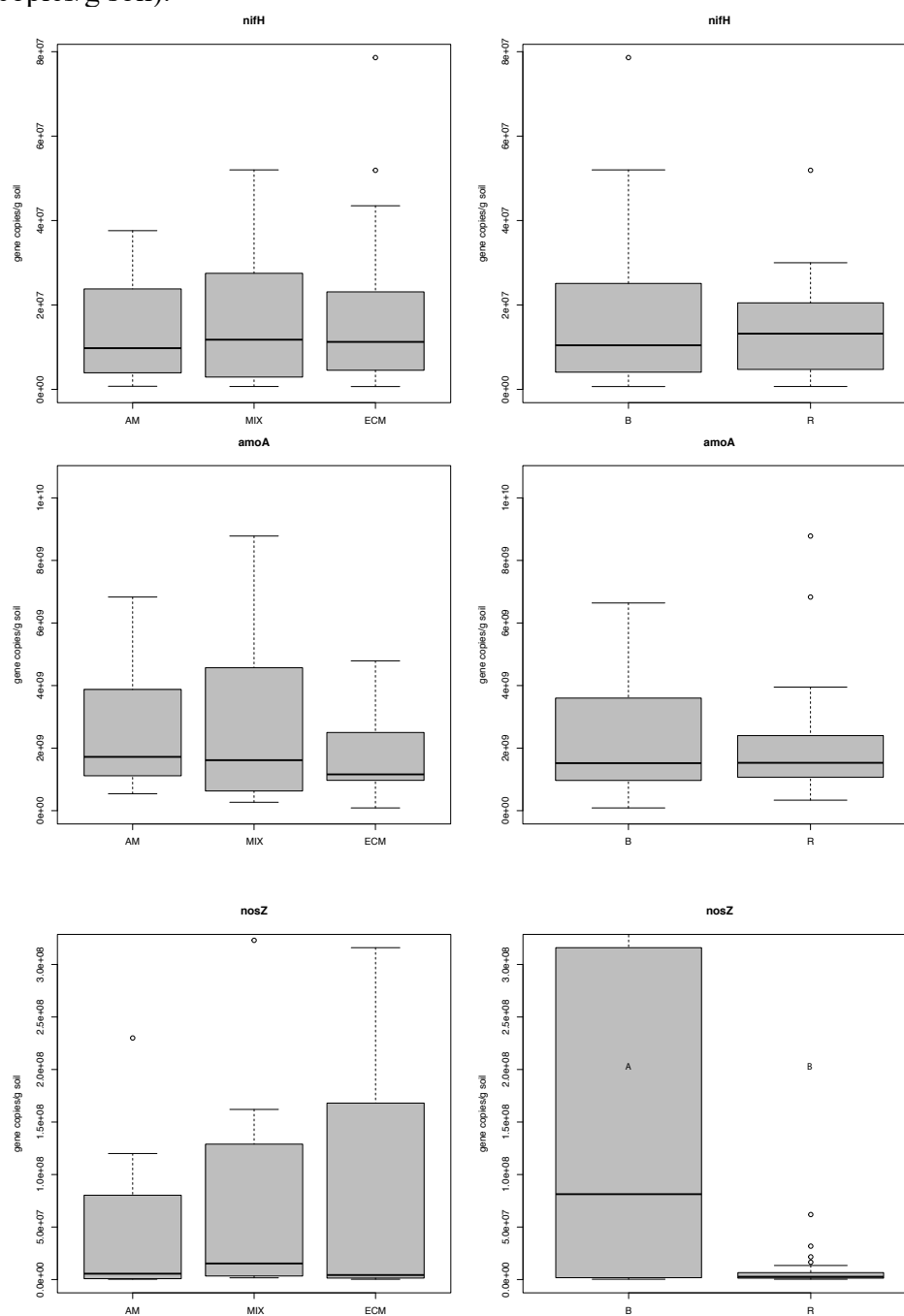


Figure 3.6 — Boxplots of gene abundances by plot type.

Boxplots of gene abundances by plot proportion of basal area: ECM ($\geq 70\%$ ECM-associated), AM ($\leq 30\%$ ECM-associated), MIX, (30-70% ECM) and soil fraction (B=Bulk, R=Rhizosphere) for (A) *nifH* (copies/g soil), (B) *amoA* (copies/g soil), and (C) *nosZ* (copies/g soil).



	ECM		AM	
	Median	IQR	Median	IQR
NO ₃ ***	0.24	0.51	2.08	2.60
NH ₄	1.31	1.74	1.31	3.06
NO ₃ :NH ₄ **	0.17	0.38	3.03	5.77
SOIL N*	0.22	0.14	0.29	0.29
SOIL C:N**	20.31	1.38	15.19	2.50
POTENTIAL N-FIXATION*	17.85	29.05	34.60	43.60
NET NITRIFICATION***	0.05	0.07	0.30	0.49
DENITRIFICATION*	0.04	0.82	4.43	5.28

Table 3.1 — N pools and fluxes summary

IQR = difference between 75% and 25% quartiles. *** p<0.001, ** p<0.01, * p<0.05 from Tukey test.

	STANDARD ERROR	T-VALUE	P-VALUE
NO ₃	0.69	4.67	2.62 x 10 ⁻⁵
NH ₄	0.12	0.02	0.98
NO ₃ :NH ₄	1.06	2.43	0.02
SOIL N (%)	1.04	2.81	0.007
SOIL C:N	4.96 x 10 ⁻³	-3.38	1.45 x 10 ⁻³
POTENTIAL N-FIXATION	0.01	2.16	0.04
NET NITRIFICATION	3.57	3.95	2.70 x 10 ⁻⁴
DENITRIFICATION	0.17	1.95	0.06

Table 3.2 — GLM regression summaries for N pools and fluxes against plot proportion ECM

CHAPTER FOUR: IDENTIFYING DATA NEEDS AND ASSESSING UNCERTAINTY IN A COUPLED MICROBIAL SOIL C AND N DECOMPOSITION MODEL

Abstract

Ecosystem models can provide useful insights into understanding coupled C and N cycling and predicting potential biogeochemical responses to global change. Recent advancements have allowed soil C and N cycle models to incorporate increasingly accurate representations of microbial physiology and enzyme-mediated depolymerization of soil organic matter. A major challenge with these model structural improvements involves the requirement for additional parameters, which are often poorly constrained, and thus sources of uncertainty. I use DAMM-MCNiP, a microbially-mediated, coupled soil C and N cycling model, as a tool to explore the influence of microbial physiological and enzyme kinetic parameters on model estimates of coupled soil C and N cycling. I then quantify the potential for constraining model parameters using empirical measurements of soil respiration, and I use simulated data to assess the potential for future data collection to constrain particular parameters. I find that modeled soil C and N pools and fluxes are disproportionately sensitive to only a few parameters (e.g. activation energies and CUE), while others exert minimal influence (e.g. Michaelis-Menten half-saturation constants). While some parameters can be constrained by the available data on heterotrophic

respiration, the collection of additional data on dissolved organic matter is identified as a potentially useful constraint on most parameters. Improving model representations of microbially-mediated soil C and N cycling will require closer consideration of model uncertainties and targeting data collection towards reducing these uncertainties.

Introduction

The global cycling of C and N are closely coupled in terrestrial ecosystems as N is a required constituent of nucleic acids, proteins, enzymes and biomass. Soil microorganisms play a central role in linking these cycles through the decomposition of soil organic matter, followed by assimilation of C and N into microbial biomass, partitioning of these elements among metabolic pathways, and through the production of extracellular enzymes. Despite the activities of ubiquitous microbial decomposers, soils sequester vast quantities of organic matter (Jobbágy and Jackson 2000, Stockmann et al. 2013, Sanderman et al. 2017). Identifying the complex drivers of changes in these stocks, including soil microbial physiology, soil chemistry, enzyme activities, and environmental factors, has remained a primary challenge in the field (Allison et al. 2010, Schmidt et al. 2011).

Soil biogeochemistry models provide a useful framework for testing hypotheses on C and N cycling and predicting ecosystem responses to change (Manzoni and Porporato 2009, Blankinship et al. 2018). These types of model have increased in complexity over the past decades with growing information on linkages between soil C and N cycling, microbial physiology, enzyme kinetics and their responses to temperature and moisture (Manzoni and Porporato 2009). Advancements in the representation of N cycling in these models to account for the potential for N limitation has been shown to improve their capacity to predict C sequestration, which is often over-estimated in C-only models (Hungate et al. 2003, Thornton et al. 2007, Zaehle 2013). An additional development in modeling C and N cycling has been the incorporation of more explicit representations of

soil microbial physiology and the role of microbial groups in producing extracellular enzymes (Schimel and Weintraub 2003). These advancements have allowed models to more accurately capture responses of soil respiration to warming (Allison et al. 2010), predict the spatial distribution of soil C (Wieder et al. 2013), and broadly improve our understanding of biogeochemical responses to environmental change (e.g. Sistla et al. 2014, Wieder et al. 2014).

With growing complexity, biogeochemical models are sometimes able to more accurately represent our understanding of microbial C and N cycling; however, the addition of new model compartments and parameters also has the potential for increasing sources of uncertainty (Manzoni and Porporato 2009, Shi et al. 2018). Several recent model comparison studies demonstrate divergent model predictions from closely-related soil C cycle models. For example, even when given identical forcing data, models disagree on steady state soil C stocks and the seasonal dynamics of respiration, due to a combination of differences in model forms and parameterization (Wieder et al. 2018). Furthermore, confronting these models with existing long-term datasets on C cycle responses to warming or litter inputs has not been useful for discriminating amongst conflicting models, making it critical to identify specific data constraints that can be used to inform models (Sulman et al. 2018).

Soil biogeochemical models are subject to a number of sources of uncertainty, including model structural uncertainty (Ajami and Gu 2010, Myrgiotis et al. 2018). For example, microbial processes are often represented by single microbial biomass pools, with more recent models shifting to include multiple functional guilds or distinguishing between

different microbial functional types (Moorhead and Sinsabaugh 2006, Waring et al. 2013, Sistla et al. 2014, Wieder et al. 2014). Similarly, the number and types of specific C and N pools represented across models also varies, with a shift towards representing measurable, rather than conceptual pools (Abramoff et al. 2018).

A second source of uncertainty, which often receives less attention than model structural uncertainty, is related to parameters used in models. Most soil biogeochemical models use fixed parameter values selected *ad hoc* or passed on through model variations following single site-specific literature measurements. While some soil physical parameters are easy to directly measure, enzyme kinetic parameters and microbial physiological parameters can be challenging to identify. Specific enzyme kinetics are often measured in lab assays, but their translation into models has been challenging due to a mismatch between the conceptual or simplified substrates represented in models and the diversity of highly-specific extracellular enzymes, each with unique stoichiometries and sensitivities (Drake et al. 2013, Sinsabaugh et al. 2014, 2015). Microbial physiological parameters, such as the carbon use efficiency of microbial biomass production, are also particularly hard to quantify due to metabolic diversity across taxa and challenges with empirically measuring microbial physiology (Geyer et al. 2016, Saifuddin et al. *in review*). Despite large uncertainties in these parameter choices, soil C and N cycle models rarely account for parameter uncertainty explicitly.

Identifying approaches for reducing parameter uncertainty in soil biogeochemical models is necessary to overcome these challenges. We use a recent microbial C-N coupled decomposition model, DAMM-MCNP, to explore how parameters associated with

enzyme kinetics, microbial uptake of C and N, and microbial physiology impact estimates of C and N cycling (Abramoff et al. 2017). Using a combination of sensitivity analyses and a Bayesian statistical approach, we explore the following questions:

1. Which model parameters have the greatest impact on estimates of soil C and N cycling?
2. Can the assimilation of existing data reduce model parameter uncertainty, and what is the impact of reducing parameter uncertainty on model estimates?
3. What specific types of data have the potential to reduce model parameter uncertainty most effectively?

Methods

DAMM-MCNI_P

The Dual Arrhenius Michaelis-Menten Microbial Carbon and Nitrogen Physiology Model (DAMM-MCNI_P) represents coupled soil C and N cycling through soil organic stocks (SOCN), dissolved organic stocks (DOCN), microbial biomass, and extracellular enzyme pools based on soil temperature and soil moisture (Figure 4.1, Davidson et al. 2012, Finzi et al. 2015, Abramoff et al. 2017). The model utilizes a combination of Arrhenius and Michaelis-Menten kinetic equations to describe the depolymerization of SOCN, resulting in the production of DOCN. A parallel set of arrhenius and Michaelis-Menten kinetic equations describe the incorporation of DOCN into microbial biomass. A series of microbial physiological parameters determine how C and N are partitioned between microbial biomass, enzyme production and losses as respiration and N

mineralization. The model has been shown to capture seasonal patterns of heterotrophic respiration at trenched (root-free) plots in a temperate hardwood forest (Abramoff et al. 2017). Additionally, the model structure is similar to those of several recent microbial biogeochemical models across ecosystem types, making inferences regarding parameter uncertainty and model structure generalizable (Schimel 2003, Manzoni and Porporato 2009, Waring et al. 2013, Sistla et al. 2014, Abramoff et al. 2018).

In DAMM-MCNP, extracellular enzymes produced by microbes depolymerize a fraction of the SOCN pool according to equilibrium approximation kinetics, which are a generalization of Michaelis-Menten kinetics:

$$\text{Depolymerization rate} = Vmax_S \times \frac{[SOM_{avail}] \times [ENZ]}{K_{M_{dep}} + [SOM_{avail}] + [ENZ]} \quad (\text{Equation 4.1}).$$

SOM_{avail} , the fraction of the total SOCN pool available for depolymerization, is dependent on soil moisture (θ) and the fraction of SOCN that is not physically or chemically occluded from enzyme-binding (Magill et al. 2000):

$$SOM_{avail} = SOM \times Frac \times dLiq \times \theta^3 \quad (\text{Equation 4.2}).$$

The maximum rate of depolymerization is determined by soil temperature according to the Arrhenius function:

$$Vmax_S = \alpha_{dep} \times e^{\frac{Ea_{dep}}{RT}} \quad (\text{Equation 4.3}).$$

Thus, a total of three parameters ($K_{M_{dep}}$, α_{dep} and Ea_{dep}) are involved in specifying depolymerization rates based on temperature and the pool sizes of SOCN and enzymes.

Following depolymerization, SOCN enters the DOCN pool, which is available for uptake by microbial biomass. Uptake rate is determined by Michaelis-Menten kinetics, which are dependent on the concentration of DOCN and the concentration of O₂ based on soil moisture:

$$Uptake\ rate = Vmax_D \times \frac{[DOM]}{Km_{upt} + [DOM]} \times \frac{[O_2]}{Km_{O_2} + [O_2]} \quad (Equation\ 4.4).$$

The maximum rate of uptake is specified by the Arrhenius function:

$$Vmax_D = \alpha_{upt} \times e^{-\frac{Ea_{upt}}{RT}} \quad (Equation\ 4.5).$$

Thus, four parameters (Km_{upt} , Km_{O_2} , a_{upt} and Ea_{upt}) are involved in specifying rates of DOCN uptake based on moisture as represented by soil O₂ concentration and DOCN pool sizes (Abramoff et al. 2017).

Following uptake, C and N are either retained within microbial biomass, allocated towards enzyme production, or lost through respiration and N mineralization. Respiration is first calculated as a fixed fraction of C uptake (1-CUE). Two separate parameters, p and q , determine maximum allocation of C and N towards enzyme production respectively, with actual allocations constrained by enzyme stoichiometry according to Liebig's law of the minimum. If there remain excess C or N from this initial allocation that cannot be incorporated into enzymes due to stoichiometric demand, remaining C and N are maximally incorporated into microbial biomass. Finally, if there remain excess C or N that

cannot be incorporated into biomass due to stoichiometric demand, it is lost as overflow respiration or N mineralization.

Parameter sensitivity

DAMM-MCNP requires a total of 25 parameters (Table 4.1). A subset of seven parameters are directly involved in the kinetic equations describing depolymerization and uptake, and three parameters are used to represent microbial physiology. The remaining 15 parameters are primarily associated with defining stoichiometries and soil physical properties. We assessed the sensitivity of model outputs to the ten depolymerization, uptake and physiological parameters. Each parameter was varied from 50% up to 150% of its default setting (Table 4.1) in 10% increments (i) for a total of 10 parameter settings (y). The associated values for respiration, SOC, SON, DOC, DON, microbial biomass C, and microbial biomass N were calculated at each parameter setting (x). The sensitivity, or change in model outputs relative to changes in parameter values was calculated over each pair of parameter settings:

$$Sensitivity_{i,x,y} = \frac{Output_{x+i} - Output_x}{Output_x} \times \left(\frac{Parameter_{y+i} - Parameter_y}{Parameter_y} \right)^{-1} \quad (Equation 4.6).$$

This approach scales rates of change by both pool size and parameter size to allow for congruent comparisons across model outputs and parameter types, which can vary by orders of magnitude.

Parameter estimation

We utilized a Bayesian statistical framework to explore the potential for constraining parameter uncertainty through data assimilation (Luo et al. 2009, Dietze et al. 2014, Shi et al. 2016). This approach involves providing prior distributions describing the potential range of values for each parameter of interest. The likelihood of the observed data given particular parameter selections from within the prior distribution is then evaluated. This process is repeated manyfold to generate a posterior distribution describing how likely particular parameter values are given prior constraints and data. In this analysis, we utilized broad, uniform prior distributions throughout all simulations ($\pm 50\%$ of default parameter values) to explore a wide range of parameter options outside of those currently incorporated in the published model (Abramoff et al. 2017). Differential Evolution Markov Chain Monte Carlo (DEzs-MCMC, Ter Braak and Vrugte 2008) simulations were performed to evaluate the likelihood across parameter space using the BayesianTools package (Hartig et al. 2018) in RStudio (R core team 2007). We used a normally-distributed likelihood function and performed simulations using three chains, each with 20,000 to 100,000 iterations as needed to achieve convergence based on Gelman-Ruben potential scale reduction factors ($\text{psrf} < 1.3$). We estimated all ten parameters of interest in tandem or estimated parameters of a given category (depolymerization, uptake, or physiology) while holding others fixed.

We first assimilated published, field measurements of heterotrophic respiration from trenched plots at the Harvard Forest Long-Term Ecological Research Site to constrain depolymerization, uptake and physiological parameters in DAMM-MCNI_P. The respiration data used in this analysis were collected by prior researchers using automated

soil respiration chambers monitoring CO₂ efflux from soil at half-hourly increments across the growing season in 2009 (Savage et al. 2008, Davidson et al. 2012). Simultaneous, automated measurements of soil moisture and soil temperature were also recorded (Savage et al. 2008, Davidson et al. 2012).

Several of the model pools and fluxes in DAMM-MCNiP represent mechanistic processes upstream of respiration which currently lack comparable, high-resolution observational data (Figure 4.1). To identify which of these datasets would be most useful for future collection and to explore how additional data constraints might impact parameter estimation, we simulated data for SOCN, DOCN and microbial biomass pools and respiration rates using default parameter values and added normally distributed observation error. We then used this simulated data to estimate parameters and compare resulting posterior parameter distributions with the known parameter values used to simulate data. We calculated the percent difference between the *maximum a posteriori* (MAP) parameter estimate and the parameter value used in simulation to measure the accuracy of parameter estimation. We measured the precision of parameter estimates by comparing the range of the posterior distribution (2.5% to 97.5% interval) to the range of the prior distribution for each parameter:

$$Uncertainty\ Reduction = \left(1 - \frac{Posterior_{97.5} - Posterior_{2.5}}{Prior_{max} - Prior_{min}}\right) \times 100 \quad (Equation\ 4.7).$$

A reduction in the range of the posterior distribution relative to the prior distribution indicates that assimilation of the provided data has provided some constraint on the range of parameter values.

Model projections

For parameters that could be constrained through assimilation of respiration data, we assessed the potential impacts of new parameter estimates on long-term model projections by comparing model outputs using posterior parameter estimates to model outputs estimated using default parameters. Annual seasonal cycles of temperature, moisture, and litterfall inputs were repeated for 100 years. Annual totals were calculated for respiration rates and annual means were calculated for pool sizes of SOC, SON, DOC, DON, microbial biomass C and microbial biomass N. Annual estimates at median parameter settings and their uncertainties based on 2.5 and 97.5% posterior parameter intervals were compared to annual estimates at default parameter settings and their uncertainties based on the range of prior distributions.

Results

Parameter sensitivity

Model outputs are disproportionately sensitive to a few select parameters and largely insensitive to others (Figure 4.2). All model outputs are most sensitive to either the activation energy of depolymerization (Ea_{dep}) or the activation energy of uptake (Ea_{upt}). DOCN pool sizes and respiration rates are most sensitive to parameter variation on average, while microbial biomass pools are the least sensitive. SOCN pool sizes are almost exclusively sensitive to Ea_{dep} , with sensitivities below 0.02 for all other parameters. The low sensitivity of SOCN pools to most parameters is amplified by the fact that only a small

fraction of total SOCN is available for enzymatic activity due to chemical and physical protection (Magill et al. 2000). DOCN pool sizes show sensitivities greater than 0.05 for all parameters except Km_{dep} , which shows the lowest sensitivity across all model outputs. Microbial biomass pools are most sensitive to Ea_{dep} , Ea_{upt} , CUE, and p . The sensitivity of respiration to Ea_{upt} is the largest sensitivity value observed, due to near-zero rates of respiration resulting when Ea_{upt} is raised above approximately 95 kJ mol⁻¹. In contrast, respiration rates are nearly insensitive to the other uptake parameters.

Respiration data assimilation

Assimilating seasonal heterotrophic respiration data resulted in large reductions in parameter uncertainty for Ea_{dep} , Ea_{upt} , and CUE (Figure 4.3, Figure 4.5A, Table 4.2). Although respiration was sensitive to a_{dep} , this parameter did not show a reduction in uncertainty when provided with respiration data. The parameters used to define allocations of C and N to enzyme production (p and q) showed reductions in uncertainty, but were highly-skewed towards the maximum values allowed in the prior specification. Thus, assimilating respiration data alone allowed for large reductions in parameter uncertainty for some of the most sensitive parameters (e.g. Ea_{dep} , Ea_{upt} , and CUE), while most other parameters require additional data constraints. Model predictions using median parameter estimates produced through the assimilation of the respiration data were able to recreate the observed seasonal pattern in heterotrophic respiration and explain over 60% of variation in the data (Figure 4.4; $R^2_{adj}=0.61$, RMSE = 0.24).

Parameter estimation with simulated data

Using simulated respiration data with known parameter values was useful for corroborating the patterns observed with field measurements of respiration and identifying the potential for constraining parameter estimates with a more complete respiration dataset (no missing timepoints). Estimating parameters following assimilation of simulated respiration data with known, default parameter values indicates that respiration data alone can reliably reduce parameter uncertainty with high accuracy for Ea_{dep} , Ea_{upt} , CUE , and q (Figure 4.5B). However, the remaining six parameters showed low accuracy and limited reductions in uncertainty, indicating that they cannot be constrained even with a more complete respiration dataset alone.

Providing the model with data on SOCN pool sizes only reduced uncertainty in Ea_{dep} , while other parameters showed a combination of low reductions in uncertainty and inaccurate MAP values (Figure 4.5C). These observations are consistent with the strong sensitivity of SOCN pool sizes to Ea_{dep} (Figure 4.2). In contrast, assimilating data on DOCN pool sizes reduced parameter uncertainty for several parameters (Km_{O_2} , Ea_{upt} , q , CUE , Ea_{dep}) with fairly high accuracy (Figure 4.5D). Microbial biomass data was primarily useful for reducing uncertainty in Ea_{dep} and the microbial physiological parameters q and CUE (Figure 4.5E).

The most useful dataset for constraining all three depolymerization parameters and all three physiological parameters was DOCN data (Figure 4.5D). Parameters associated with uptake were the most challenging to constrain, with Km_{upt} largely unidentifiable by any provided data. While reductions in uncertainty in uptake parameters were relatively

small regardless of data type provided, DOCN showed the greatest overall uncertainty reduction and accuracy.

Model projections

Incorporating parameter estimates for Ea_{dep} , Ea_{upt} , and CUE based on assimilated respiration data results in major changes in estimates of specific pools and fluxes (Figure 4.7). A 4% decrease in Ea_{dep} results in large declines in SOCN stocks over time (Figure 4.7A), while other pools and fluxes equilibrate towards similar values. This reduction in Ea_{dep} represents a reduced barrier for SOCN depolymerization, which frees a larger fraction of the initial SOCN pool to be released as DOCN before eventually stabilizing. The increased availability of SOCN is temporarily associated with increases in DOCN and microbial biomass. A 17% increase in Ea_{upt} results in large, sustained accumulations of DOCN (Figure 4.7B), with minimal impact on other model outputs. A 20% reduction of CUE results in sustained declines in microbial biomass, declining SOC stores due to reduced microbial inputs, and temporary increases in respiration, N mineralization, and DOC accumulation due to reduced uptake (Figure 4.7C). Reductions in parameter uncertainty result in narrower credible intervals for model projections compared to model projections based on prior parameter intervals for both C (Figure 4.8) and N pools and fluxes (Figure 4.9).

Discussion

Modeling coupled soil microbial C and N cycling is critical for understanding the drivers of soil organic matter sequestration, losses of C and N through mineralization, and their responses to global change. In particular, incorporating explicit representations of microbial physiology and extracellular enzyme activities has been helpful for developing more representative biogeochemical models (Schimel 2003, Allison et al. 2010, Wieder et al. 2013). A major challenge with modeling these processes involves accounting for multiple sources of uncertainty, and few models explicitly account for the potential impacts of parameter uncertainty when making model estimates (Luo et al. 2016, 2017).

We used the ecosystem model DAMM-MCNP to explore how parameters associated with the enzyme kinetics of depolymerization, microbial uptake of C and N, and microbial metabolism impact estimates of C and N cycling. Second, we used a Bayesian statistical approach to constrain parameter estimates through data assimilation, and we identified specific parameters that could be constrained through existing respiration data alone. As several parameters required additional data constraints outside of the available respiration data, we used simulated data to identify which alternative datasets would be most useful targets for future data collection to reduce model parameter uncertainty. Last, we updated model parameters based on data assimilation and explored the potential impacts of these shifts in parameter values on long-term model projections.

Which model parameters have the greatest impact on estimates of soil C and N cycling?

DAMM-MCNP requires a total of 25 parameters, with 10 parameters involved in depolymerization, uptake and physiology specifically (Table 4.1). We focused our analyses on these parameters as they are some of the most challenging parameters to measure directly, in contrast to soil biophysical parameters (e.g. bulk density, stoichiometries of soil organic matter, litter and microbial biomass) which are routinely measured. These parameters are associated with some of the most recent advancements in soil biogeochemical modeling, as they are related to the direct representation of enzyme-mediated decomposition and microbial processing (Manzoni and Porporato 2009, Allison et al. 2010, Tang 2015).

Efforts to constrain model parameter uncertainty should prioritize parameters that have the largest impacts on model outputs of interest. In DAMM-MCNP, model outputs were disproportionately sensitive to the activation energies of depolymerization and uptake (Ea_{dep} , Ea_{upt}), while other parameters, including the half-saturation constant of depolymerization (Km_{dep}) showed minimal impact on C and N cycling (Figure 4.2). Holding temperature and moisture constant, varying Km_{dep} by $\pm 50\%$ of its default value only alters estimated rates of depolymerization by 3%. Furthermore, depolymerization acts only on a small fraction of the total SOCN pools, based on the available fraction of unprotected soil organic matter and diffusion dictated by soil moisture (only 0.0015% of the total SOCN pool is available for decomposition at mean soil moisture). Therefore, the large stocks of total SOCN are essentially insensitive to this parameter. In contrast, varying Ea_{dep} by $\pm 50\%$ of its default value leads to variation in depolymerization rates over several

orders of magnitude, allowing this parameter to have detectable impacts on SOCN pool sizes.

In DAMM-MCNP, activation energies ($E_{a_{dep}}$, $E_{a_{upt}}$) are used to calculate the maximum reaction rates (V_{max}) for depolymerization and uptake, which are sensitive to temperature. In the context of global change, it is particularly important to constrain these enzyme kinetic parameters as they directly impact estimates of how SOCN stocks respond to warming. As V_{max} increases exponentially in response to temperature, models parameterized according to this formulation predict positive feedbacks to warming. However, enzyme kinetic parameters, such as V_{max} , may also show local adaptation to temperature, with greater temperature sensitivities in cooler climates (German et al. 2012, Allison et al. 2018). This potential for local, thermal adaptation of enzyme kinetics can reduce predictions of SOCN losses from soil in response to warming. Collectively, these observations demonstrate the critical importance of constraining enzyme kinetic parameters for understanding SOCN sequestration and predicting its response to global change.

Can the assimilation of existing data reduce model parameter uncertainty, and what is the impact of reducing parameter uncertainty on model estimates?

In addition to analyzing the sensitivity of model outputs to variation in parameter settings, it is important to also pair these analyses with an assessment of parameter uncertainty. Focusing solely on parameter sensitivity can be misleading as highly-sensitive,

but tightly-constrained parameters can potentially be smaller sources of uncertainty to model estimates than less-sensitive, but poorly-constrained parameters (Dietze et al. 2014).

Assimilating field data on heterotrophic respiration data was primarily useful for constraining three parameters (Ea_{dep} , Ea_{upt} , CUE), while the remaining parameters were not reliably identifiable (Figure 4.3). These parameters also had some of the largest impacts on modeled estimates of C and N. Thus, although respiration data alone could not constrain all parameters, it was associated with uncertainty reductions in some of the most critical parameters to constrain. However, parameter estimates identified through data assimilation were sometimes very different from default parameter settings currently used in the published model. Even small adjustments to these parameter choices, justified by reductions in uncertainty, could have major consequences for model projections (Figure 4.7) and their associated uncertainties (Figure 4.8, Figure 4.9). For example, assimilation of the respiration data supported a small reduction in Ea_{dep} by 4%. This minimal change in a single parameter resulted in increased rates of depolymerization, which resulted in sizable depletions of SOCN stocks over time (Figure 4.7A).

What specific types of data have the potential to reduce model parameter uncertainty most effectively?

As respiration data alone could not be used to constrain all model parameters, we explored which potential future data sources would be most useful for reducing parameter uncertainty. We found that identifying changes in DOCN pool sizes could reduce parameter uncertainty for most parameters (Figure 4.6D), while measuring SOCN pool

sizes had limited utility (Figure 4.6C). DOCN pool sizes were also sensitive to variation in most parameters, in contrast to SOCN pool sizes which were relatively stable across parameter variation (Figure 4.2).

DOCN pools are produced through depolymerization and consumed through uptake, placing them centrally in the model and making them directly responsive to both depolymerization and uptake parameters. Additionally, microbial physiological parameters indirectly impact DOCN pool sizes as the size of the microbial biomass pool impacts rates of uptake. Thus, due to their high connectedness to other pools and parameters in the model, DOCN data showed the greatest potential as a single source for constraining multiple parameters. While soil decomposition models differ in their specific representation of C and N pools, our analysis suggests that dynamic pools and fluxes that are located more centrally within model structures are most likely to constrain parameters and are a high priority for data collection.

Classical decomposition models represent conceptual C pools with constant decay rates, which can be difficult to measure directly, in contrast to a growing trend of representing measurable pools such as DOCN (Manzoni and Porporato 2009, Abramoff et al. 2018). DOCN is composed of amino acids, peptides and other compounds which can be rapidly consumed by soil microorganisms as both C and N sources (Finzi and Berthrong 2005, Farrell et al. 2011, 2014, Warren 2014). A variety of methods exist for measuring this pool in the field, including collection the use of lysimeters, microdialysis, and aqueous extraction (McDowell and Likens 1988, Currie et al. 1996, Warren 2014) making it

possible to pair measurements of DOCN along with respiration measurements for constraining model parameters in the future.

Field observations in forest ecosystems indicate that DOCN pools are highly dynamic, with rapid turnover times, sometimes at hourly timescales (Bengtson and Bengtsson 2007). Additionally, these pools are highly sensitive to factors likely to shift under global change. Experimental treatments of elevated CO₂ and soil warming both suggest that DOCN availabilities may increase in the future, with feedbacks to more stable SOCN pool sizes (Toosi et al. 2014, Fang et al. 2015). Thus, accurately representing DOCN dynamics will be critical for modeling ecosystem responses to global change.

Implications for modeling soil biogeochemistry

Several recent soil biogeochemical models share similar structures to DAMM-MCNiP, featuring multiple distinct soil C pools with transfers among pools mediated by microbially-produced enzymes according to Michaelis-Menten type kinetics (Manzoni and Porporato 2009). DAMM-MCNiP is unique among these models in that it combines the effects of temperature, soil moisture and N on C cycling through the SOCN-microbial system. Parameter estimation efforts in biogeochemical modeling have largely focused on aboveground processes, plant-related parameters, or simple decay rate constants for decomposition, while similar approaches in microbe-explicit coupled C-N soil biogeochemical models are lagging (Luo et al. 2009). The assimilation of simulated data has previously been used to assess the identifiability of decay rate parameters in the two-pool Introductory Carbon Balance Model (ICBM; Luo et al. 2017) and initial pool sizes in

the Rothamsted C model (Scharnagl et al. 2010). Both the ICBM and Rothhamsted C model lack a mechanistic representation of microbial processes, coupled C and N cycling, and enzyme-mediated depolymerization, making it necessary to extend these approaches to explore parameters in more recent soil biogeochemical models.

Current microbe-explicit coupled C-N soil biogeochemical models predict widely divergent model projections in response to global change due to differences in structure and parameterization (Sulman et al. 2018, Wieder et al. 2018). A comparison of recent microbial biogeochemical models found that confronting conflicting models with a large synthesis of available data from field experiments on SOC responses to global change was unable to identify which models are most representative, and also highlighted a consistent failure among models to predict frequent empirical observations of increased SOC accumulation under warming (Sulman et al. 2018). These challenges indicate a need for more closely integrating data collection efforts with model development.

Efforts on improving model representations of C and N cycling tend to focus on increasing model structural complexity to include more realistic representations of microbial physiology and enzyme-mediated decomposition. These advancements have been critical for understanding the direct controls of enzyme activities on soil organic matter storage and depolymerization and microbial contributions to C and N cycling; however, it is equally important to consider model uncertainties associated with parameterization. Improving our ability to model the interactions of soil microbial physiology, soil chemistry, enzyme activities, and environmental factors on C and N

cycling will require closely considering model uncertainties and integrating future data collection with model needs.

Tables and Figures

Parameter	Units	Default Value	Description
E_{dep}	kJ mol^{-1}	61.77	Activation energy for SOCN depolymerization
K_{m_{dep}}	mg cm^{-3}	0.0025	Half-saturation constant for SOCN depolymerization
a_{dep}	$\text{mg SOCN cm}^{-3} (\text{mg Enz cm}^{-3})^{-1} \text{ h}^{-1}$	1.0815×10^{11}	Pre-exponential constant for SOCN depolymerization
E_{upt}	kJ mol^{-1}	61.77	Activation energy for DOC uptake
a_{upt}	$\text{mg DOCN cm}^{-3} (\text{mg biomass cm}^{-3})^{-1} \text{ h}^{-1}$	1.0815×10^{11}	Pre-exponential constant for DOCN uptake
K_{m_{upt}}	mg cm^{-3}	0.3	Half-saturation constant for DOCN uptake
K_{m_{O2}}	$\text{cm}^3 \text{ O}_2 \text{ cm}^{-3} \text{ air}$	0.121	Michaelis constant for O_2
CUE	mg mg^{-1}	0.31	Carbon use efficiency
p	-	0.5	proportion of assimilated C allocated to enzyme production
q	-	0.5	proportion of assimilated N allocated to enzyme production
CN_s	-	27.6	C:N of soil
CN_l	-	27.6	C:N of litter
CN_m	-	10	C:N of microbial biomass
CN_e	-	3	C:N of enzymes
BD	g cm^{-3}	0.8	bulk density
PD	g cm^{-3}	2.52	particle density

r_death	hr ⁻¹	0.00015	microbial turnover rate
r_EnzLoss	hr ⁻¹	0.001	enzyme turnover rate
MICtoSOCN	mg mg ⁻¹	0.5	fraction of dead microbial biomass allocated to SOCN
a	-	0.5	proportion of enzyme pool acting on SOC pool (1-a = proportion acting on SON pool)
frac	g C cm ⁻³ / g C cm ⁻³	0.000414	fraction of unprotected SOCN, using soluble substrate estimated from Magill et al., 2000
sat	cm ³ H ₂ O cm ⁻³ soil	1	Moisture saturation level
O₂airfrac	L O ₂ / L air	0.209	volume fraction of O ₂ air
D_{liq}	-	3.17	diffusion coefficient for unprotected SOCN and DOCN in liquid
D_{gas}	-	1.67	diffusion coefficient for O ₂ in air

Table 4.1 — DAMM-MCNIp parameters, units, default values and definitions.

Shaded rows show parameters not estimated in the present analysis.

PARAMETER	POSTERIOR 2.50%	POSTERIOR MEDIAN	POSTERIOR 97.5%	PRIOR LOWER	PRIOR UPPER
<i>A_{DEP}</i>	5.52×10^{10}	9.53×10^{10}	1.58×10^{11}	5.40×10^{10}	1.62×10^{11}
<i>A_{UPT}</i>	6.07×10^{10}	1.20×10^{11}	1.60×10^{11}	5.40×10^{10}	1.62×10^{11}
CUE	3.21E-01	0.35	0.40	0.16	0.47
<i>EA_{DEP}</i>	59.3	60.6	61.9	30.9	92.7
<i>EA_{UPT}</i>	32.8	69.6	72.0	30.9	92.7
<i>KM_{DEP}</i>	1.00×10^{-3}	2.00×10^{-3}	3.00×10^{-3}	1.25×10^{-3}	3.75×10^{-3}
<i>KM_{O2}</i>	6.30×10^{-2}	0.11	0.18	6.05×10^{-2}	0.18
<i>KM_{UPT}</i>	0.15	0.23	0.42	0.15	0.45
P	0.30	0.63	0.74	0.25	0.75
Q	0.53	0.70	0.75	0.25	0.75

Table 4.2 — Parameter estimates given field observations of heterotrophic respiration

Parameter estimates for all ten parameters estimated together using field observations of heterotrophic respiration (100000 iteration MCMC with 3 chains and burn-in of 1000).

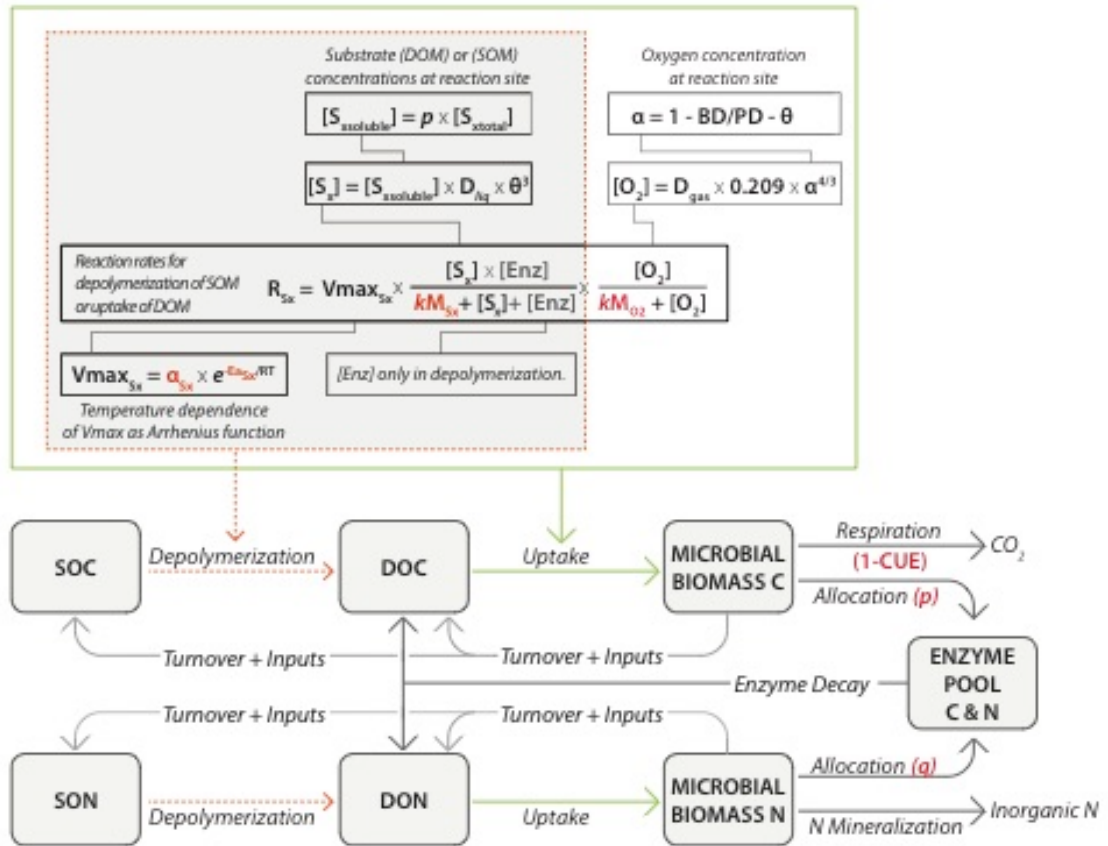


Figure 4.1 — Major pools, fluxes and equations in DAMM-MCNiP.

Dashed orange lines enclose equations used for depolymerization kinetics. Solid green lines enclose equations for uptake of DOCN (with the exception of $[\text{Enz}]$, which only appears in depolymerization kinetics). Microbial parameters are shown above associated arrows. Parameters estimated in the present analysis are highlighted in red.

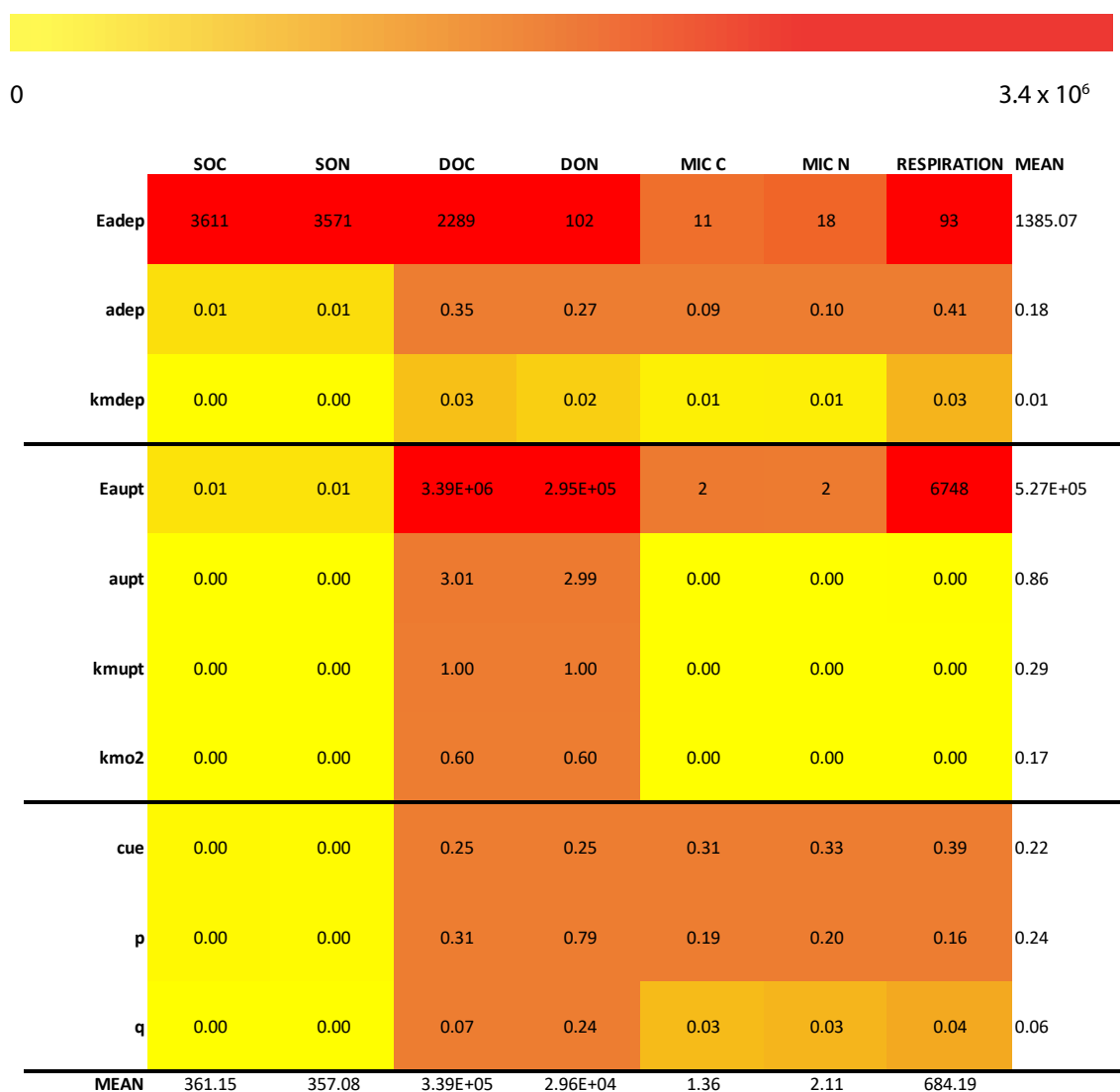


Figure 4.2 — Heatmap showing sensitivity of model outputs to parameters.

Sensitivities are calculated as proportional change in output (columns) relative to proportional change in parameter (rows; equation 4.6). Final row shows mean sensitivity of model output across the ten parameters of interest. Final column shows mean sensitivity across model outputs for each parameter individually.

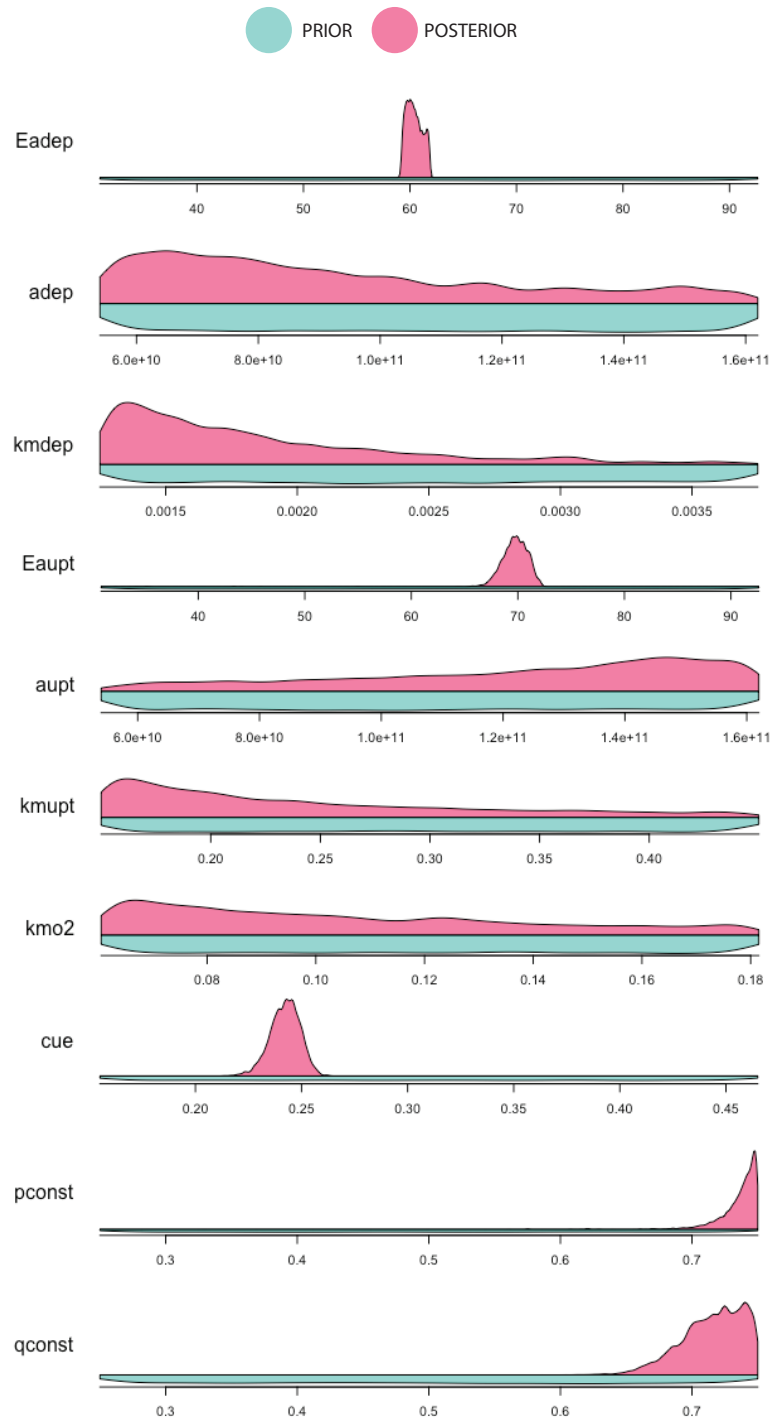


Figure 4.3 — Marginal parameter uncertainties following incorporation of observed respiration data.

Prior distributions are uniform spanning the full range of x-axes.

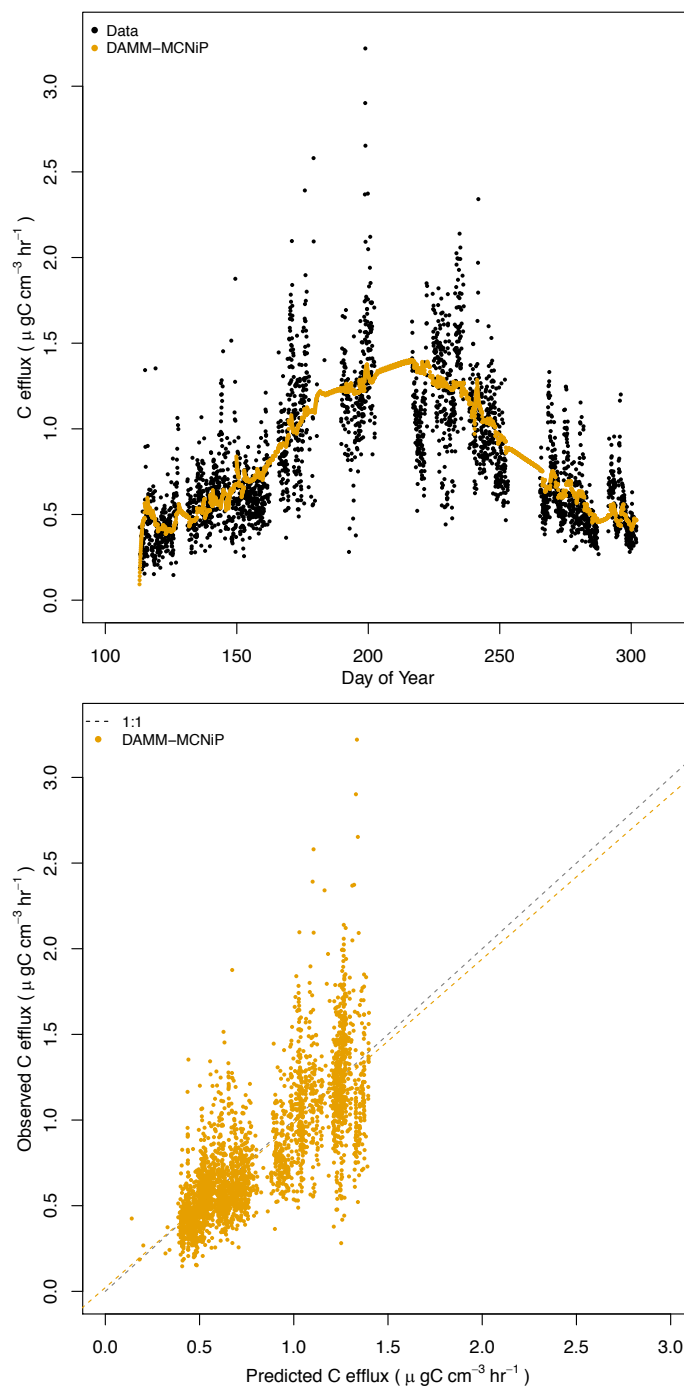
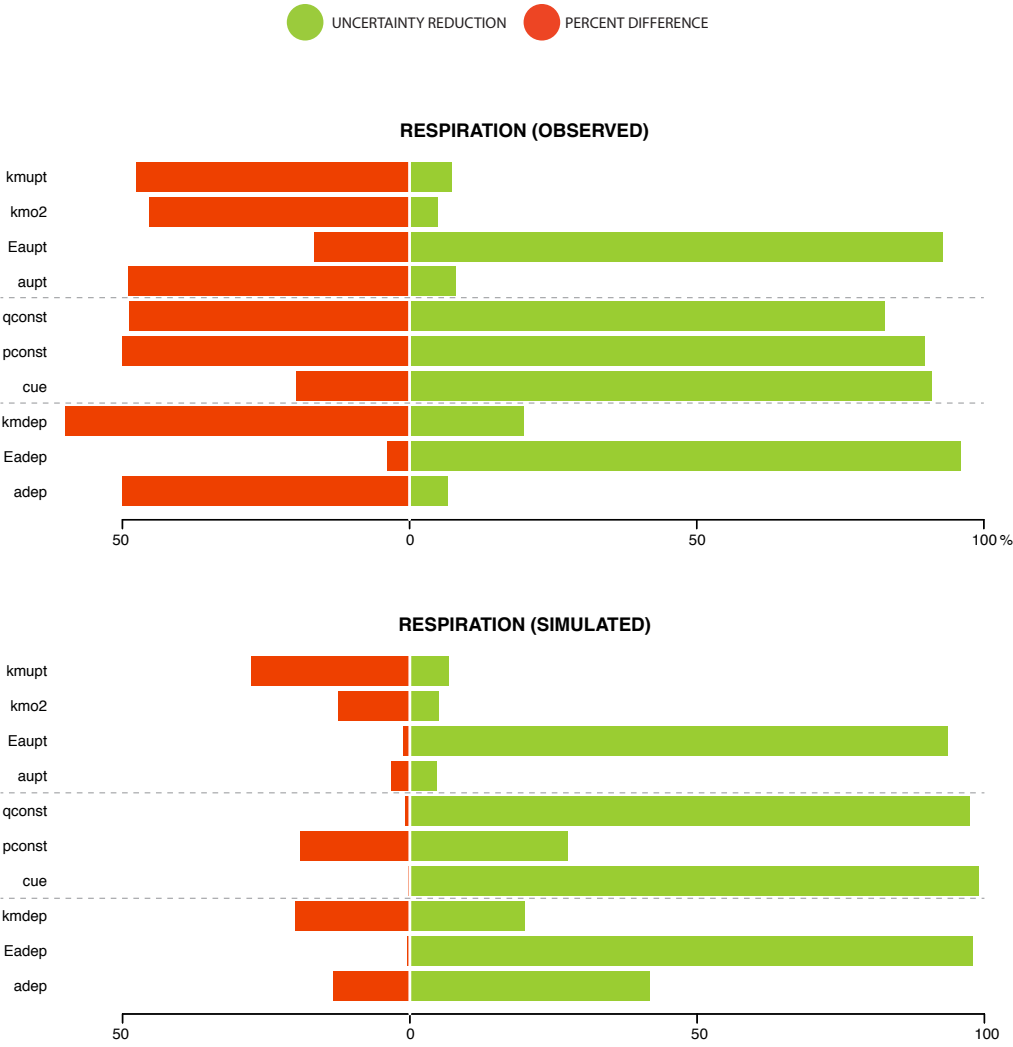


Figure 4.4 — Model predictions compared to heterotrophic respiration data.

(a) Time series of observed (black) and predicted (yellow) heterotrophic respiration rates and (b) linear regression of observed and predicted heterotrophic respiration rates using median parameter values from data assimilation (Table 4.2; Slope = 0.96, $R^2_{\text{adj}}=0.61$, RMSE = 0.24).



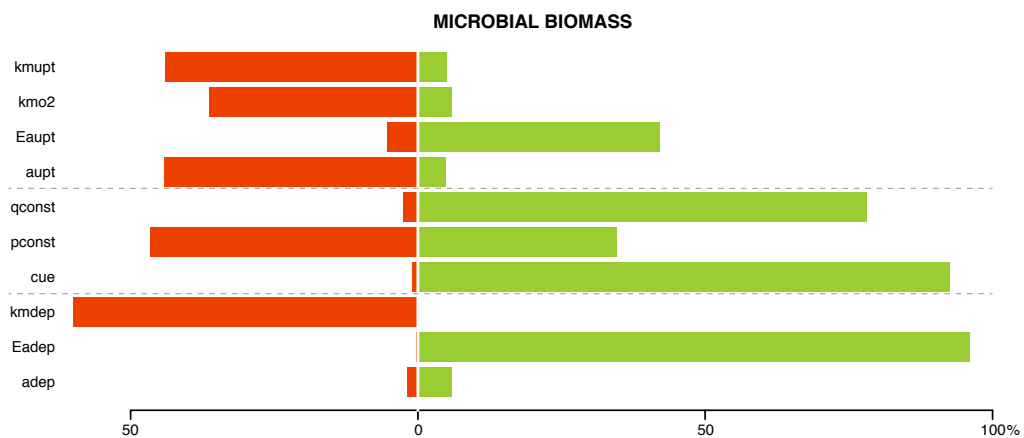
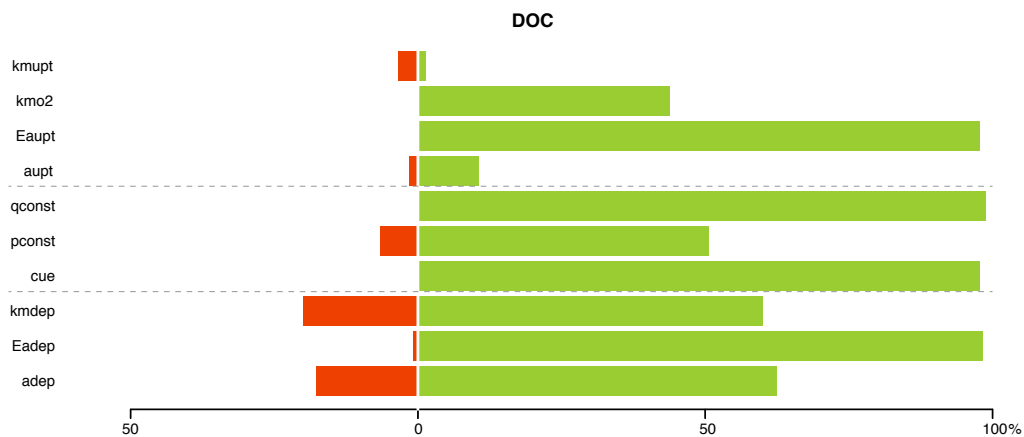
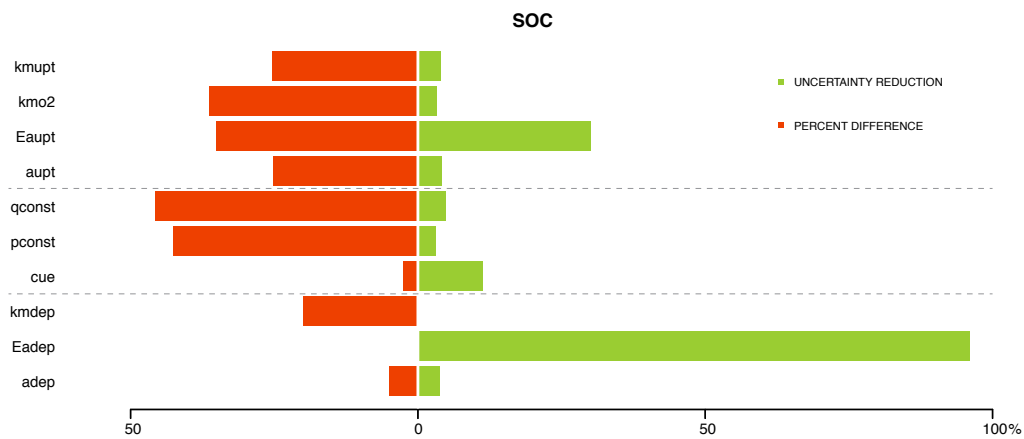


Figure 4.5 — Uncertainty reductions and accuracies for parameters estimated provided different types of data.

Uncertainty reduction (green) and percent difference between *maximum a posteriori* (MAP) estimate and default parameter values (red) for parameters when provided with (A) observed data for respiration, or simulated data for (B) respiration, (C) SOC and SON, (D) DOC and DON, (E) Microbial Biomass C and N. Dashed lines separate between parameter types for uptake, physiology, and depolymerization. Uncertainty reductions approaching 100% indicate narrowing of posterior distribution relative to prior distribution based on data provided. Percent differences approaching 0% indicate MAP values close to parameter setting used in simulation.

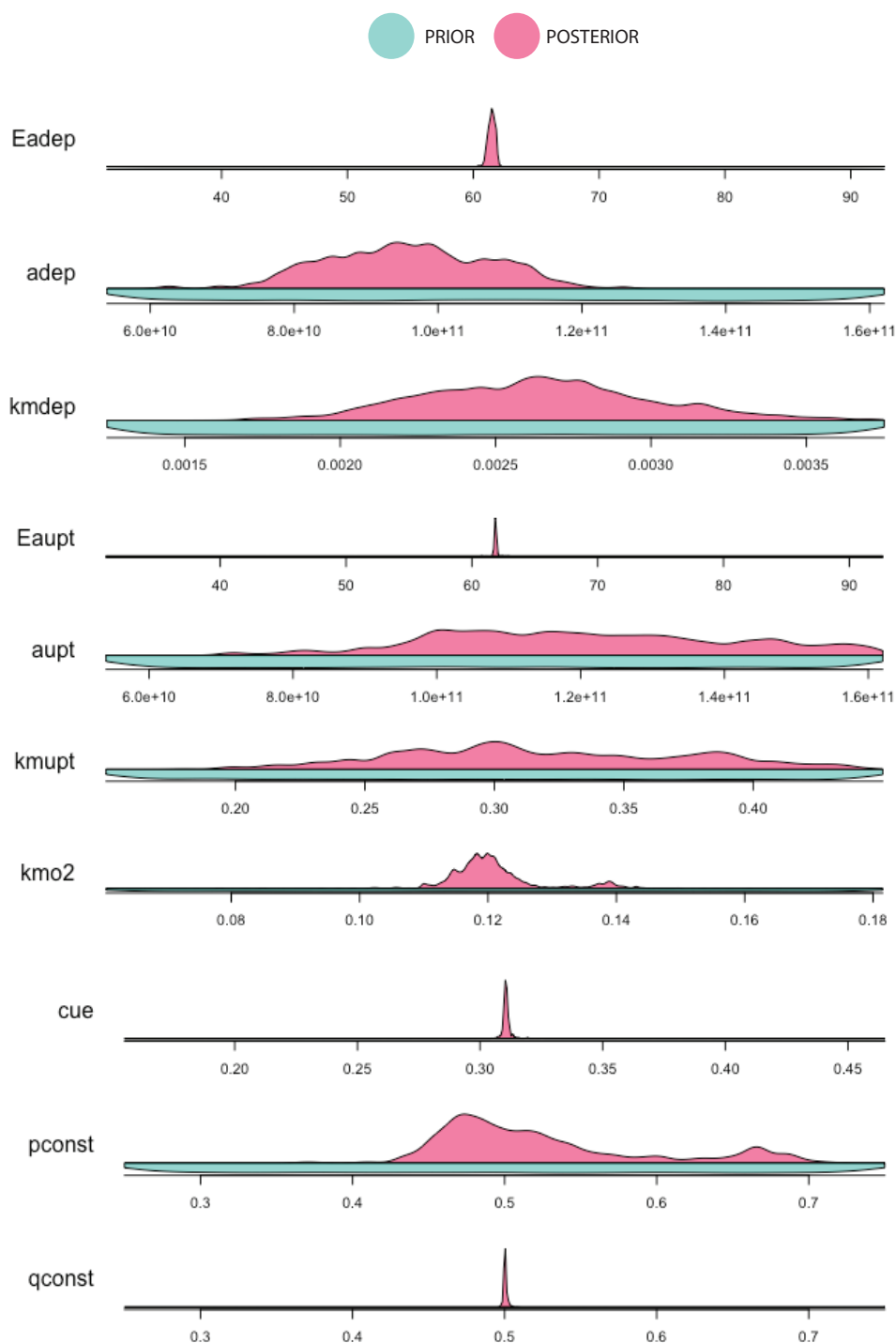


Figure 4.6 — Marginal parameter uncertainties following incorporation of simulated DOC data.

Prior distributions are uniform spanning the full range of x-axes.

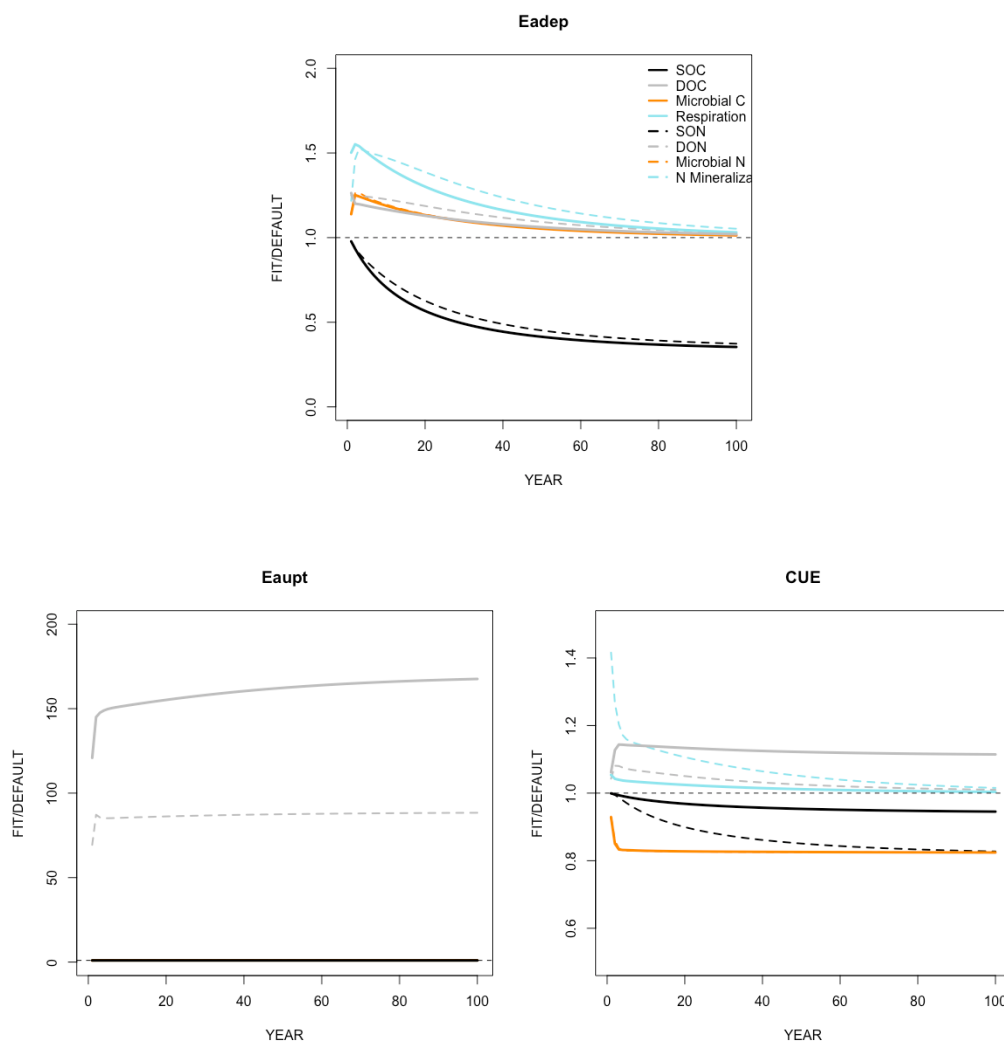


Figure 4.7. – MODELED OUTPUTS USING MAP PARAMETER ESTIMATES.

Ratio of model outputs over 100-year projections using MAP values for (A) Ea_{dep} , (B) Ea_{upt} and (C) CUE following incorporation of observed respiration data relative to model outputs based on default parameter value.

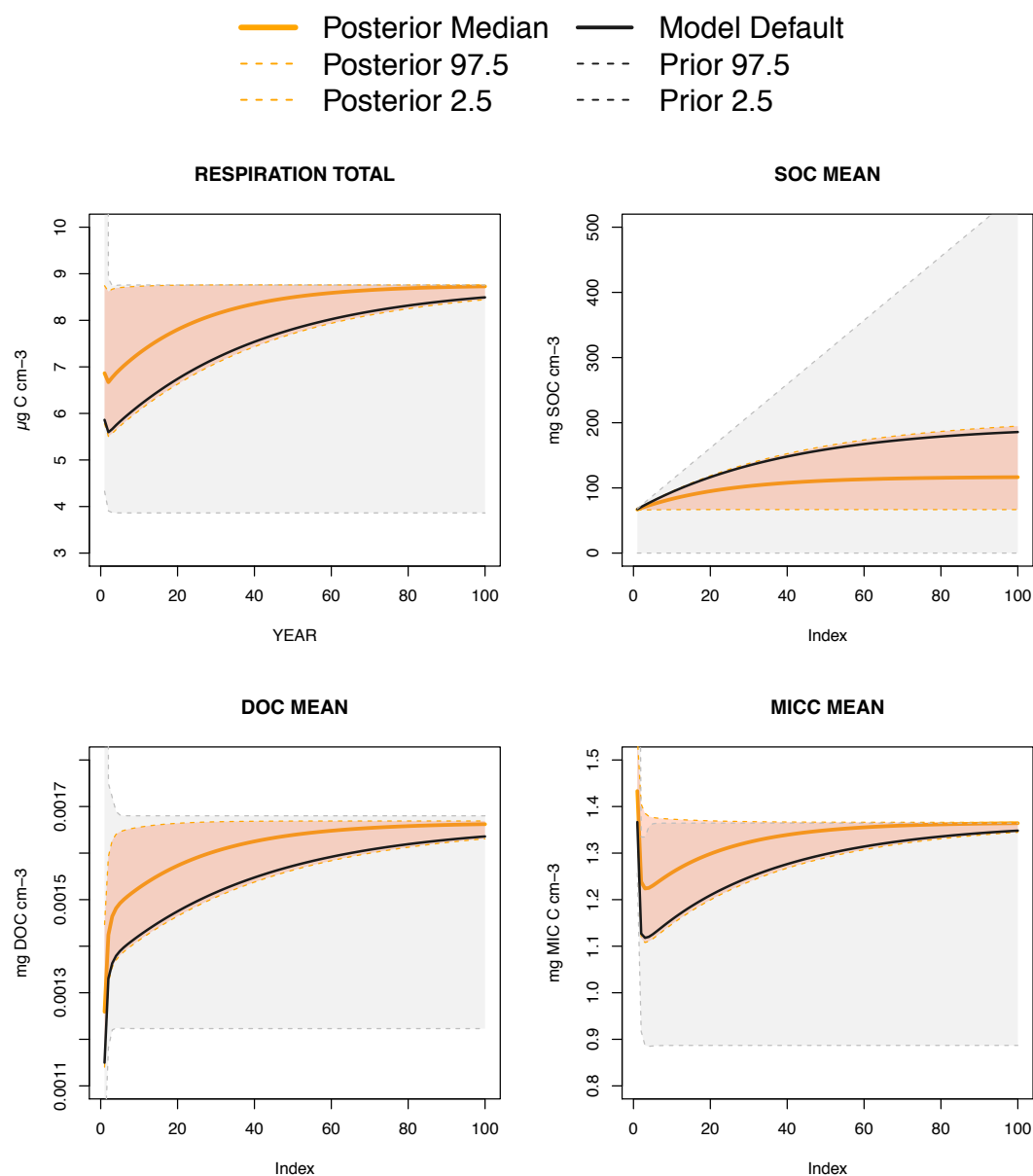


Figure 4.8 – MODELED C OUTPUTS WITH PRIOR AND POSTERIOR ESTIMATES OF Ea_{dep} .

Model estimates of (a) annual total respiration, (b) mean SOC, (c) mean DOC, and (d) mean microbial biomass N based on posterior median parameter estimates for Ea_{dep} (solid orange) and 2.5 – 97.5% credible interval for posterior parameter estimate (orange interval). Solid black line shows model estimate based on default parameter setting. Gray interval shows range of estimates using prior parameter distribution.

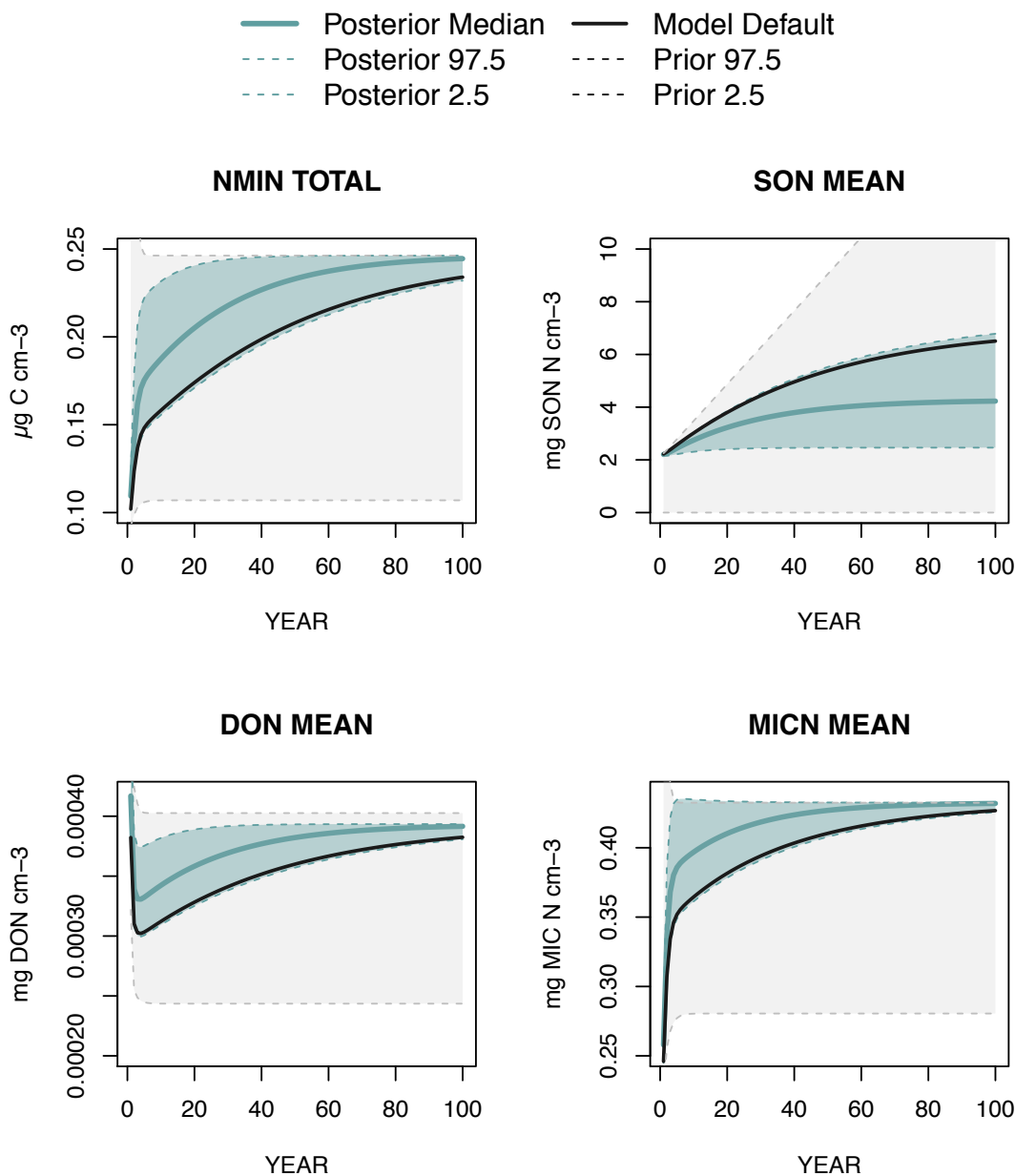


Figure 4.9 – MODELED N OUTPUTS WITH PRIOR AND POSTERIOR ESTIMATES OF Ea_{dep}

Model estimates of (a) annual total N mineralization, (b) mean SON, (c) mean DON, and (d) mean microbial biomass N based on posterior median parameter estimates for Ea_{dep} (solid blue) and 2.5 – 97.5% credible interval for posterior parameter estimate (blue interval). Solid black line shows model estimate based on default parameter setting. Gray interval shows range of estimates using prior parameter distribution.

CHAPTER FIVE: CONCLUSIONS

Advancing our understanding of soil C and N cycling requires bridging across vastly different scales to identify how the physiology of individual microbial cells impacts biogeochemical cycling at the ecosystem level. This dissertation seeks to characterize relationships between soil microbial physiology, microbial community structure and the biogeochemical cycling of C and N, by integrating a range of techniques from genome-scale metabolic modeling, quantitative PCR, stable isotope-based flux measurements, and ecosystem modeling. Through the combination of these approaches, I find that variation in microbial physiology can have quantitatively important implications for C cycling in soil, microbial community interactions can have significant implications for N cycling in temperate forest ecosystems, and that our ability to model coupled soil C and N cycling is highly sensitive to representations of microbial physiology and the kinetics of extracellular enzymes produced by microbes.

In this dissertation, I first characterized variation in C metabolism across bacterial taxa. The approach I used was intentionally informed by model representations of microbial metabolism making it easy to directly integrate into soil biogeochemical models (Manzoni et al. 2012). While respiration is a ubiquitous metabolic pathway across microbial taxa, physiological variation in C metabolism can have important ramifications for C cycling (Allison et al. 2010, Frey et al. 2013, Waring et al. 2013, Wieder et al. 2014). Previous studies have primarily focused on the impact of abiotic factors such as temperature and substrate chemistry on bacterial C metabolism, while less is known about variation across bacterial taxa. Using a novel approach to predict bacterial C-use efficiency

(CUE) for over 200 species with genome-specific constraint-based metabolic modeling, I found that accounting for variation in C metabolism across bacterial taxa had large impacts on ecosystem-level estimates of C cycling. Intrinsic differences between taxa alone resulted in >300% variation in CUE (0.2-0.9), indicating important functional differences across bacterial taxa not accounted for in current modeled representations of microbial C cycling.

Characterizing the phylogenetic organization of bacterial traits is one approach that may be helpful for classifying sources of variation across highly diverse microbial taxa (Martiny et al. 2013). I found that CUE has a significant phylogenetic signal, with variation among taxa structured primarily at the class and order levels. These findings contribute to a growing discussion on identifying the appropriate level of phylogenetic resolution required to understand biogeochemically-relevant variation in microbial traits (Fierer et al. 2007, Martiny et al. 2013). Future studies will benefit from empirically measuring taxa-specific bacterial CUE using consistent methodologies. Exciting new work utilizing quantitative stable isotope probing show promise for helping understand the phylogenetic structure of bacterial activities (Morrissey et al. 2016, 2017).

I was also able to relate variation in CUE to specific genomic traits, such as GC content and genome size, providing a potential method for predicting CUE from genomic information in the future. The negative correlation between genome size and potential CUE observed in my analysis could reflect an interesting tradeoff between efficiency and access to diverse substrate types, as taxa with larger genomes were able to access a wide range of C sources at the expense of efficiency compared to taxa with smaller genomes. These

findings build on work to link genomic traits with bacterial metabolism, which have provided useful hypotheses on how CUE may be predicted from compositional data (Roller et al. 2016).

When I incorporated the full range of CUE values generated from my analysis into a next-generation model of soil biogeochemistry, I found that differences among taxa significantly alter estimates of soil-C storage and microbial biomass over 100 year projections. These large differences in ecosystem-level C cycling driven solely by phylogenetic variation in CUE indicate that an improved understanding of microbial physiology is critical to accurate representations of C cycling.

At the microbial community level, there remain many unanswered questions on relationships between community structure, microbial interactions and ecosystem function (Bier et al. 2015). I used the N cycle as a case study to explore these questions as soil microbial groups play distinct roles in controlling the availability and form of N in soil. However, factors regulating the abundances and activities of these functional groups in temperate forest ecosystems have been challenging to consistently describe (Wallenstein and Vilgalys 2005, Rocca et al. 2015). I hypothesized that interactions among microbial groups and their access to inorganic N might structure communities and their associated fluxes. In particular, I was interested in how competition between free-living soil bacteria and ectomycorrhizal (ECM) fungi might impact the structure and activity of N cycling bacterial groups.

I explored how rates of these N cycle fluxes varied across gradients of mycorrhizal abundance at four temperate forest sites in Massachusetts and Indiana, USA. I paired

measurements of N-fixation, net nitrification and denitrification rates with gene abundance data for specific bacterial functional groups associated with each process, and I found that the availability of NO_3^- and rates of N-fixation, net nitrification and denitrification all declined with increasing abundances of plants associated with ECM fungi. Despite the strong structuring of activity in relation to mycorrhizal abundance, I did not find consistent patterns in the abundances of functional groups across the gradients, and gene abundances were not correlated with process rates.

Overall, these findings indicated that ECM-associated vegetation was correlated with suppressed rates of N-fixation, net nitrification and denitrification, but I was unable to detect differences in the structure of the bacterial community. Thus, inter-domain community interactions had a dramatic impact on ecosystem function, while the structure of the bacterial community itself was not clearly linked to N-cycling rates. There are several challenges associated with linking community structure with ecosystem function, including the high diversity of soil microbial communities, potential functional redundancy across taxa, dormancy, and abiotic controls on activity (Bier et al. 2015). Future work will benefit from continuing to couple measurements of N cycling microbial abundances and associated fluxes to help identify the mechanisms by which microbial functional group abundances translate into N cycling rates.

The biogeochemical cycles of C and N are closely coupled, making it important to consider both in tandem. Ecosystem models are useful tools for representing coupled C and N cycling and can provide useful insights on how complex interactions among soil chemistry, microbial processes and abiotic factors impact biogeochemical functioning. Soil

biogeochemical models have increasingly incorporated more accurate representations of microbial physiology and enzyme-mediated depolymerization of soil organic matter (Manzoni and Porporato 2009, Finzi et al. 2015, Abramoff et al. 2017, 2018). A major challenge with these model structural improvements involves the requirement for additional parameters, which are often poorly constrained sources of uncertainty.

I used DAMM-MCNiP, a microbially-mediated, coupled soil C and N cycling model, as a tool to explore the influence of microbial physiological and enzyme kinetic parameters on model estimates of coupled soil C and N cycling. I quantified the potential for constraining model parameters using empirical measurements of soil respiration, and I used simulated data to assess the potential for future data collection to constrain particular parameters. This approach represents a useful framework for designing future studies to directly improve modeling efforts. I found that modeled soil C and N pools and fluxes are disproportionately sensitive to only a few parameters (e.g. activation energies and CUE), while others exert minimal influence (e.g. Michaelis-Menten half-saturation constants). While some parameters could be constrained by the available data on heterotrophic respiration, the collection of additional data on dissolved organic matter was identified as a potentially useful constraint on most parameters. Improving model representations of microbially-mediated soil C and N cycling will require closer consideration of model uncertainties and targeting data collection towards reducing these uncertainties.

In the context of climate change, which may alter the availability and distribution of resources as well as abiotic controls on microbial activity, it will be critical to characterize microbial functions and their roles in global C and N cycling. Collectively, the

work in this dissertation builds toward an improved understanding of the linkages between microbial physiology, community structure and biogeochemical cycling. Continued integrative work bridging approaches from the scale of genomes to greenhouse gases fluxes will be necessary to advance our understanding of the factors regulating microbial abundances, microbial activities and soil C and N cycling.

BIBLIOGRAPHY

- Abramoff, R., X. Xu, M. Hartman, S. O'Brien, W. Feng, E. Davidson, A. Finzi, D. Moorhead, J. Schimel, M. Torn, and M. A. Mayes. 2018. The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry*, 137:51–71.
- Abramoff, R. Z., E. A. Davidson, and A. C. Finzi. 2017. A parsimonious modular approach to building a mechanistic belowground carbon and nitrogen model. *Journal of Geophysical Research: Biogeosciences*, 122: 2418-2434.
- Ajami, N. K., and C. Gu. 2010. Complexity in microbial metabolic processes in soil nitrogen modeling: A case for model averaging. *Stochastic Environmental Research and Risk Assessment*, 24:831–844.
- Allison, S. D., and J. B. H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences*, 105: 11512-11519.
- Allison, S. D., A. L. Romero-Olivares, Y. Lu, J. W. Taylor, and K. K. Treseder. 2018. Temperature sensitivities of extracellular enzyme V_{max} and K_m across thermal environments. *Global Change Biology*, 24:2884-2897.
- Allison, S. D., M. D. Wallenstein, and M. A. Bradford. 2010. Soil-carbon response to warming dependent on microbial physiology. *Nature Geoscience*, 3:336–340.

- Arkin, A. P., R. W. Cottingham, C. S. Henry, N. L. Harris, R. L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M. W. Sneddon, M. L. Henderson, W. J. Riehl, D. Murphy-Olson, S. Y. Chan, R. T. Kamimura, S. Kumari, M. M. Drake, T. S. Brettin, E. M. Glass, D. Chivian, D. Gunter, D. J. Weston, B. H. Allen, J. Baumohl, A. A. Best, B. Bowen, S. E. Brenner, C. C. Bun, J. M. Chandonia, J. M. Chia, R. Colasanti, N. Conrad, J. J. Davis, B. H. Davison, M. Dejongh, S. Devoid, E. Dietrich, I. Dubchak, J. N. Edirisinghe, G. Fang, J. P. Faria, P. M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H. L. Haun, F. He, R. Jain, M. P. Joachimiak, K. P. Keegan, S. Kondo, V. Kumar, M. L. Land, F. Meyer, M. Mills, P. S. Novichkov, T. Oh, G. J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S. S. Poon, G. A. Price, S. Ramakrishnan, P. Ranjan, P. C. Ronald, M. C. Schatz, S. M. D. Seaver, M. Shukla, R. A. Sutormin, M. H. Syed, J. Thomason, N. L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, and D. Yu. 2018. KBase: The United States department of energy systems biology knowledgebase.
- Averill, C., M. C. Dietze, and J. M. Bhatnagar. 2018. Continental-scale nitrogen pollution is shifting forest mycorrhizal associations and soil carbon stocks. *Global Change Biology*, 24:4544-4553.
- Averill, C., and C. V. Hawkes. 2016. Ectomycorrhizal fungi slow soil carbon cycling. *Ecology Letters*, 19:937-947.
- Averill, C., B. L. Turner, and A. C. Finzi. 2014. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature*, 505:543–5.

Barberán, A., K. S. Ramirez, J. W. Leff, M. A. Bradford, D. H. Wall, and N. Fierer.

2014. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecology Letters*, 17:794-802.

Barnard, R., P. W. Leadley, and B. A. Hungate. 2005. Global change, nitrification, and denitrification: A review. *Global Biogeochemical Cycles*, 19:GB1007.

Bartón, K. 2018. MuMIn: Multi-Model Inference. R Package Version 0.12.2/r18.

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

Bengtson, P., and G. Bengtsson. 2007. Rapid turnover of DOC in temperate forests accounts for increased CO₂ production at elevated temperatures. *Ecology Letters*, 10:783-790.

Bernard, L., C. Mougél, P. A. Maron, V. Nowak, J. Lévêque, C. Henault, F. E. Z.

Haichar, O. Berge, C. Marol, J. Balesdent, F. Gibiat, P. Lemanceau, and L. Ranjard. 2007. Dynamics and identification of soil microbial populations actively assimilating carbon from ¹³C-labelled wheat residue as estimated by DNA- and RNA-SIP techniques. *Environmental Microbiology*, 9:752-764.

Berthrong, S. T., C. M. Yeager, L. Gallegos-Graves, B. Steven, S. A. Eichorst, R. B.

Jackson, and C. R. Kuske. 2014. Nitrogen fertilization has a stronger effect on soil nitrogen-fixing bacterial communities than elevated atmospheric CO₂. *Applied and*

Environmental Microbiology, 80:3103-3112.

Bier, R. L., E. S. Bernhardt, C. M. Boot, E. B. Graham, E. K. Hall, J. T. Lennon, D. R. Nemergut, B. B. Osborne, C. Ruiz-González, J. P. Schimel, M. P. Waldrop, and M. D. Wallenstein. 2015. Linking microbial community structure and microbial processes: An empirical and conceptual overview. *FEMS Microbiology Ecology*, 91:fiv113.

Blankinship, J. C., A. A. Berhe, S. E. Crow, J. L. Druhan, K. A. Heckman, M. Keiluweit, C. R. Lawrence, E. Marín-Spiotta, A. F. Plante, C. Rasmussen, C. Schädel, J. P. Schimel, C. A. Sierra, A. Thompson, R. Wagai, and W. R. Wieder. 2018. Improving understanding of soil organic matter dynamics by triangulating theories, measurements, and models. *Biogeochemistry*, 140:1-13.

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

Bölscher, T., L. Wadsö, G. Börjesson, and A. M. Herrmann. 2016. Differences in substrate use efficiency: impacts of microbial community composition, land use management, and substrate complexity. *Biology and Fertility of Soils*, 52: 547-559.

Bond-Lamberty, B., and A. Thomson. 2010. Temperature-associated increases in the global soil respiration record. *Nature*, 464:579–582.

Bond-Lamberty, B., C. Wang, and S. T. Gower. 2004. A global relationship between the

- heterotrophic and autotrophic components of soil respiration? *Global Change Biology*, 10:1756-1766.
- Bormann, B. T. 1993. Rapid N^2 fixation in pines, alder, and locust: evidence from the sandbox ecosystem study. *Ecology*, 74:583-598.
- Bormann, B. T., C. K. Keller, D. Wang, and F. H. Bormann. 2002. Lessons from the sandbox: Is unexplained nitrogen real? *Ecosystems*, 5:0727-0733.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and inference: a practical information-theoretic approach. Second edition. Springer, New York.
- Cheeke, T. E., R. P. Phillips, E. R. Brzostek, A. Rosling, J. D. Bever, and P. Fransson. 2017. Dominant mycorrhizal association of trees alters carbon and nutrient cycling by selecting for microbial groups with distinct enzyme function. *New Phytologist*, 10.1111/nph.14343.
- Cornelissen, J., R. Aerts, B. Cerabolini, M. Werger, and M. Van der Heijden. 2001. Carbon cycling traits of plant species are linked with mycorrhizal strategy. *Oecologia*, 129:611-619.
- Craig, M. E., B. L. Turner, C. Liang, K. Clay, D. J. Johnson, and R. P. Phillips. 2018. Tree mycorrhizal type predicts within-site variability in the storage and distribution of soil organic matter. *Global Change Biology*, 24:3317-3330.
- Currie, W. S., J. D. Aber, W. H. McDowell, R. D. Boone, and A. H. Magill. 1996.

Vertical transport of dissolved organic C and N under long-term N amendments in pine and hardwood forests. *Biogeochemistry*, 35:471-505.

Davidson, E. A., S. Samanta, S. S. Caramori, and K. Savage. 2012. The Dual Arrhenius and Michaelis-Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, 18:371-384.

DeAngelis, K. M., G. Pold, B. D. Topçuoglu, L. T. A. van Diepen, R. M. Varney, J. L. Blanchard, J. Melillo, and S. D. Frey. 2015. Long-term forest soil warming alters microbial communities in temperate forest soils. *Frontiers in Microbiology*, 6:104.

Dietze, M. C., S. P. Serbin, C. Davidson, A. R. Desai, X. Feng, R. Kelly, R. Kooper, D. LeBauer, J. Mantooth, K. McHenry, and D. Wang. 2014. A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. *Journal of Geophysical Research: Biogeosciences*, 119:286–300.

Dijkstra, P., E. Salpas, D. Fairbanks, E. B. Miller, S. B. Hagerty, K. J. van Groenigen, B. A. Hungate, J. C. Marks, G. W. Koch, and E. Schwartz. 2015. High carbon use efficiency in soil microbial communities is related to balanced growth, not storage compound synthesis. *Soil Biology and Biochemistry*, 89: 35-43.

Doane, T. A., and W. R. Horwáth. 2003. Spectrophotometric determination of nitrate with a single reagent. *Analytical Letters*, 36:2713-2722.

Drake, J. E., M. A. Giasson, K. J. Spiller, and A. C. Finzi. 2013. Seasonal plasticity in the

temperature sensitivity of microbial activity in three temperate forest soils.

Ecosphere, 4:77.

Edwards, J. S., R. U. Ibarra, and B. O. Palsson. 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19:125-130.

Enanga, E. M., N. J. Casson, T. A. Fairweather, and I. F. Creed. 2017. Nitrous Oxide and Dinitrogen: The Missing Flux in Nitrogen Budgets of Forested Catchments? *Environmental Science and Technology*, 51:6036-6043.

Fang, H., S. Cheng, E. Lin, G. Yu, S. Niu, Y. Wang, M. Xu, X. Dang, L. Li, and L. Wang. 2015. Elevated atmospheric carbon dioxide concentration stimulates soil microbial activity and impacts water-extractable organic carbon in an agricultural soil. *Biogeochemistry*, 122:253-267.

Farrell, M., P. W. Hill, S. D. Wanniarachchi, J. Farrar, R. D. Bardgett, and D. L. Jones. 2011. Rapid peptide metabolism: A major component of soil nitrogen cycling? *Global Biogeochemical Cycles*, 25:GB3014.

Farrell, M., M. Prendergast-Miller, D. L. Jones, P. W. Hill, and L. M. Condron. 2014. Soil microbial organic nitrogen uptake is regulated by carbon availability. *Soil Biology and Biochemistry*, 77:261-267.

Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification

of soil bacteria. *Ecology*, 88:1354-1364.

Fierer, N., J. W. Leff, B. J. Adams, U. N. Nielsen, S. T. Bates, C. L. Lauber, S. Owens, J.

A. Gilbert, D. H. Wall, and J. G. Caporaso. 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, 109:21390-21395.

Finzi, A. C., R. Z. Abramoff, K. S. Spiller, E. R. Brzostek, B. A. Darby, M. A. Kramer, and R. P. Phillips. 2015. Rhizosphere processes are quantitatively important components of terrestrial carbon and nutrient cycles. *Global Change Biology*.

Finzi, A. C., and S. T. Berthrong. 2005. The uptake of amino acids by microbes and trees in three cold-temperate forests. *Ecology*, 21:2082-2094.

Finzi, A. C., N. Van Breemen, and C. D. Canham. 1998. Canopy tree-soil interactions within temperate forests: Species effects on soil carbon and nitrogen. *Ecological Applications*, 8:447-454.

Frey, S. D., R. Drijber, H. Smith, and J. Melillo. 2008. Microbial biomass, functional capacity, and community structure after 12 years of soil warming. *Soil Biology and Biochemistry*, 40:2904–2907.

Frey, S. D., J. Lee, J. M. Melillo, and J. Six. 2013. The temperature response of soil microbial efficiency and its feedback to climate. *Nature Climate Change*, 3:395–398.

- Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti. 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27:511-526.
- Gadgil, R. L., and P. D. Gadgil. 1975. Suppression of litter decomposition by mycorrhizal roots of *Pinus radiata*. *New Zealand Journal of Forestry Science*, 5:33-41.
- Gallet-Budynek, A., E. Brzostek, V. L. Rodgers, J. M. Talbot, S. Hyzy, and A. C. Finzi. 2009. Intact amino acid uptake by northern hardwood and conifer trees. *Oecologia*, 160:129–38.
- Galloway, J. N., F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, C. C. Cleveland, P. A. Green, E. A. Holland, D. M. Karl, A. F. Michaels, J. H. Porter, A. R. Townsend, and C. J. Vörösmarty. 2004. Nitrogen cycles: Past, present, and future. *Biogeochemistry*, 70:153-226.
- Garza, D. R., M. C. Van Verk, M. A. Huynen, and B. E. Dutilh. 2018. Towards predicting the environmental metabolome from metagenomics with a mechanistic model. *Nature Microbiology*, 3:456-460.
- German, D. P., K. R. B. Marcelo, M. M. Stone, and S. D. Allison. 2012. The Michaelis-Menten kinetics of soil extracellular enzymes in response to temperature: A cross-latitudinal study. *Global Change Biology*, 10.1111/j.1365-2486.2011.02615.x.

- Geyer, K. M., E. Kyker-Snowman, A. S. Grandy, and S. D. Frey. 2016. Microbial carbon use efficiency: accounting for population, community, and ecosystem-scale controls over the fate of metabolized organic matter. *Biogeochemistry*, 127:173-188.
- Grafen, A. 1989. The phylogenetic regression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 326:119-157.
- Groffman, P. M., and J. M. Tiedje. 1989. Denitrification in north temperate forest soils: Spatial and temporal patterns at the landscape and seasonal scales. *Soil Biology and Biochemistry*, 21:613-620.
- Harcombe, W. R., W. J. Riehl, I. Dukovski, B. R. Granger, A. Betts, A. H. Lang, G. Bonilla, A. Kar, N. Leiby, P. Mehta, C. J. Marx, and D. Segrè. 2014. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Reports*, 7:1104-1115.
- Hashimoto, S., N. Carvalhais, A. Ito, M. Migliavacca, K. Nishina, and M. Reichstein. 2015. Global spatiotemporal distribution of soil respiration modeled using a global database. *Biogeosciences*, 12:4121-4132.
- Henry, C. S., M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, 28:977-982.
- Houlton, B. Z., and E. Bai. 2009. Imprint of denitrifying bacteria on the global terrestrial

- biosphere. *Proceedings of the National Academy of Sciences*, 106:21713-21716.
- Hsu, S. F., and D. H. Buckley. 2009. Evidence for the functional significance of diazotroph community structure in soil. *ISME Journal*, 3:124-136.
- Hungate, B. A., J. S. Dukes, M. R. Shaw, Y. Luo, and C. B. Field. 2003. Nitrogen and Climate Change, 302: 1512-1513.
- Jobbágy, E. G., and R. B. Jackson. 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications*, 10:423-436.
- Keiblinger, K. M., E. K. Hall, W. Wanek, U. Szukics, I. Hämmerle, G. Ellersdorfer, S. Böck, J. Strauss, K. Sterflinger, A. Richter, and S. Zechmeister-Boltenstern. 2010. The effect of resource quantity and resource stoichiometry on microbial carbon-use-efficiency. *FEMS Microbiology Ecology*, 73:430-440.
- Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26:1463-1464.
- King, Z. A., J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44:D515-522.
- Kögel-Knabner, I. 2017. The macromolecular organic composition of plant and microbial

residues as inputs to soil organic matter: Fourteen years on. *Soil Biology and Biochemistry*, 105:A3-A8.

Konstantinidis, K. T., and J. M. Tiedje. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences*, 101:3160-3165.

Latendresse, M. 2014. Efficiently gap-filling reaction networks. *BMC Bioinformatics*, 15:225.

Lauro, F. M., D. McDougald, T. Thomas, T. J. Williams, S. Egan, S. Rice, M. Z. DeMaere, L. Ting, H. Ertan, J. Johnson, S. Ferriera, A. Lapidus, I. Anderson, N. Kyrpides, A. C. Munk, C. Detter, C. S. Han, M. V. Brown, F. T. Robb, S. Kjelleberg, and R. Cavicchioli. 2009. The genomic basis of trophic strategy in marine bacteria. *Proceedings of the National Academy of Sciences*, 106:15527-15533.

LeBauer, D. S., and K. K. Treseder. 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology*, 89:371-379.

Leff, J. W., S. E. Jones, S. M. Prober, A. Barberán, E. T. Borer, J. L. Firn, W. S. Harpole, S. E. Hobbie, K. S. Hofmockel, J. M. H. Knops, R. L. McCulley, K. La Pierre, A. C. Risch, E. W. Seabloom, M. Schütz, C. Steenbock, C. J. Stevens, and N. Fierer. 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.

Lennon, E. F. E., and B. Z. Houlton. 2017. Coupled molecular and isotopic evidence for denitrifier controls over terrestrial nitrogen availability. *ISME Journal*, 11:727-740.

Levy-Booth, D. J., C. E. Prescott, and S. J. Grayston. 2014. Microbial functional genes involved in nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biology and Biochemistry*, 75:11-25.

Li, J., G. Wang, S. D. Allison, M. A. Mayes, and Y. Luo. 2014. Soil carbon sensitivity to temperature and carbon use efficiency compared across microbial-ecosystem models of varying complexity. *Biogeochemistry*, 119:67-84.

Lovett, G. M., K. C. Weathers, M. A. Arthur, and J. C. Schultz. 2004. Nitrogen cycling in a northern hardwood forest: Do species matter? *Biogeochemistry*, 67: 289-308.

Luo, Y., A. Ahlström, S. D. Allison, N. H. Batjes, V. Brovkin, N. Carvalhais, A. Chappell, P. Ciais, E. A. Davidson, A. Finzi, K. Georgiou, B. Guenet, O. Hararuk, J. W. Harden, Y. He, F. Hopkins, L. Jiang, C. Koven, R. B. Jackson, C. D. Jones, M. J. Lara, J. Liang, A. D. McGuire, W. Parton, C. Peng, J. T. Randerson, A. Salazar, C. A. Sierra, M. J. Smith, H. Tian, K. E. O. Todd-Brown, M. Torn, K. J. Van Groenigen, Y. P. Wang, T. O. West, Y. Wei, W. R. Wieder, J. Xia, X. Xu, X. Xu, and T. Zhou. 2016. Toward more realistic projections of soil carbon dynamics by Earth system models. *Global Biogeochemical Cycles*, 30:40-56.

- Luo, Y., E. Weng, X. Wu, C. Gao, X. Zhou, and L. Zhang. 2009. Parameter identifiability, constraint, and equifinality in data assimilation with ecosystem models. *Ecological Applications*, 19:571–574.
- Luo, Z., E. Wang, and O. J. Sun. 2017. Uncertain future soil carbon dynamics under global change predicted by models constrained by total carbon measurements. *Ecological Applications*, 27:1001–1009.
- Lynch, M. 2006. Streamlining and simplification of microbial genome architecture. *Annual Review of Microbiology*, 60:327–349.
- Magill, A. H., J. D. Aber, G. M. Berntson, W. H. McDowell, K. J. Nadelhoffer, J. M. Melillo, and P. Steudler. 2000. Long-term nitrogen additions and nitrogen saturation in two temperate forests. *Ecosystems*, 3:238–253.
- Magnúsdóttir, S., A. Heinken, L. Kutt, D. A. Ravcheev, E. Bauer, A. Noronha, K. Greenhalgh, C. Jäger, J. Baginska, P. Wilmes, R. M. T. Fleming, and I. Thiele. 2017. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35:81–89.
- Mann, S., and Y. P. P. Chen. 2010. Bacterial genomic G + C composition-eliciting environmental adaptation. *Genomics*, 95:7–15.
- Manzoni, S., and A. Porporato. 2009. Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry*, 41:1355–1379.

- Manzoni, S., P. Taylor, A. Richter, A. Porporato, and G. I. Ågren. 2012. Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. *New Phytologist*, 196:79-91.
- Martiny, A. C., K. Treseder, and G. Pusch. 2013. Phylogenetic conservatism of functional traits in microorganisms. *ISME Journal*, 7:830-838.
- McDowell, W. H., and G. E. Likens. 1988. Origin, composition, and flux of dissolved organic carbon in the Hubbard Brook valley. *Ecological Monographs*, 58:177-195.
- McGuire, K. L., and K. K. Treseder. 2010. Microbial communities and their relevance for ecosystem models: Decomposition as a case study. *Soil Biology and Biochemistry*, 42:529-535.
- Midgley, M. G., and R. P. Phillips. 2014. Mycorrhizal associations of dominant trees influence nitrate leaching responses to N deposition. *Biogeochemistry*, 117:241-253.
- Moles, A. T., D. D. Ackerly, C. O. Webb, J. C. Twiddle, J. B. Dickie, and M. Westoby. 2005. A brief history of seed size. *Science*, 307:576-580.
- Moorhead, D. L., and R. L. Sinsabaugh. 2006. A theoretical model of litter decay and microbial interaction. *Ecological Monographs*, 76:151-174.
- Mooshammer, M., W. Wanek, I. Hämmerle, L. Fuchslueger, F. Hofhansl, A. Knoltsch, J. Schneckner, M. Takriti, M. Watzka, B. Wild, K. M. Keiblinger, S. Zechmeister-Boltenstern, and A. Richter. 2014. Adjustment of microbial nitrogen use efficiency

to carbon: Nitrogen imbalances regulates soil nitrogen cycling. *Nature Communications*, 5:3694.

Morrissey, E. M., R. L. Mau, E. Schwartz, J. G. Caporaso, P. Dijkstra, N. Van Gestel, B. J. Koch, C. M. Liu, M. Hayer, T. A. McHugh, J. C. Marks, L. B. Price, and B. A. Hungate. 2016. Phylogenetic organization of bacterial activity. *ISME Journal*, 10:2336-2340.

Morrissey, E. M., R. L. Mau, E. Schwartz, T. A. McHugh, P. Dijkstra, B. J. Koch, J. C. Marks, and B. A. Hungate. 2017. Bacterial carbon use plasticity, phylogenetic diversity and the priming of soil organic matter. *ISME Journal*, 11:1890-1899.

Myrriotis, V., R. M. Rees, C. F. E. Topp, and M. Williams. 2018. A systematic approach to identifying key parameters and processes in agroecosystem models. *Ecological Modelling*, 368:344–356.

Navarro, D. J. 2015. *Learning statistics with R: A tutorial for psychology students and other beginners*. University of Adelaide, Adelaide, Australia.

Norman, J. S., and M. L. Friesen. 2017. Complex N acquisition by soil diazotrophs: How the ability to release exoenzymes affects N fixation by terrestrial free-living diazotrophs. *ISME Journal*, 11:315-326.

O'Brien, E. J., J. M. Monk, and B. O. Palsson. 2015. Using genome-scale models to predict biological capabilities. *Cell*, 161:971-987.

- Orwin, K. H., M. U. F. Kirschbaum, M. G. St John, and I. A. Dickie. 2011. Organic nutrient uptake by mycorrhizal fungi enhances ecosystem carbon storage: A model-based assessment. *Ecology Letters*, 14:493-502.
- Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. Xia, and R. Stevens. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42:D206-D214.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20:289-290.
- Phillips, R. P., E. Brzostek, and M. G. Midgley. 2013. The mycorrhizal-associated nutrient economy: a new framework for predicting carbon-nutrient couplings in temperate forests. *New Phytologist*, 199:41–51.
- Pinheiro, J., D. Bates, S. DebRoy, and D. Sarkar. 2007. nlme: Linear and Nonlinear Mixed Effects Models. R package version.
- Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, 5:e9490.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Read, D. J., and J. Perez-Moreno. 2003. Mycorrhizas and nutrient cycling in ecosystems -

- A journey towards relevance? *New Phytologist*, 157:475-492.
- Reed, S. C., C. C. Cleveland, and A. R. Townsend. 2011. Functional ecology of free-living nitrogen fixation: A contemporary perspective. *Annual Review of Ecology, Evolution, and Systematics*, 42:489-512.
- Reed, S. C., A. R. Townsend, C. C. Cleveland, and D. R. Nemergut. 2010. Microbial community shifts influence patterns in tropical forest nitrogen fixation. *Oecologia*, 164:521-531.
- Regan, K., B. Stempfhuber, M. Schlöter, F. Rasche, D. Prati, L. Philippot, R. S. Boeddinghaus, E. Kandeler, and S. Marhan. 2017. Spatial and temporal dynamics of nitrogen fixing, nitrifying and denitrifying microbes in an unfertilized grassland soil. *Soil Biology and Biochemistry*, 109:214-226.
- Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3:217-223.
- Robertson, G. P., and P. M. Groffman. 2015. Nitrogen transformations. *Soil Microbiology, Ecology and Biochemistry*, 421-446.
- Rocca, J. D., E. K. Hall, J. T. Lennon, S. E. Evans, M. P. Waldrop, J. B. Cotner, D. R. Nemergut, E. B. Graham, and M. D. Wallenstein. 2015. Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *ISME Journal*, 9:1693-1699.

- Roller, B. R. K., and T. M. Schmidt. 2015. The physiology and ecological implications of efficient growth. *ISME Journal*, 9:1481-1487.
- Roller, B. R. K., S. F. Stoddard, and T. M. Schmidt. 2016. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nature Microbiology*, 1:16160.
- Sanderman, J., T. Hengl, and G. J. Fiske. 2017. Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences*, 114:9575-9580.
- Savage, K., E. A. Davidson, and A. D. Richardson. 2008. A conceptual and practical approach to data quality and analysis procedures for high-frequency soil respiration measurements. *Functional Ecology*, 22:1000-1007.
- Scharnagl, B., J. A. Vrugt, H. Vereecken, and M. Herbst. 2010. Information content of incubation experiments for inverse estimation of pools in the Rothamsted carbon model: A Bayesian perspective. *Biogeosciences*, 7:763–776.
- Schellenberger, J., R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. Palsson. 2011. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols*, 2:727-738.
- Schimel, J. 2003. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. *Soil Biology and Biochemistry*,

35:549–563.

- Schmidt, M. W. I., M. S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I. A. Janssens, M. Kleber, I. Kögel-Knabner, J. Lehmann, D. A. C. Manning, P. Nannipieri, D. P. Rasse, S. Weiner, and S. E. Trumbore. 2011. Persistence of soil organic matter as an ecosystem property. *Nature*, 478:49-56.
- Shade, A., H. Peter, S. D. Allison, D. L. Baho, M. Berga, H. Bürgmann, D. H. Huber, S. Langenheder, J. T. Lennon, J. B. H. Martiny, K. L. Matulich, T. M. Schmidt, and J. Handelsman. 2012. Fundamentals of microbial community resistance and resilience. *Frontiers of Microbiology*, 3:417.
- Shi, Z., S. Crowell, Y. Luo, and B. Moore. 2018. Model structures amplify uncertainty in predicted soil carbon responses to climate change. *Nature Communications*, 9:1–11.
- Shi, Z., Y. Yang, X. Zhou, E. Weng, A. C. Finzi, and Y. Luo. 2016. Inverse analysis of coupled carbon-nitrogen cycles against multiple datasets at ambient and elevated CO₂. *Journal of Plant Ecology*, 9:285–295.
- Sims, G. K., T. R. Ellsworth, and R. L. Mulvaney. 1995. Microscale determination of inorganic nitrogen in water and soil extracts. *Communications in Soil Science and Plant Analysis*, 26:303-316.
- Sinsabaugh, R. L., J. Belnap, S. G. Findlay, J. J. Follstad Shah, B. H. Hill, K. A. Kuehn, C. R. Kuske, M. E. Litvak, N. G. Martinez, D. L. Moorhead, and D. D. Warnock.

2015. Scaling microbial biomass, metabolism and resource supply. *Biogeochemistry*, 122:175–190.
- Sinsabaugh, R. L., J. Belnap, S. G. Findlay, J. J. F. Shah, B. H. Hill, K. A. Kuehn, C. R. Kuske, M. E. Litvak, N. G. Martinez, D. L. Moorhead, D. D. Warnock, J. J. Follstad, S. Brian, and H. H. Kevin. 2014. Extracellular enzyme kinetics scale with resource availability. *Biogeochemistry*, 121:287–304.
- Sinsabaugh, R. L., S. Manzoni, D. L. Moorhead, and A. Richter. 2013. Carbon use efficiency of microbial communities: Stoichiometry, methodology and modelling. *Ecology Letters*, 16:930-939.
- Sinsabaugh, R. L., B. L. Turner, J. M. Talbot, B. G. Waring, J. S. Powers, C. R. Kuske, D. L. Moorhead, and J. J. F. Shah. 2016. Stoichiometry of microbial carbon use efficiency in soils. *Ecological Monographs*, 86:172-189.
- Sistla, S. A., E. B. Rastetter, and J. P. Schimel. 2014. Responses of a tundra system to warming using SCAMPS: A stoichiometrically coupled, acclimating microbeplantsoil model. *Ecological Monographs*, 84:151-170.
- Sridhara, V., a. G. Meyer, P. Rai, J. E. Barrick, P. Ravikumar, D. Segre, and C. O. Wilke. 2014. Predicting bacterial growth conditions from bacterial physiology. *bioRxiv*:0–26.
- Sterkenburg, E., K. E. Clemmensen, A. Ekblad, R. D. Finlay, and B. D. Lindahl. 2018.

Contrasting effects of ectomycorrhizal fungi on early and late stage decomposition in a boreal forest. *ISME Journal*, 12:2187-2197.

- Stockmann, U., M. A. Adams, J. W. Crawford, D. J. Field, N. Henakaarchchi, M. Jenkins, B. Minasny, A. B. McBratney, V. de R. de Courcelles, K. Singh, I. Wheeler, L. Abbott, D. A. Angers, J. Baldock, M. Bird, P. C. Brookes, C. Chenu, J. D. Jastrow, R. Lal, J. Lehmann, A. G. O'Donnell, W. J. Parton, D. Whitehead, and M. Zimmermann. 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems and Environment*, 164:80-99.
- Sulman, B. N., J. A. M. Moore, R. Abramoff, C. Averill, S. Kivlin, K. Georgiou, B. Sridhar, M. D. Hartman, G. Wang, W. R. Wieder, M. A. Bradford, Y. Luo, M. A. Mayes, E. Morrison, W. J. Riley, A. Salazar, J. P. Schimel, and J. Tang. 2018. Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry*, 141:109-123.
- Tang, J. Y. 2015. On the relationships between the Michaelis-Menten kinetics, reverse Michaelis-Menten kinetics, equilibrium chemistry approximation kinetics, and quadratic kinetics. *Geoscientific Model Development*, 8:3823-3835.
- Templer, P. H., G. M. Lovett, K. C. Weathers, S. E. Findlay, and T. E. Dawson. 2005. Influence of tree species on forest nitrogen retention in the Catskill Mountains, New York, USA. *Ecosystems*, 8:1-16.

- Thiele, I., and B. Palsson. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5:93-121.
- Thornton, P. E., J. F. Lamarque, N. A. Rosenbloom, and N. M. Mahowald. 2007. Influence of carbon-nitrogen cycle coupling on land model response to CO₂ fertilization and climate variability. *Global Biogeochemical Cycles*, 21:1–15.
- Tian, H., C. Lu, J. Yang, K. Banger, D. N. Huntzinger, C. R. Schwalm, A. M. Michalak, R. Cook, P. Ciais, D. Hayes, M. Huang, A. Ito, A. K. Jain, H. Lei, J. Mao, S. Pan, W. M. Post, S. Peng, B. Poulter, W. Ren, D. Ricciuto, K. Schaefer, X. Shi, B. Tao, W. Wang, Y. Wei, Q. Yang, B. Zhang, and N. Zeng. 2015. Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochemical Cycles*, 29:10.1002/2014GB005021.
- Todd-Brown, K. E. O., F. M. Hopkins, S. N. Kivlin, J. M. Talbot, and S. D. Allison. 2012. A framework for representing microbial decomposition in coupled climate models. *Biogeochemistry*, 109:19-33.
- Toosi, E. R., J. P. Schmidt, and M. J. Castellano. 2014. Soil temperature is an important regulatory control on dissolved organic carbon supply and uptake of soil solution nitrate. *European Journal of Soil Biology*, 61:68-71.
- Trivedi, P., M. Delgado-Baquerizo, C. Trivedi, H. Hu, I. C. Anderson, T. C. Jeffries, J. Zhou, and B. K. Singh. 2016. Microbial regulation of the soil carbon cycle: evidence

- from gene–enzyme relationships. *ISME Journal*, 10:2593.
- Varma, A., and B. O. Palsson. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Applied and Environmental Microbiology*, 60:3724-3731.
- Venables, W. N., and B. D. Ripley. 2002. *Modern applied statistics with S*. Springer.
- Vitousek, P. M., K. Cassman, C. Cleveland, T. Crews, C. B. Field, N. B. Grimm, R. W. Howarth, R. Marino, L. Martinelli, E. B. Rastetter, and J. I. Sprent. 2002. Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry*, 57:1-45.
- Vitousek, P. M., and R. W. Howarth. 1991. Nitrogen limitation on land and in the sea: How can it occur? *Biogeochemistry*, 13:87-115.
- Wallenstein, M. D., and R. J. Vilgalys. 2005. Quantitative analyses of nitrogen cycling genes in soils. *Pedobiologia*, 49:665-672.
- Waring, B. G., C. Averill, and C. V. Hawkes. 2013. Differences in fungal and bacterial physiology alter soil carbon and nitrogen cycling: Insights from meta-analysis and theoretical models. *Ecology Letters*, 16:887–94.
- Warren, C. R. 2014. Organic N molecules in the soil solution: What is known, what is unknown and the path forwards. *Plant and Soil*, 375:1-19.

- Webb, C. O., D. D. Ackerly, and S. W. Kembel. 2008. Phylocom: Software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics*, 24:2098-2100.
- Wieder, W. R., G. B. Bonan, and S. D. Allison. 2013. Global soil carbon projections are improved by modelling microbial processes. *Nature Climate Change*, 3:909-912.
- Wieder, W. R., A. S. Grandy, C. M. Kallenbach, and G. B. Bonan. 2014. Integrating microbial physiology and physio-chemical principles in soils with the MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences*, 11:3899-3917.
- Wieder, W. R., M. D. Hartman, B. N. Sulman, Y. P. Wang, C. D. Koven, and G. B. Bonan. 2018. Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Change Biology*, 24:1563-1579.
- Xia, J., and S. Wan. 2008. Global response patterns of terrestrial plant species to nitrogen addition. *New Phytologist*, 179:428-439.
- Yanai, R. D., C. R. Levine, M. B. Green, and J. L. Campbell. 2012. Quantifying uncertainty in forest nutrient budgets. *Journal of Forestry*, 110:448-456.
- Yanai, R. D., M. A. Vadeboncoeur, S. P. Hamburg, M. A. Arthur, C. B. Fuss, P. M. Groffman, T. G. Siccama, and C. T. Driscoll. 2013. From missing source to missing sink: Long-term changes in the nitrogen budget of a northern hardwood forest. *Environmental Science and Technology*, 47:11440-11448.

Zaehle, S. 2013. Terrestrial nitrogen-carbon cycle interactions at the global scale.
Philosophical Transactions of the Royal Society B: Biological Sciences,
368:20130125

CURRICULUM VITAE

