UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title
Shifting terrain: Soil microbial communities in precarious climates and methodological contestations

Permalink
https://escholarship.org/uc/item/5qq0q0jt

Author
Qin, Clara

Publication Date
2023

Copyright Information
This work is made available under the terms of a Creative Commons Attribution-ShareAlike License, available at https://creativecommons.org/licenses/by-sa/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SHIFTING TERRAIN: SOIL MICROBIAL COMMUNITIES IN PRECARIOUS CLIMATES AND METHODOLOGICAL CONTESTATIONS

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY in

ENVIRONMENTAL STUDIES

By

Clara Qin

September 2023

The Dissertation of Clara Qin is approved:

Professor Stacy Philpott, chair

Professor Julie Guthman

Professor Kathleen Treseder

Peter Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by
Clara Qin
2023

TABLE OF CONTENTS

List of Tables iv

List of Figures v

Abstract vii

Dedication ix

Acknowledgments x

Introduction 1

Chapter 1 – From DNA sequences to microbial ecology: Wrangling NEON soil microbe data with the *neonMicrobe* R package 4

Chapter 2 – Niche modeling predicts that soil fungi occupy a precarious climate in boreal forests 62

Chapter 3 – "It could be anyway": Ambivalence to the knowing of soilborne plant disease in post-fumigant agriculture 108

Supplementary Files 145

References 146

LIST OF TABLES

Table 1.1. ANOVA results for the sensitivity of ASV Shannon diversity to quality filtering parameters *truncLen*$_R$ and *maxEE*$_R$ 37

Table 1.2. ANOVA results for the sensitivity of observed ASV richness to quality filtering parameters *truncLen*$_R$ and *maxEE*$_R$ 38

Table 1.3. PERMANOVA results for the sensitivity of Bray-Curtis dissimilarity to quality filtering parameters *truncLen*$_R$ and *maxEE*$_R$ 39

Table 1.S1. Permutational analysis of variance (PERMANOVA) results for the sensitivity of Bray-Curtis dissimilarity to quality filtering parameters *truncLen*$_R$ and *maxEE*$_R$, after removal of communities produced with *truncLen*$_R$ = 170 51

Table 1.S2. Permutational analysis of variance (PERMANOVA) results for the sensitivity of Bray-Curtis dissimilarity to quality filtering parameters *truncLen*$_R$, *maxEE*$_R$, and *sampleID* 52

Table 2.1. Variable importance of the full set of covariates across ridge regression models for all fungal operational taxonomic units (OTUs; n = 8,597), in order of median variable importance 87

Table 2.S1. Out-of-sample predictive performance of logistic ridge regression models and maximum entropy models for the presence and absence of 8,597 fungal operational taxonomic units (OTUs) present in at least 10 operational sites 98

LIST OF FIGURES

Figure 1.1. Overview of the *neonMicrobe* package's marker gene sequence processing pipeline 41

Figure 1.2. Default organization of input data for the *neonMicrobe* R package 43

Figure 1.3. Simple plots produced using the outputs of the *neonMicrobe* pipeline for analyses of 16S sequences in soil samples collected from NEON grassland sites 45

Figure 1.4. Retention of NEON 16S reads throughout the processing pipeline across different values of quality filtering parameters $truncLen_R$ and $maxEE_R$ 47

Figure 1.5. Shannon diversity of ASV tables produced from NEON 16S marker gene sequences across different values of quality filtering parameters $truncLen_R$ and $maxEE_R$ 48

Figure 1.6. Community composition of ASV tables produced by NEON 16S marker gene sequences across different values of quality filtering parameters $truncLen_R$ and $maxEE_R$ 50

Figure 1.S1. Generalized approach to analyzing NEON microbial sequence data for ecological analyses using the DADA2 pipeline 55

Figure 1.S2. Cumulative expected error profiles for NEON 16S sequencing runs BDNB6 and BFDG8, based on the first 10 samples from each run 57

Figure 1.S3. Standard ANOVA diagnostic plots for the sensitivity of ASV Shannon diversity to quality filtering parameters 58

Figure 1.S4. Standard ANOVA diagnostic plots for the sensitivity of observed ASV richness to quality filtering parameters 59

Figure 1.S5. Observed richness of ASV tables produced from NEON 16S marker gene sequences across different values of quality filtering parameters $maxEE_R$ and $truncLen_R$ 60

Figure 1.S6. Tukey's HSD test for the homogeneity of group dispersions (variances) for different values of $truncLen_R$ 61

Figure 2.1. Conceptual diagram of the relationship between niche distributions and the climate sensitivity of soil fungal assemblage composition, i.e. Sørensen climate sensitivity 88

Figure 2.2. Distribution of soil sampling sites in geographic space and climate space 89

Figure 2.3. Fungal species niche edges across temperature and precipitation gradients in North America 91

Figure 2.4. Standardized niche metrics and climate sensitivity of soil fungal assemblage composition across North American climates 93

Figure 2.5. Niche metrics and climate sensitivity of soil fungi across North America 95

Figure 2.6. Standardized Sørensen climate sensitivity of soil fungi across major North American biomes 97

Figure 2.S1. Fungal species presence–absence transitions across a temperature gradient, by genus 99

Figure 2.S2. Fungal species niche edges across temperature and precipitation gradients in North America under the null model 101

Figure 2.S3. Niche metrics and climate sensitivity of soil fungi across a two-dimensional slice of North American climate space, including null and standardized distributions 103

Figure 2.S4. Niche metrics and climate sensitivity of soil fungi across North America 105

Figure 2.S5. Variable importance of the full set of predictors across all ridge regression models, in order of median variable importance 106

Figure 2.S6. Model validation metrics for the logistic ridge regression models including all predictors except soil pH and total soil nitrogen content 107

Figure 3.1. Sociomaterial knowledge practices enacting multiple ontologies of soilborne plant disease in the agricultural sciences, and their configuration with(in) production agriculture 118

Figure 3.2. An ontological contestation between pathogenic disease and microbiomic disease and its temporary resolution in deference to an existing disease management practice 137

ABSTRACT

SHIFTING TERRAIN: SOIL MICROBIAL COMMUNITIES IN PRECARIOUS CLIMATES AND METHODOLOGICAL CONTESTATIONS

Clara Qin

Soil microbial communities play critical roles as decomposers, plant pathogens, and plant mutualists in terrestrial ecosystems. Understanding how these communities will respond to a changing environment requires well-replicated studies across diverse ecological contexts and a careful assessment of the methodological dependencies of large-scale microbiological research. In my dissertation, I develop methods for the study of soil microbial communities across three registers: informatic, ecological, and anthropological. First, I present a software package to facilitate access to an open-access, continental-scale dataset of soil microbial DNA metabarcoding sequences for ecological research (Chapter 1). Second, I use this dataset to predict the spatial distributions of fungal taxa across the U.S. and Canada and quantify the sensitivity of soil fungal community composition to ongoing climate change (Chapter 2). Finally, taking an ethnographic approach, I observe how soilborne plant disease comes to be known through a variety of scientific practices that stabilize distinct objects of study (Chapter 3). Together, this work illustrates that while soil microbial communities everywhere may be shifting, these shifts are not the same everywhere. The North American boreal forest occupies a particularly precarious climate in which even slight warming can create major shifts in the composition of its soil fungal communities. Additionally, the enrollment of soil microbial communities in the transition to post-

fumigant agriculture contributes to a complex political terrain wherein the possibilities for knowing soil microbial communities are entangled with the possibilities for managing soilborne plant disease. By showcasing diverse methods for studying soil microbial communities, this dissertation puts forth an ontological approach to scientific inquiry that does not presuppose objects as a given but rather enacts them anew through methodological innovations.

DEDICATION

To Adrienne and Sucy

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my committee members: Julie Guthman, Kathleen Treseder, and especially the chair of my committee, Stacy Philpott. Their insightful feedback and professional mentorship were pivotal in the completion of this dissertation. My gratitude extends to my co-authors, who contributed to the content of the first two chapters of this dissertation. These include Peter Pellitier, Michael Van Nuland, Kabir Peay, Kai Zhu, Ryan Bartelme, Anny Chung, Dawson Fairbanks, Yang Lin, Daniel Liptzin, Chance Muscarella, Kusum Naithani, Ayanna St. Rose, Lee Stanish, and Zoey Werbin.

I recognize the unwavering support of the mentors and colleagues who bolstered my sense of purpose and belonging throughout my graduate studies. I am grateful to Kabir Peay, Jorge Ramos, Julie Lee-Yaw, and Fushcia Hoover for offering invaluable perspectives beyond my home institution. I am indebted to Carol Shennan, Joji Muramoto, Margherita Zavatta, Peter Henry, Polly Goldman, Jade Ermita, Jason Hong, and Erin Rosskopf for helping me gain a footing in agricultural science. I also appreciate the support of fellow graduate students in the Environmental Studies department, including my cohort, labmates, and housemates. Special mention goes to Azucena Lucatero and Anna Nisi for always lending an ear. Additionally, I am proud to be a member of the various professional organizations and affinity groups that welcomed me: the UC Santa Cruz Women of Color Graduate Community, the WOC in EEB Slack, and Women in Soil Ecology.

Thanks are due to the scholars who introduced me to Science and Technology Studies and the social sciences. This acknowledgment extends to Julie Guthman, Anna Tsing, Madeleine Fairbairn, Maywa Montenegro, Emily Reisman, members of the Agri-Food Working Group at UC Santa Cruz, and participants in the Science and Technology Food and Agriculture Network (STSFAN). Charlie Mather, Katie Ulrich, Summer Sullivan, Halie Kampman, Indiana Reid-Shaw, Edith Gonzales, and Abby Cunniff provided invaluable insights during various iterations of my third chapter, helping me realize the value of my writing. Appreciations also go to the agricultural scientists who generously agreed to be interviewed for my third chapter.

This work was made possible through external funding from the National Science Foundation, Out To Innovate, the Sonoma County Mycological Association, and the NEON Early Career Scholars Program, as well as internal funding from the Department of Environmental Studies, Division of Social Sciences, and Division of Graduate Studies at UC Santa Cruz.

I am profoundly grateful to my labor union, especially fellow organizers at UC Santa Cruz, for safeguarding student-workers' rights and building a movement for a fair and accessible higher education system.

Finally, I wish to express my heartfelt appreciation to the members of my family, by blood or by choice, who never stopped believing in me.

The text of this dissertation includes reprints of the following previously published material:

Qin, C., R. Bartelme, Y. A. Chung, D. Fairbanks, Y. Lin, D. Liptzin, C. Muscarella, K. Naithani, K. Peay, P. Pellitier, A. St. Rose, L. Stanish, Z. Werbin, and K. Zhu. 2021. From DNA sequences to microbial ecology: Wrangling NEON soil microbe data with the *neonMicrobe* R package. Ecosphere 12.

Qin, C., P. T. Pellitier, M. E. Van Nuland, K. G. Peay, and K. Zhu. 2023. Niche modelling predicts that soil fungi occupy a precarious climate in boreal forests. Global Ecology and Biogeography:geb.13684.

The co-author listed in this publication directed and supervised the research which forms the basis for the dissertation.

Clara Qin led the analysis, software development, and writing of the previously published material with input from all co-authors.

INTRODUCTION

Everything has come to depend on soil microbial communities. Upheld as a natural climate solution (Anthony et al., 2022), the key to sustainable food production (Busby et al., 2017), and a model for hospitable relations with the more-than-human (Puig de la Bellacasa, 2015), soil microbial communities have emerged as an impossibly potent ally in overcoming the interlocking crises of the Anthropocene. Yet when challenged, the instability of soil microbial communities suggests that they cannot support humanity's hopes on their own. Soil microbial communities are transformed under changing climates (Cavicchioli et al., 2019); their beneficial associations with plants are bred out of modern crops (Cordovez et al., 2019); and their ecological relations turn out to be different from the speculations that made them morally relevant (Paxson and Helmreich, 2014). In methodological considerations, the solidity of soil microbial communities begins to unravel into fractal possibilities. Soil microbes cultured in a Petri dish look and behave differently from soil microbes identified in a bioinformatic analysis of environmental DNA. Within the bioinformatic analysis, too, even minor perturbations in the sample collection process, the DNA sequencing protocol, and the data processing script can have an outsized influence on the resulting microbiome metrics. It would seem, then, that soil microbial communities also depend on everything else.

In the traditional ordering logic of Western science, the instability of soil microbial communities should amount to no less than a scientific crisis (Kuhn, 1962).

To stabilize the object of study, it should be necessary to reinscribe clear boundaries between *interpretations* of soil microbial communities and soil microbial communities *in themselves*—perhaps by establishing common methodological practices that unite a community of researchers. Once settled, it should finally be possible to study objective matters, such as the resilience of soil microbial communities to a changing climate, or the proliferation of soilborne fungal pathogens and their associated plant diseases in California specialty crop agriculture. To an extent, the structure of this dissertation reflects this logic. But its disciplinary breadth suggests a different reading that challenges distinctions between subject and object, social and natural.

In Chapter 1, I introduce *neonMicrobe*, a suite of computational tools for downloading, processing, and conducting sensitivity analyses on microbial DNA metabarcoding data from soil samples of the National Ecological Observatory Network (NEON). NEON is a federally-funded platform for open ecological data, making it a standard-bearer for ecological research in North American ecosystems. Thus, as an open-source software library designed to facilitate access into this platform, *neonMicrobe* is a bid to establish a standardized package (Fujimura, 1988) for soil microbial ecology.

In Chapter 2, I combine NEON data with a second soil sampling network and spatially gridded climate data to map the distributions of thousands of fungal taxa across the U.S. and Canada. With these distributions, I develop a novel method for measuring the sensitivity of soil fungal community composition to ongoing climate

change. This analysis predicts that fungal communities in boreal forest soils occupy an especially precarious climate, painting a grim picture for the climate resilience of an already threatened ecosystem.

In Chapter 3, I describe a shift in the methods by which soilborne plant disease is known, especially as it relates to the phaseout of methyl bromide, a soil fumigant, from California specialty crop agriculture. Through an ethnographic method (Mol, 2002), I show how the variety of methods used to study soilborne plant disease in an agricultural research laboratory stabilizes a multiplicity of objects "in themselves," not merely their "interpretations" —as if these could ever be separate. This reveals new possibilities for how soil microbial communities can be known. On the other hand, it also risks sustaining certainties about how industrial agriculture can be done.

It is possible to read this dissertation with the understanding that statements of fact are always contingent on the practices that bring objects into being. Attending to knowledge practices recasts this dissertation as a knowledge-making project among many possible knowledge-making projects. In other words, this dissertation is not a point of closure in the study of soil microbial communities. It is yet another point of departure.

CHAPTER 1

From DNA sequences to microbial ecology: Wrangling NEON soil microbe data with the *neonMicrobe* R package

Clara Qin, Ryan Bartelme, Y. Anny Chung, Dawson Fairbanks, Yang Lin, Daniel Liptzin, Chance Muscarella, Kusum Naithani, Kabir Peay, Peter Pellitier, Ayanna St. Rose, Lee Stanish, Zoey Werbin & Kai Zhu

Published in *Ecosphere*, 2021

Introduction

Microbial life on earth is ubiquitous and essential in critical ecosystem processes (Cavicchioli et al. 2019). Soils are among the most diverse microbial habitats known, and recent surveys at continental (Fierer et al. 2012, Ladau et al. 2013, Talbot et al. 2014, Prober et al. 2015, Thompson et al. 2017, Wang et al. 2018) and global scales (Serna-Chavez et al. 2013, Thompson et al. 2017, Chu et al. 2020) have shed light on the diversity and distribution of soil microbes. Such large-scale studies often identify abiotic environmental factors, such as climate and edaphic characteristics, to be strong predictors of soil microbial community composition. For example, soil fungal richness is strongly determined by climate (Tedersoo et al. 2014, Větrovský et al. 2019, Steidinger et al. 2020), soil protist composition by annual

precipitation (Oliverio et al. 2020), and bacterial composition and function by edaphic characteristics such as pH and soil carbon (Lauber et al. 2009, Delgado-Baquerizo et al. 2016). Enough data has accumulated that global meta-analyses of environmental controls of microbial biogeography have recently emerged (Větrovský et al. 2019), and we are just beginning to discern the influence of biotic interactions on microbial diversity and distribution (Bahram et al. 2018, Steidinger et al. 2019).

While molecular microbial surveys have contributed much to our understanding of microbial ecology, they have also highlighted unique problems. One conceptual challenge in microbial ecology is scale. Classical ecological theories and sampling techniques were developed with macro-organisms in mind, and may not apply well to microbes (Levin 1992). Thus, drivers of microbial diversity and distribution are highly scale-dependent (Martiny et al. 2011, Peay et al. 2016). Sampling techniques that preserve sample scale-dependency are important to broaden our understanding of microbial community ecology. For example, the strength of positive and negative interactions in microbial community assembly processes is predicted to occur at distinct spatial scales (Mod et al. 2020). However, spatially explicit tests of assembly rules for microbial groups are lacking (Talbot et al. 2014, Maynard et al. 2017), and cross-study comparisons are stymied by widely varying sample granularity and survey extent (Zinger et al. 2019). In addition to variation in sampling scale, the differences between protocols of sequence-based microbial ecological research can make the interpretation of meta-analyses challenging. For example, differences in sample collection, such as soil core size, storage method,

DNA extraction, sequencing, and bioinformatic approaches, can all have oversized effects on our final understanding of microbial abundance, diversity, and distribution (Lindahl et al. 2013, Pauvert et al. 2019). Thus, laboratory and bioinformatic standardization must complement field sampling designs to truly empower ecological inferences.

The National Ecological Observatory Network (NEON) is a multi-scale ecological observation platform spanning the United States for understanding and forecasting the impacts of climate change, land use change, and invasive species on ecosystems (Schimel and Keller 2015). NEON is designed to enable users, including scientists, educators, policymakers, and the general public, to assess large-scale and long-term ecological changes and address their major drivers. The NEON Terrestrial Observation System monitors environmental drivers and key taxonomic groups at multiple trophic levels in order to quantify the responses of biodiversity and biogeochemical cycles to climate and land use changes. A component of this data collection program, the NEON Microbial Ecology Sampling Program, measures the diversity and abundances of microbiota, and archives raw samples and DNA for public research use. Sampling and analysis for microbes are performed using standardized and freely available methods (Stanish et al. 2018) that help eliminate confounding factors in cross-site or cross-study analyses. The data collected by NEON are processed into documented, calibrated, and quality-controlled data products, and are openly available through the NEON Data Portal and API (Application Programming Interface).

Given the high diversity of microbial communities relative to most macroorganisms, NEON microbial data also presents unique challenges for the end user. For those new to "big data," accessing the NEON API may seem unintuitive: acquiring metadata entails downloading data from multiple NEON products and preprocessing metadata through table joins before analysis. For high-throughput microbial marker gene (amplicon) sequencing data, files are relatively large and cannot be readily visualized, and often advanced bioinformatic tools and computing capabilities are needed to analyze the data. Our goal is to lower the barriers of entry to utilizing NEON microbial data. We provide a data processing pipeline and software package to wrangle NEON soil microbial community data and promote its wider accessibility and use in a standardized manner, thereby maximizing its potential for developing ecological insights.

In this paper, we introduce neonMicrobe, a novel, quality-tested data pipeline that standardizes the processing of NEON soil bacterial and fungal amplicon sequence data into abundance tables for microbial ecology research. While the current scope of this paper and the accompanying neonMicrobe R package is on soil microbes and the soil environment, our package provides the scaffolding for the analysis of surface water and benthic microbe marker gene sequence data. Our pipeline builds on existing validated protocols (Tedersoo, Ramirez, et al. 2015, Callahan et al. 2016, Lunch et al. 2021) to create a reproducible way to download, quality-control, and process sequence data into sequence tables all within the R statistical computing environment (R Core Team 2021). Acknowledging the

complexities behind selecting appropriate parameters for various stages of the pipeline, we use a sensitivity analysis to demonstrate how changes in read quality filtering parameters may affect downstream ecological inferences. We conclude with lessons learned related to the provisioning of microbial data by NEON, as well as the use of NEON data products by researchers to generate new insights about microbial ecology.

NEON Soil Microbe Marker Gene Sequence Data Products

This paper utilizes soil microbial 16S and ITS sequence data from NEON (NEON DP1.10108.001), which primarily target bacteria and fungi, respectively. A full description of the sampling design and analysis methods can be found in the documentation available through the NEON Data Portal website (data.neonscience.org).

Sampling design

The NEON domain encompasses 47 terrestrial field sites across the United States including Puerto Rico, covering 20 eco-climatic domains as defined by NEON. Sites are strategically located in ecosystems across the U.S. so that site-level measurements can be used to extrapolate across the continent (Barnett et al. 2017). Each terrestrial field site contains 10 plots for soil microbial sampling: four tower plots within the airshed of an instrumentation tower and six distributed plots that are designed to be spatially balanced while reflecting the dominant vegetation type at the

site (Stanish et al. 2018). Each plot is divided into four subplots, three of which are randomly chosen for sampling. Random coordinates with a 1-m buffer zone are generated for each subplot. Up to three soil cores may be collected within 0.5 m of each set of coordinates and combined to provide a sufficient sample volume for downstream processing and analyses.

Sampling events are broadly designed to capture periods when microbial activity is expected to be at its highest, or when activity may be rapidly changing, such as the transition from the dry to wet seasons or during spring soil thawing. Sampling does not occur when the ground is frozen or covered in snow due to logistical and safety concerns, which may miss critical periods in snow-covered ecosystems (Schadt et al. 2003). Most sites are sampled for microbial community characterization three times per year, with one event corresponding to peak plant productivity as measured using remote sensing data (Stanish et al. 2018, Stanish and Parker 2019). In sites where the activity is more strongly driven by precipitation than temperature, historical precipitation data are used to determine sampling periods.

Soil cores are collected down to a maximum depth of 30 cm. If an organic horizon is present, it is collected separately from the mineral horizon. The cores are co-located with other critical soil physical and biogeochemical measurements, including litter depth, temperature, moisture, pH, and nutrients. The metadata associated with NEON samples exceed the minimum standards defined by the Genomics Standards Consortium (Yilmaz et al. 2011). Once collected, the cores are separated by horizon, homogenized, and subsampled for microbial and chemical

analyses. The microbial samples are frozen in the field and shipped to analytical facilities for DNA sequencing analysis.

Molecular methods

Sample processing and analyses are performed using standardized methodologies to the extent possible to ensure comparability of data over time. However, methodologies and technologies will change and improve, and adapting to changes over time is critical. For full transparency, any changes in laboratory methods are captured in the freely available external laboratory standard operating procedures (SOPs), which are listed in the metadata for every downloaded sequence dataset.

The processing methods for generating 16S and ITS sequence data used in this analysis are detailed in the 16S and ITS Sequencing Standard Operating Procedure (Battelle Memorial Institute 2018). Genomic DNA from thawed soil samples is extracted using Qiagen DNeasy Powersoil HTP 96 Kits, and quantified with Quantifluor ONE dsDNA Kits. The marker genes targeted are the V3-V4 region of the 16S ribosomal RNA (rRNA) gene for bacteria and archaea (primers Pro341F and Pro805R, Takahashi et al. 2014) and the internal transcribed spacer (ITS) region of the rRNA operon for fungal identification (primers ITS1f and ITS2, Walters et al. 2015). Additional details on PCR processing and quality assurance can be found in the associated laboratory SOP (Battelle Memorial Institute 2019) and in the marker gene sequencing data product tables "mmg_soilPcrAmplification_16s" and

"mmg_soilPcrAmplification_ITS" (NEON DP1.10108.001). All sequencing runs are performed on an Illumina MiSeq v3 600-cycles cartridge as 300-bp paired-end reads, resulting in one set of forward and reverse sequencing reads for each sample. A sequencing run usually consists of a library of samples pooled from multiple sites and collection dates, as well as DNA extraction and PCR controls (Battelle Memorial Institute 2018). At the time of this study, only samples with a minimum of 3000 reads and post-trimming mean quality score of 20 pass the quality filter for the NEON Data Portal (Battelle Memorial Institute 2018).

The *neonMicrobe* Marker Gene Sequence Processing Pipeline

The *neonMicrobe* R package promotes data accessibility by allowing users, especially those lacking extensive bioinformatics experience, to wrangle NEON's soil microbe marker gene data products with greater ease and reproducibility. The pipeline begins by downloading NEON marker gene sequence data from the NEON Data Portal and produces amplicon sequence variant (ASV) abundance tables linked to associated taxonomic and soil abiotic data in a *Phyloseq* data structure. The pipeline, which builds on existing validated pipelines (Lindahl et al. 2013), explicitly considers the unique properties of NEON data with the goal of maximizing ecological insight of microbial communities. While there is no consensus on an optimal bioinformatic processing method for microbial amplicon sequence data (Pauvert et al. 2019), bioinformatic choices are consequential to downstream ecological analysis (Tedersoo, Anslan, et al. 2015, Tedersoo, Ramirez, et al. 2015). We preferred

bioinformatics approaches that would allow comparability across disparate datasets and require a relatively low level of programming knowledge for the end user (Callahan et al. 2016, 2017). These criteria align with a major goal of NEON: to use standard methods across datasets from large temporal and spatial scales, enabling many different studies to answer myriad questions.

The *neonMicrobe* R package includes functions for downloading, renaming, and subsetting NEON sequencing data, as well as custom wrappers for the *DADA2* algorithms (Callahan et al. 2017, Callahan et al. 2016). Briefly, *DADA2* allows analysis of microbial taxa at the resolution of exact ASVs, in contrast to the more traditional use of operational taxonomic units (OTUs) that are based on a user-defined nucleotide sequence similarity (e.g., 97-98.5% pairwise sequence identity). In addition to the biological benefits of finer DNA sequence resolution by employing ASVs, exact sequences are highly advantageous to OTUs because they can be directly compared across datasets, making them critical to NEON's coordinated network sampling design (Callahan et al. 2017). Another benefit is that the *DADA2* pipeline allows sequence processing and data analysis steps to be conducted in the R statistical computing environment (R Core Team 2021), which enhances reproducibility and lowers barriers to entry for those who are less familiar with command-line bioinformatics tools. *DADA2* has also demonstrated its compatibility with other methods and platforms in biological interpretation related to the assembly of paired-end reads, the treatment of chimeras, and the final filtering of the ASV tables (Pauvert et al. 2019). Our processing pipeline creates a ready-to-use soil

microbial dataset of unprecedented spatio-temporal range and taxonomic resolution. In the following subsections, we describe our processing pipeline. Each of the following subsections has a corresponding vignette in the *neonMicrobe* R package (Fig. 1.1), which can be accessed at https://github.com/claraqin/neonMicrobe.

Downloading and quality-controlling NEON soil microbe marker gene sequence data

The *neonMicrobe* data processing pipeline begins by leveraging the NEON Data API via the neonUtilities R package (Lunch et al. 2021) to acquire soil microbe marker gene sequencing data. First, the downloadSequenceMetadata function downloads and joins the tables within NEON data product DP1.10108.001 (Soil microbe marker gene sequences), which includes information about DNA extraction, PCR amplification, marker gene sequencing, and sequence file metadata. Because the output of this function contains information about sample processing but does not include the raw sequence files themselves, we refer to this output as *sequence metadata*. downloadSequenceMetadata can be parameterized to download a subset of raw sequence data according to a specific date range, site range, sequencing run, or target gene (16S or ITS). Metadata can be further filtered to remove records that include quality flags or fail certain quality tests, as described in greater detail below. These steps take place before the user downloads the raw sequence files, saving processing time and disk space. Next, the downloadRawSequenceData function references the metadata to download the desired raw sequence files. By default,

NEON data will be organized into a directory structure illustrated by Fig. 1.2. These functions are implemented in the vignette "Download NEON Data."

As with any analysis, ensuring that the downloaded data are high-quality, correctly formatted, and directly comparable is a critical data processing step. Performing quality control steps prior to entering the data analysis workflow can reduce downstream processing errors due to incomplete or improperly formatted data, and improves efficiency by conserving CPU time processing low-quality data that may ultimately be discarded. The NEON microbial data products contain data quality flags in which known quality issues are reported. In addition to quality issues, the sequence metadata contain other crucial details, some of which can significantly affect the comparability of sequencing runs, such as specific laboratory protocols, oligonucleotide primer sets, and sequencing platforms.We strongly recommend that users review the sequence metadata and consider whether additional data filtering should be performed based on the research needs and data stringency requirements.

We have implemented a number of basic quality control steps in the function qcMetadata. In this function, users can opt to (1) remove samples that are flagged as having low read quality or being legacy data, (2) check for and remove duplicate samples, and (3) prepare for a paired-reads analysis by removing samples for which only one read orientation is available.

Generating sequence tables and taxonomy tables using DADA2

16S and ITS sequences are processed by different variations of the *DADA2* workflow. Processing is done on a sequencing-run basis to allow for variable error rates between sequencing runs to optimize amplicon sequence variant (ASV) calling and chimera detection. For each sequencing run, the generalized steps are: (1) filtering samples to remove all reads containing ambiguous ("N") base calls; (2) removing PCR primers, using Cutadapt (Martin 2011) for ITS sequences but not for 16S sequences; (3) truncating (for 16S reads only) and filtering reads to ensure a minimum quality score; (4) building an error model for each sequencing run to describe the probability that a given read was produced from a given sample sequence; (5) denoising reads into ASVs through the DADA divisive partitioning algorithm based on an underlying nucleotide sequence error rate model; (6) optionally, removing chimeric sequences; and (7) joining sequence tables across all sequencing runs using the *DADA2* function mergeSequenceTables, which performs a simple merge, and the DADA2 function collapseNoMismatch, which performs 100% clustering on the ASVs. (Note that Cutadapt is not supported on Windows computers. For Windows users, we recommend running the ITS pipeline in another computer, or in a Docker container, as outlined in the section "Extending Scientific Workflow Reproducibility with Container Technology.") As an alternative to step 7 for ITS sequences, it may be prudent to cluster ASVs to a lower, user-specified sequence similarity threshold (e.g. 97-98.5%) using the VSEARCH or DECIPHER programs (Rognes et al. 2016, Wright 2016), because the same ITS ASV may have different

length variants across different sequencing runs. This step is not needed for 16S reads, for which 100%-similar ASVs can instead be combined using the *collapseNoMismatch* command in *DADA2*. Finally, a taxonomic reference database can be used to assign taxonomy to the ASVs. Many of these processing steps are wrapped into the novel functions trimPrimers16S, qualityFilter16S, and runDada16S, and their ITS-specific analogues.

Linking sequence data and soil abiotic data in a Phyloseq object

By taking advantage of different NEON data products, users can draw inferences regarding the relationships between soil microbial community characteristics, soil physical and chemical properties, climate variables, and other spatiotemporal processes. As demonstrated in the "Add Environmental Variables" vignettes, the end product of the *neonMicrobe* pipeline is a *Phyloseq* object linking the ASV table, its (optional) taxonomy table, and associated soil abiotic data (NEON DP1.10086.001) downloaded using the downloadSoilData function, creating a data structure that is ready for ecological analysis (McMurdie and Holmes 2013). While an overview of statistical microbial community analysis is beyond the scope of this paper, excellent reviews of the subject (Hugerth and Andersson 2017) and tutorials using *Phyloseq* are widely available (McMurdie and Holmes 2013) (https://www.bioconductor.org/packages/release/bioc/vignettes/phyloseq/inst/doc/phyloseq-analysis.html).

Example: Analysis of soil bacterial diversity in grasslands

The processed data are immediately usable in analyses to answer ecological questions. To demonstrate this, we present a relatively simple analysis of soil bacterial diversity using NEON 16S sequence data that has been processed and assembled by *neonMicrobe* (Fig. 1.3). The code for this analysis is available as Supplementary File 1.1.

In this example analysis we asked, what controls soil bacterial communities within and across sites in a grassland ecosystem? We included three sites—Central Plains Experimental Range (CPER), Konza Prairie Biological Station (KONZ), and Northern Great Plains Research Laboratory (NOGP)—as these sites share Argiustoll soils, but vary in climate and belong to different NEON eco-climatic domains (Fig. 1.3a). We used soil samples that were collected at peak plant productivity in 2017 ($n = 86$ samples), in order to minimize temporal effects. Across these sites, we examined the effects of soil pH and soil moisture on soil bacterial composition, as these were previously found to explain substantial variation in soil bacterial community composition across NEON sites (Docherty et al. 2015). We additionally included mean annual temperature (MAT) and mean annual precipitation (MAP) as climatic covariates. We used the adonis2 function in the vegan R package (Oksanen et al. 2020) to conduct permutational analysis of variance (PERMANOVA). We found that significant drivers of bacterial community composition in these grassland sites included soil pH (PERMANOVA, $P < 0.001$), MAT ($P < 0.001$), and MAP ($P < 0.001$), while the effect of soil moisture was not significant ($P = 0.089$). Our results

suggest that climatic variables drive between-site variation, while soil pH drives within-site variation, in soil bacterial community composition across grasslands (Fig. 1.3d).

Extending Scientific Workflow Reproducibility with Container Technology

Reproducibility is a major principle of the scientific approach and is critical for successful application of the bioinformatics pipeline. One challenge that hinders reproducibility in computational tools is the staggeringly large number of possible combinations of operating systems, programming languages, and package versions a user may have installed locally, which allows variability to creep into analysis pipelines. Furthermore, there is a significant time and energy investment required to install dependencies, check operating system compatibilities, and implement code at a large scale. To minimize this cognitive load, efforts like the Open Container Initiative started by Docker (https://opencontainers.org/) enable the deployment of discretized applications to cloud computing infrastructure. This container paradigm extends into bioinformatics tools through the BioContainers initiative (da Veiga Leprevost et al. 2017). The offering of these computational biology tools as containers allows users to move away from user-specific workflow generation on high-performance computing (HPC) systems and into cloud native scientific computing.

Due to the network of data products, supporting R packages, bioinformatic tools, computational resources, and operating system compatibility requirements associated with *neonMicrobe,* the *neonMicrobe* R package cannot encapsulate a

reproducible scientific workflow on its own (Boettiger 2015). To extend its reproducibility, two Docker container images were created for *neonMicrobe*. First, an RStudio Server instance was created from the Rocker Group's RStudio Server *tidyverse* base image (Nüst et al. 2020). This RStudio image is freely available on the CyVerse Docker Hub (https://hub.docker.com/repository/docker/cyversevice/rstudio-neon-dada2), as well as through the CyVerse Discovery Environment's (DE) Visual Interactive Computing Environment (VICE) as the "rstudio_neon_microbiome" application. The CyVerse DE allows users to interact with data and docker containers on VICE without explicitly requiring mastery of Docker in the command line. The second Docker container image of *neonMicrobe* is strictly command-line based and designed for scaling to larger cloud systems; it is also available on Docker Hub (https://hub.docker.com/r/rbartelme/neonmicrobe). Therefore, users may utilize either of these containers on their local systems, increasing both access to the tools and creating a more easily reproduced environment to conduct microbial ecology experimental analyses.

Sensitivity Analysis of Quality Filtering Parameters

The choice of bioinformatic software and processing parameters can have implications for the accuracy of the inferred microbial community (Pauvert et al. 2019, Prodan et al. 2020). While we make some recommendations for processing the NEON marker gene sequences, such as the use of *DADA2* over OTU-based processing pipelines, we leave other decisions to the researcher depending on their

research needs and computing capacity (Fig. 1.S1). These decisions include but are not limited to: the removal or retention of reverse reads, the choice of parameters for the quality filter, and the choice of partitioning, alignment, and sequence comparison heuristics for *DADA2*.

Exploring all combinations of these decision points to arrive at an optimal processing pipeline is beyond the scope of this paper. We expect that the combination of choices that creates the most accurate representation of the NEON soil microbial communities will change depending on the specific set of samples being processed or the metrics of interest to the researcher. However, for any given instance of the NEON marker gene sequence data, it should be possible to evaluate how the choice of processing parameters influences some benchmark metrics related to the pipeline outputs. Here, we provide a framework for conducting a sensitivity analysis on the processing pipeline, using the quality filtering parameters for 16S amplicons as an example.

We investigated how the choice of parameters for the quality filter—which truncates or removes low-quality reads—would influence our downstream ecological inference. This represents one of the first such sensitivity analyses to compare multiple sequencing runs and bioinformatic platforms in an ecologically robust manner. To assess parameter sensitivity, we considered the effects of quality filtering parameters on the following outcomes: (1) number of reads remaining at each step of the pipeline, (2) estimated alpha diversity, and (3) estimated beta diversity, using a subset of 16S sequences as a test case.

We varied quality filtering parameters for the reverse reads while holding them constant for the forward reads. The reverse reads are typically lower in quality, and thus represent the limiting factor for read retention in paired-end marker gene sequence analyses. Specifically, we tested the effects of the following quality filtering parameters used by DADA2:

- ***$truncLen_R$***: Reverse reads that do not meet or exceed *$truncLen_R$* in length will be discarded. Reverse reads that exceed *$truncLen_R$* will be truncated to *$truncLen_R$*.
- ***$maxEE_R$***: After truncation, reverse reads with higher than *$maxEE_R$* expected errors will be discarded. Expected errors are calculated from the nominal definition of the quality score: $EE = \sum_{l=1}^{L} 10^{-Q/10}$, where *l* is a base position index extending to the length of the sequence, *L*.

The code used to conduct this sensitivity analysis is available in the Supporting Information (Supplementary File 1.2). In summary, the sensitivity analysis evaluates variation in our benchmark metrics with respect to variation in parameter values. It does this by randomly selecting 10 samples from each of the 20 available 16S sequencing runs as of January 2021, and processes these samples on a sequencing-run basis through the 16S pipeline under a variety of quality filtering parameter combinations. The *$truncLen_R$* parameter was assigned values of 170, 220, and 250 base pairs (bp), representing short, medium, and long truncation lengths for the 2 × 300 bp reads produced by Illumina MiSeq. The *$maxEE_R$* parameter was

assigned values of 4, 8, and 16 maximum allowable errors as invoked in the core *DADA2* algorithm; the preferred values may vary substantially between sampling locations, soil types, and lab protocols, so these values were intended to cover a wide range of desirable values. Together, these values resulted in 9 parameter combinations. The following processing decisions were held constant over all pipeline iterations: minimum required length of reads after trimming and truncating (*minLen*) was set to 50 bp; forward reads were processed with truncation length of 240 bp ($truncLen_F$) and a maximum of 8 allowable expected errors ($maxEE_F$); all other quality filtering parameters were set to their default values; and all sequence alignment heuristics for *DADA2* were set to their default values.

Alpha diversity was calculated using the Phyloseq function estimate_richness (McMurdie and Holmes 2013) for Shannon diversity and observed richness. Beta diversity was assessed by joining sequencing tables from across all parameter combinations into one combined sequence table without collapsing sequence-length variants and calculating the pairwise Bray-Curtis distance between all versions of all samples. Samples with a sequencing depth below 1000 were removed prior to ordination and permutational analysis of variance (PERMANOVA). PERMANOVA was conducted via the adonis2 function in the *vegan* R package (Oksanen et al. 2020).

Sensitivity analysis results

The 200 selected samples represented 37 terrestrial NEON sites, collected between May 2014 through November 2018. Sequence read retention throughout the

processing pipeline varied across both parameters, though the degree to which they varied depended on the sequencing run. As expected, higher values of $maxEE_R$ resulted in greater read retention at the quality filtering step (Fig. 1.4). Differences in read retention between sequencing runs could be explained by differences in the quality scores of the reads from each sequencing run. For example, sequencing run BDNB6, whose read retention is relatively sensitive to $maxEE_R$, accumulates more expected errors across its read length than sequencing run BFDG8, whose read retention is relatively insensitive to $maxEE_R$ (Fig. 1.S2.) Read retention was relatively insensitive to $truncLen_R$ except when reverse reads were truncated to 170 bp—this would cause a large drop-off in the pair-merging step, likely due to insufficient overlap between forward and reverse reads (Fig. 1.4). Overall, we found that a moderate value for $truncLen_R$ (220 bp) led to the highest rates of read retention.

The alpha diversity metrics used in the sensitivity analysis were ASV Shannon diversity and observed ASV richness. Shannon diversity (ANOVA, $P < 0.001$; Table 1.1) and observed richness ($P < 0.001$; Table 1.2) were both sensitive to variation in $truncLen_R$, and the effect of $truncLen_R$ varied across sequencing runs ($P < 0.001$; Table 1.1; Table 1.2; diagnostic plots for ANOVA in Fig. 1.S3 and Fig. 1.S4). Consistent with our finding that read retention was highest for moderate values of $truncLen_R$, we also found the highest estimates of Shannon diversity and observed richness when $truncLen_R$ was 220 bp (Fig. 1.5). In contrast, $maxEE_R$ had no significant effects on Shannon diversity ($P = 0.242$; Fig. 1.5; Table 1.1) or observed richness ($P = 0.151$; Fig. 1.S5; Table 1.2). Although $maxEE_R$ does affect read

retention at the quality filtering stage for some sequencing runs, our results suggest that for the NEON 16S sequences in general, differences in $maxEE_R$ have relatively inconsequential effects on estimates of soil microbial alpha diversity. However, researchers who extend this pipeline to other datasets, such as the NEON ITS sequences, should conduct a similar sensitivity analysis before proceeding to make ecological inferences about the processed data.

Variation in the parameters resulted in a small degree of variation in inferred community composition. $truncLen_R$ had a significant effect on Bray-Curtis dissimilarity (PERMANOVA with 999 permutations, $P < 0.001$, $R^2 = 0.012$), while $maxEE_R$ had no significant effect ($P = 1.000$, $R^2 = 3 \times 10^{-5}$; Table 1.3). Upon inspection, setting $truncLen_R = 170$ produced communities with significantly less group dispersion (variance) than at higher values of $truncLen_R$ (*betadisper* multivariate test for homogeneity of group dispersions in the *vegan* R package, $P < 0.001$; Fig. 1.S6). To confirm that the significant effect of $truncLen_R$ in the PERMANOVA analysis was attributable to differences in group means rather than differences in group dispersions, PERMANOVA was repeated on the dataset after removal of communities produced with $truncLen_R = 170$. Within this subset, the sensitivities of Bray-Curtis dissimilarity to the quality filtering parameter remained largely the same: $truncLen_R$ had a significant effect (PERMANOVA with 999 permutations, $P < 0.001$, $R^2 = 6.8 \times 10^{-4}$) while $maxEE_R$ did not ($P = 0.697$, $R^2 = 5 \times 10^{-5}$; Table 1.S1). Since group dispersion did not vary significantly between the remaining values of $truncLen_R$ in the subset (*betadisper*, $P = 0.388$), the results of this

repeated analysis confirm that community composition of the NEON 16S sequences is sensitive to $truncLen_R$. Nevertheless, the amount of variation explained by $truncLen_R$ (R^2 = 0.012) is small compared to that explained by soil sample ID (R^2 = 0.771; Fig. 1.6, Table 1.S2), suggesting that variation in quality filtering parameters is unlikely to obscure real between-sample variation in community composition.

Based on these results, we advise against varying $truncLen_R$ between sequencing runs, as it may lead to inconsistent standards of ecological inference across datasets consisting of samples from multiple runs. However, varying $maxEE_R$ to suit the overall quality of each sequencing run may be appropriate depending on the metrics of interest. The sensitivity analysis framework above can be generalized to test the robustness of ecological inference to other processing decisions, such as paired-end read merging, *DADA2* sequence alignment heuristics, and incorporation of data from different sequencing runs or sequencing platforms.

Lessons Learned and Future Directions

Technical challenges associated with processing large-scale marker gene sequence datasets

There is significant technical variation among NEON sequencing runs that inevitably impacts subsequent bioinformatic processing. While most NEON sequencing runs produced high-quality data, certain runs generated substantially fewer sequences that passed quality filtering (Fig. 1.4). Accordingly, the quality filtering parameters recommended here necessarily represent a compromise, given the

goal of compiling dozens of sequencing runs generated by different sequencing centers over many years. A critical step of the pipeline proposed here requires that the same portion of rRNA is used to denoise ASV across all sequencing runs; for compatibility, we recommend future studies employ identical primers for ease of cross-study comparison. Variation across Illumina sequencing runs necessarily generates variation in the behavior of quality filtering parameters employed in *DADA2*; however, these parameters must be standardized across sequencing runs in order to join ASV tables and to cluster artefactual sequence-length variants. As Illumina sequencing chemistry changes and new platforms emerge, we expect that the filtering steps employed here will need to be updated. Notably, because reverse reads were consistently low-quality across ITS sequencing runs, paired-end ITS read processing is not explicitly supported by our pipeline. Low-quality ITS reverse reads are typical of Illumina MiSeq data. While the 250-bp unmerged forward read sequences may potentially bias against certain fungal taxa (Truong et al. 2019), the extent of this bias is likely small (Nguyen et al. 2015, Pauvert et al. 2019).

NEON's continental-scale sample network captures a remarkably broad phylogenetic range of microbial taxa. Accordingly, analyzing the effect of geographic and ecological distance among samples depends on the taxonomic scale of investigation. Perhaps unsurprisingly, certain samples derived from distinct habitats share no ASV in common, creating disjunctions in community dissimilarity matrices that can complicate distance-based analyses, such as ordination. Clustering ASV at the OTU level, however defined (e.g., 97-98.5% sequence similarity), can reduce

these statistical disjunctions and allow for more meaningful continental-scale analyses of community dissimilarity. Finally, access to sufficient computing resources represents a challenge inherent to datasets of this size. Although recent R packages such as SpeedySeq (McLaren 2020) can expedite some commands run with the popular *Phyloseq* package (McMurdie and Holmes 2013), we expect more future developments.

Finally, rapid advances in high-throughput sequencing technologies may allow for the generation of long reads that span the entire ITS1, ITS2, and 18S region for fungi, and the entire 16S for bacteria, thereby allowing enhanced resolution of fine scale taxonomic boundaries and more accurate phylogenetic placement. In order to ensure compatibility with the extant NEON data presented here, large portions of read overlap with the existing ITS1 and 16S V3-V4 regions analyzed here will be necessary for sufficient sequence alignment and ASV and OTU clustering.

Future directions for NEON-enabled microbial ecology

Spatial and temporal dynamics of soil microbial communities. NEON's extensive soil sampling network fulfills a pressing need for standardized microbial data in advancing research on the spatial and temporal dynamics of soil microbial communities. Soil microbial communities are known to display rapid turnover in space (Franklin and Mills 2003, Nemergut et al. 2013) and time (Ferrenberg et al. 2013, Lauber et al. 2013, Shade et al. 2013). However, there has historically been a trade-off between spatial and temporal sampling intensity, limiting the

generalizability of biogeographic patterns to unsampled regions or timespans. Spatially nested sampling designs like that of NEON's soil data products allow researchers to quantify the spatial scaling of microbial diversity from soil cores to continents and to identify its drivers at each scale (Talbot et al. 2014). Furthermore, because NEON is committed to multiple decades of data collection, its steady accumulation of microbial sequence data will facilitate research on the scaling of microbial diversity over time, from intra-annual to decadal scales. In combination with other NEON data products—such as climate, soil physical and chemical properties, and vegetation cover—the soil microbe data will also help to elucidate the fundamental drivers of temporal scaling (Guo et al. 2019).

NEON soil microbe data may also be informative in comparing the rates of scaling across intersecting gradients of spatial, temporal, and taxonomic scales. For example, a recent study suggests that intra-annual variation in soil fungal communities is comparable to that occurring over hundreds to thousands of kilometers of space (Averill et al. 2019). One of the implications of this rapid spatial and temporal turnover is that sample-pairwise compositional similarity may drop off rapidly, creating a technical challenge for dissimilarity-based analyses when two samples in the dataset have no taxa in common. This challenge can be partially addressed by shifting the unit of taxonomic analysis, e.g., from ASVs to OTUs, or by using phylogenetic measures of beta-diversity (Lozupone and Knight 2005). Future studies, then, may also explore how turnover in soil microbial communities interacts with taxonomic scale.

In addition, the NEON sampling network provides a unique opportunity to observe lags in response time between abiotic variables and changes in soil microbes and to detect the influence of history on community structure. In traditional ecosystem modeling, microbial communities have been assumed to be resilient to disturbance and to return quickly to a state of equilibrium (Allison and Martiny 2008). However, a growing body of evidence suggests that this is not the case; microbes experience legacy effects from historical precipitation regimes (Evans and Wallenstein 2012), plant communities (Elgersma et al. 2011), and wildfires (Qin et al. 2020) that may last several years after the change from prior conditions. In the case of historical contingencies such as priority effects, the equilibrium state may also change (Hawkes and Keitt 2015). By comparing the temporal dynamics of soil microbial communities with other variables recorded in NEON data products, we can ask how long it takes for soil microbes to react to environmental shifts (e.g., in mean precipitation, in mean temperature), how resilient the microbial constituents are to this change, and whether historical events modify the equilibrium states of the microbial community. Furthermore, the NEON sampling network allows researchers to ask questions about synchrony in the spatio-temporal dynamics of microbial communities, and to link these dynamics to stability in microbe-mediated ecosystem processes (Hall et al. 2018, Wang et al. 2019).

Finally, the breadth of NEON soil microbe data now allows researchers to compare the biogeographic patterns and processes of soil microbial communities with those of plants and animals, for which abundance data are also being collected at

NEON sites. This can be used to test the generalizability of macroecological patterns (Xu et al. 2020, Dickey et al. 2021) or temporal patterns (Shade et al. 2013, Guo et al. 2019) that have traditionally been developed for macroorganisms. It may also be leveraged to understand whether the assembly "rules" that govern the distributions of macroorganisms apply equally as well to microbial community assembly across multiple nested spatial scales.

Microbial community composition and ecosystem processes. Soil microbes are important regulators of ecosystem processes such as nitrogen and carbon cycling (Cavicchioli et al. 2019). However, climate-soil feedbacks remain critical sources of uncertainty in ecosystem models (Wieder et al. 2015). This uncertainty arises due to poor understanding of soil microbial diversity and function, and the challenge of applying macroorganismal functional concepts to microorganisms (Escalas et al. 2019). There has been some progress, however, in improving ecosystem models by assigning microbial species into functional groups (Fry et al. 2019, Sulman et al. 2019). To enrich our understanding of microbial communities from marker genes to ecosystem processes, future studies could take advantage of NEON data to investigate the generalizability of the link between microbial taxonomic or functional composition and ecosystem function or services (Box 1). For example, Werbin et al. (2021) present a pipeline for coupling soil shotgun metagenomic sequence data with NEON nitrogen cycling data at each site; this pipeline complements *neonMicrobe*, allowing for an investigation into the linkages between microbial composition and

function. From a biogeochemical modeling perspective, the standardized continental scale sampling provides unique opportunities to couple plant and microbial dynamics with gross primary productivity or soil organic matter dynamics (Fry et al. 2019). Moreover, forecasting the distribution of microbial communities under future climatic regimes is similarly enabled by standardized continental scale sampling efforts coupled with fine-scale soil, plant, and climatic metadata.

Expanded use of NEON samples and infrastructure. For research needs that are not precisely met by existing NEON data streams, NEON also offers two programs to help researchers leverage NEON sample collections and field infrastructure. The NEON Biorepository Data Portal allows researchers to request access to biological samples, including frozen subsamples of the soil and extracted DNA used to generate the soil microbe marker gene sequence data product, in order to conduct their own laboratory analyses. Researchers interested in using a different sequencing protocol or conducting a functional assay, for example, may take advantage of this program. As another example, a researcher interested in studying food webs may request biorepository samples to identify arthropods in pitfall traps beyond beetles—in addition to NEON soil microbe amplicon, abiotic, and metagenomic datasets. Furthermore, the NEON Assignable Assets Program allows researchers to request the use of specialized NEON data collection infrastructure for their own research, temporarily adding to the sampling design of NEON field sites. To return to our example, the researcher interested in studying food webs may conduct some of their

own on-site sampling using similar designs to survey nematodes, via the Assignable Assets Program's Observational Sampling Infrastructure. The flexibility built into the NEON data stream infrastructure greatly expands the potential to accommodate future research directions that were not part of the original design.

Going beyond NEON data. Using networks to synthesize ecological and environmental research offers promising new avenues for scientists to create holistic representations of natural processes—particularly in fields that account for complex, large-scale phenomena such as biogeography (Schrodt et al. 2019). As a nationwide monitoring network, NEON provides broad coverage for the collection of ecological and environmental data; however, it is limited in its ability to provide sites for field experiments. The US Long-Term Ecological Research network (LTER) provides complementary infrastructure for experimental studies in a variety of ecosystems, and may help to elucidate the processes driving patterns observed in NEON data (Jones et al. 2020). At the time of writing, twelve NEON sites are co-located with LTER sites.

Although scientific research aims to explain natural processes, it is also an inherently social process in which tacit, socially-transferred knowledge is especially important for the extension of methods to novel or synthetic contexts (Collins 1974). The unprecedented spatial and temporal scales of the soil microbe marker gene sequence data provided by NEON represent such a context to develop best practices for team science. Methodological and epistemological challenges involved in using

this data led the authors of this paper to recognize the necessity of having a team of collaborators to validate methods and test results before formally embedding them into a standard algorithmic process. While there is some research on the social and technical factors that allow for effective team science (Rhoten 2003, Oliver et al. 2018), there is room to consider how to best foster collaborations that can synthesize the wide variety of NEON data products to address interdisciplinary problems (e.g., Nagy et al. 2021). Interdisciplinary collaborations have been identified as avenues for fruitful and novel research in ecology and the environment as discussed above, but especially for understanding complex socio-environmental issues (Palmer et al. 2016). They also provide opportunities for graduate students in ecology to realize and develop the unique expertise they bring to the team (Giorgio et al. 2020). Factors that may have contributed to our ability to complete this project include the diversity of expertise across our team members, which included soil microbial ecologists, molecular biologists, and statisticians, as well as a diversity of career stages that allowed graduate student members to receive real-time feedback from an informal community of mentors. One of the main challenges to our project was the inability for authors to hold meetings in person after the Summit—a challenge which was exacerbated by the COVID 19 pandemic, and which previous studies have identified as a potential hindrance to information sharing (Rhoten 2003). Future studies should seek to understand what types of social and technical configurations facilitate or hinder data-intensive, interdisciplinary team science, and how data-sharing centers

like NEON can take advantage of these findings to make their data more accessible and useful across diverse research contexts.

Conclusions

We present *neonMicrobe*, a processing pipeline for the R statistical computing environment that streamlines access to NEON microbe marker gene sequence data. Our approach adapts state-of-the-art sequence processing pipelines for current NEON marker gene sequencing approaches. We have validated the efficacy of recommended quality filtering parameters in our pipeline. The collaborative effort represented here speaks to the utility of open science, and our publicly available data wrangling tools can be adopted for user-specific applications. We expect this community resource will expedite NEON-enabled science and herald a new era of continental-scale analysis for microbial community dynamics.

Box 1: Future Questions for NEON-Enabled Microbial Ecology

Spatial and temporal dynamics

- How do soil microbial communities vary across spatial scales (sites, ecoregions, continents) and temporal scales (seasonal, annual, decadal)? What are the important drivers?
- How does microbial diversity scale over space, time, and taxonomic resolution?
- What are the patterns of temporal or spatial autocorrelation in soil microbial communities?
- How do macroecological and biogeographical patterns of soil microbial communities vary across spatial scales? Do they follow the same "rules" as for macro-organisms?

Ecosystem processes

- How can we effectively include microbial communities in ecosystem and earth systems models?
- Can we predict ecosystem functions and services (e.g., C flux) from microbial taxonomic or functional composition?
- How can we forecast future changes in these processes?

Going beyond NEON data

- How can we design future studies to take advantage of and complement NEON observatory, biorepository, and assignable assets data?
- What are some best practices for fostering interdisciplinary team science,

synthesizing a variety of NEON data products to answer complex ecological problems?

Table 1.1. ANOVA results for the sensitivity of ASV Shannon diversity to quality filtering parameters $truncLen_R$ and $maxEE_R$.

Covariate	***df***	**Sum Sq**	**Mean Sq**	***F***	**P(>*F*)**
$truncLen_R$	1	81.616	81.616	403.053	<0.001
$maxEE_R$	1	0.277	0.277	1.369	0.242
runID	19	321.151	16.903	83.473	<0.001
$truncLen_R \times$ runID	19	36.442	1.918	9.472	<0.001
$maxEE_R \times$ runID	19	0.132	0.007	0.034	1.000
Residuals	1740	352.339	0.202		

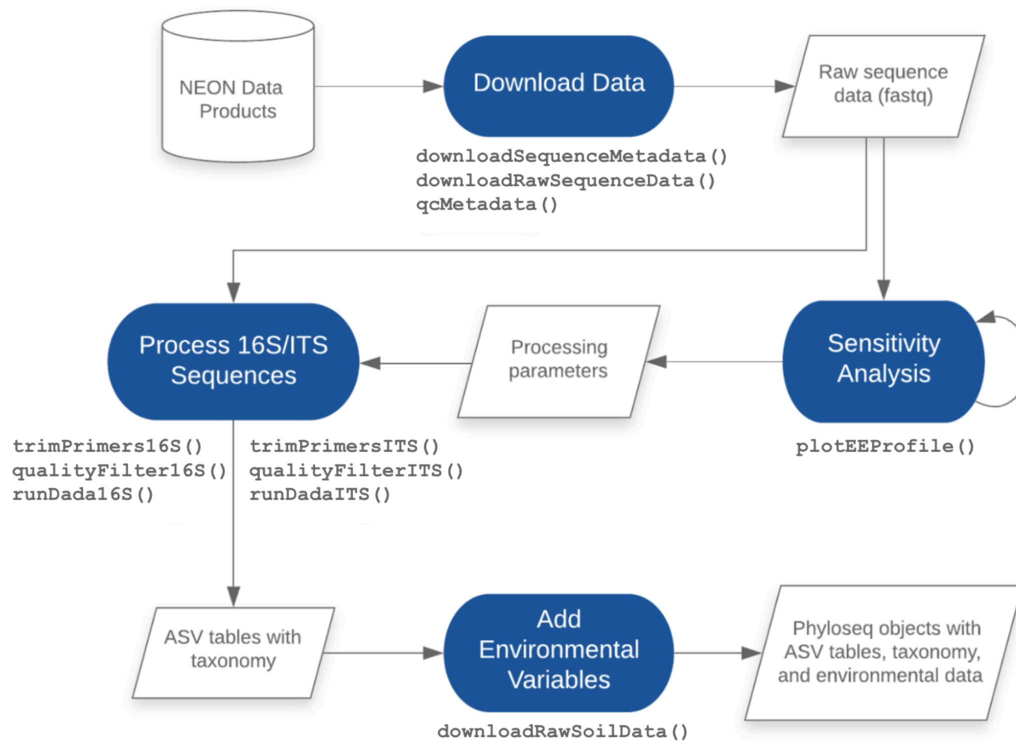
Table 1.2. ANOVA results for the sensitivity of observed ASV richness to quality filtering parameters $truncLen_R$ and $maxEE_R$.

Covariate	*df*	Sum Sq	Mean Sq	*F*	P(>*F*)
$truncLen_R$	1	14.598	14.598	66.805	<0.001
$maxEE_R$	1	0.452	0.452	2.069	0.151
runID	19	485.837	25.570	117.020	<0.001
$truncLen_R \times$ runID	19	48.997	2.579	11.802	<0.001
$maxEE_R \times$ runID	19	0.270	0.014	0.065	1.000
Residuals	1740	380.213	0.219		

Table 1.3. PERMANOVA results for the sensitivity of Bray-Curtis dissimilarity to quality filtering parameters $truncLen_R$ and $maxEE_R$. Communities were permuted 999 times within sample IDs, i.e. each community was compared against other communities produced from the same sample. For a PERMANOVA analysis that includes sample ID as a permuted variable, see Table 1.S2.

Covariate	***df***	**Sum Sq**	R^2	F	**P(>*F*)**
$truncLen_R$	1	9.61	0.01235	21.373	<0.001
$maxEE_R$	1	0.02	0.00003	0.050	1.000
Residuals	1709	768.75	0.98762		

Figure 1.1. Overview of the *neonMicrobe* package's marker gene sequence processing pipeline. Blue rounded rectangles correspond to main vignettes in the R package; white parallelograms represent locally stored data or parameters. Fixed-width text below blue-rounded rectangles denotes the *neonMicrobe* functions belonging to each vignette. The processing pipeline begins with the download of NEON data products containing microbe marker gene sequences. The raw sequence data can then be processed into ASV tables with taxonomy using either default or custom processing parameters. The sensitivity analysis can be used to determine the range of custom processing parameters, if any, that can be used across an analysis without creating artifacts in ecological metrics. Finally, the ASV tables can be joined with environmental variables and sampling information. The output data is structured in the form of one or more *Phyloseq* objects. See Fig. 1.S1 for finer-level details on the functions and parameters in the pipeline.



NEON Data Products
Download Data
downloadSequenceMetadata()
downloadRawSequenceData()
qcMetadata()
Raw sequence data (fastq)
Process 16S/ITS Sequences
Processing parameters
Sensitivity Analysis
plotEEProfile()
trimPrimers16S()
qualityFilter16S()
runDada16S()
trimPrimersITS()
qualityFilterITS()
runDadaITS()
ASV tables with taxonomy
Add Environmental Variables
downloadRawSoilData()
Phyloseq objects with ASV tables, taxonomy, and environmental data

Figure 1.2. Default organization of input data for the *neonMicrobe* R package. The tree structure in the upper-left represents the data directory structure constructed within the project root directory. Red dotted lines represent explicit linkages between NEON data products via shared data fields. (a) Sequence metadata is downloaded from NEON data product DP1.10108.001 (Soil microbe marker gene sequences) using the downloadSequenceMetadata function. (b) Raw microbe marker gene sequence data is downloaded from NEON based on the sequence metadata using the downloadRawSequenceData function. (c) Soil physical and chemical data is downloaded from NEON data product DP1.10086.001 using the downloadSoilData function. (d) Taxonomic reference datasets (e.g., SILVA, UNITE) are added separately by the user.

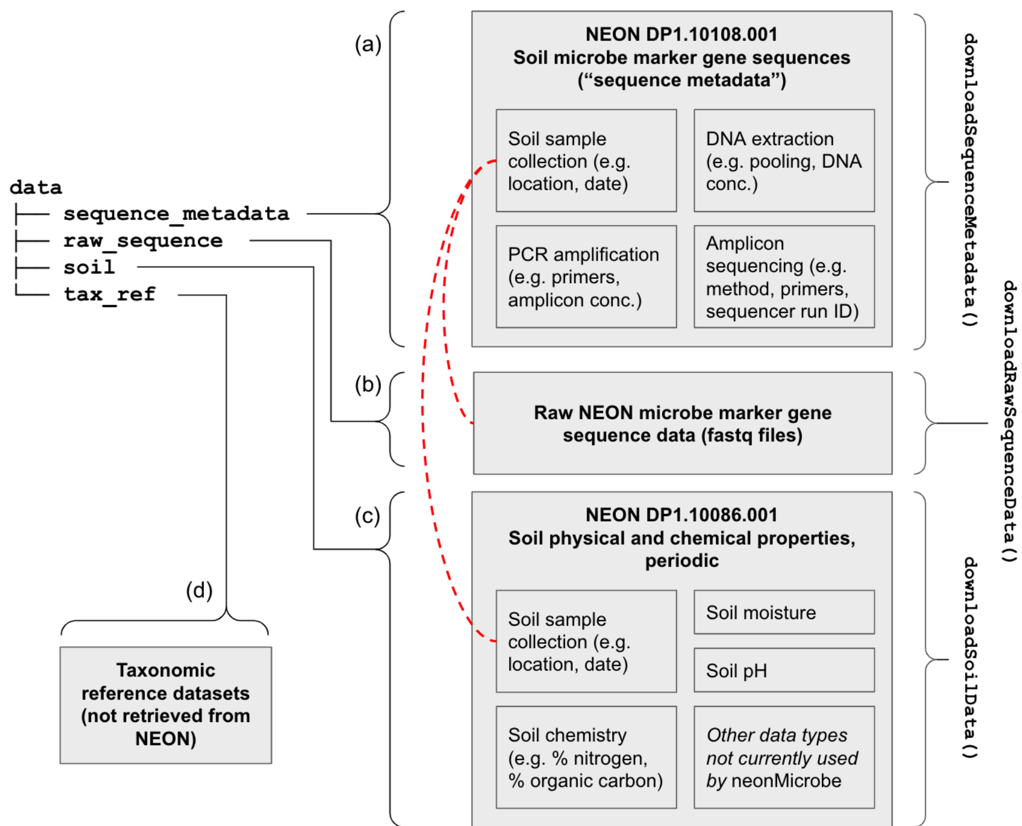
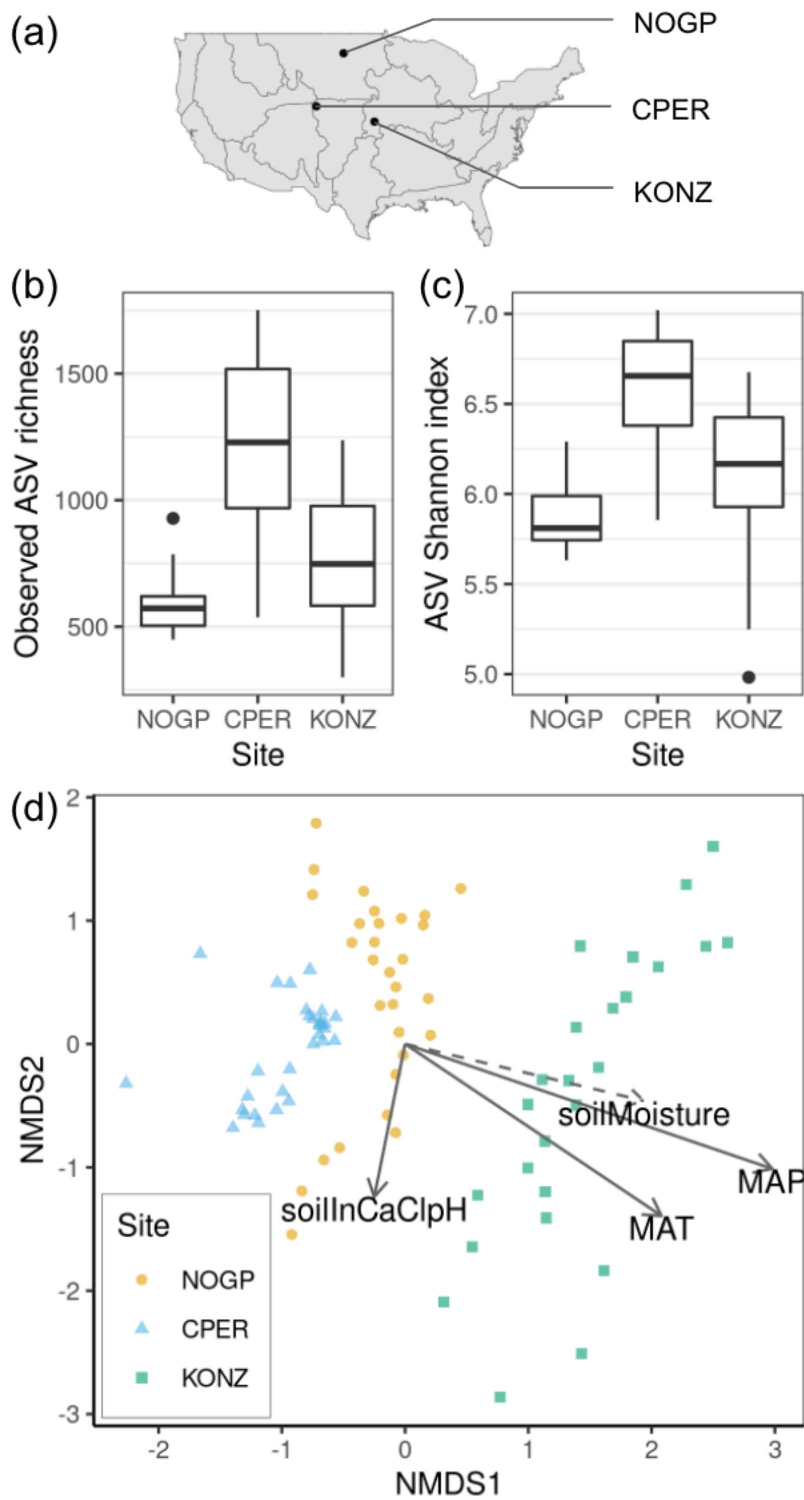
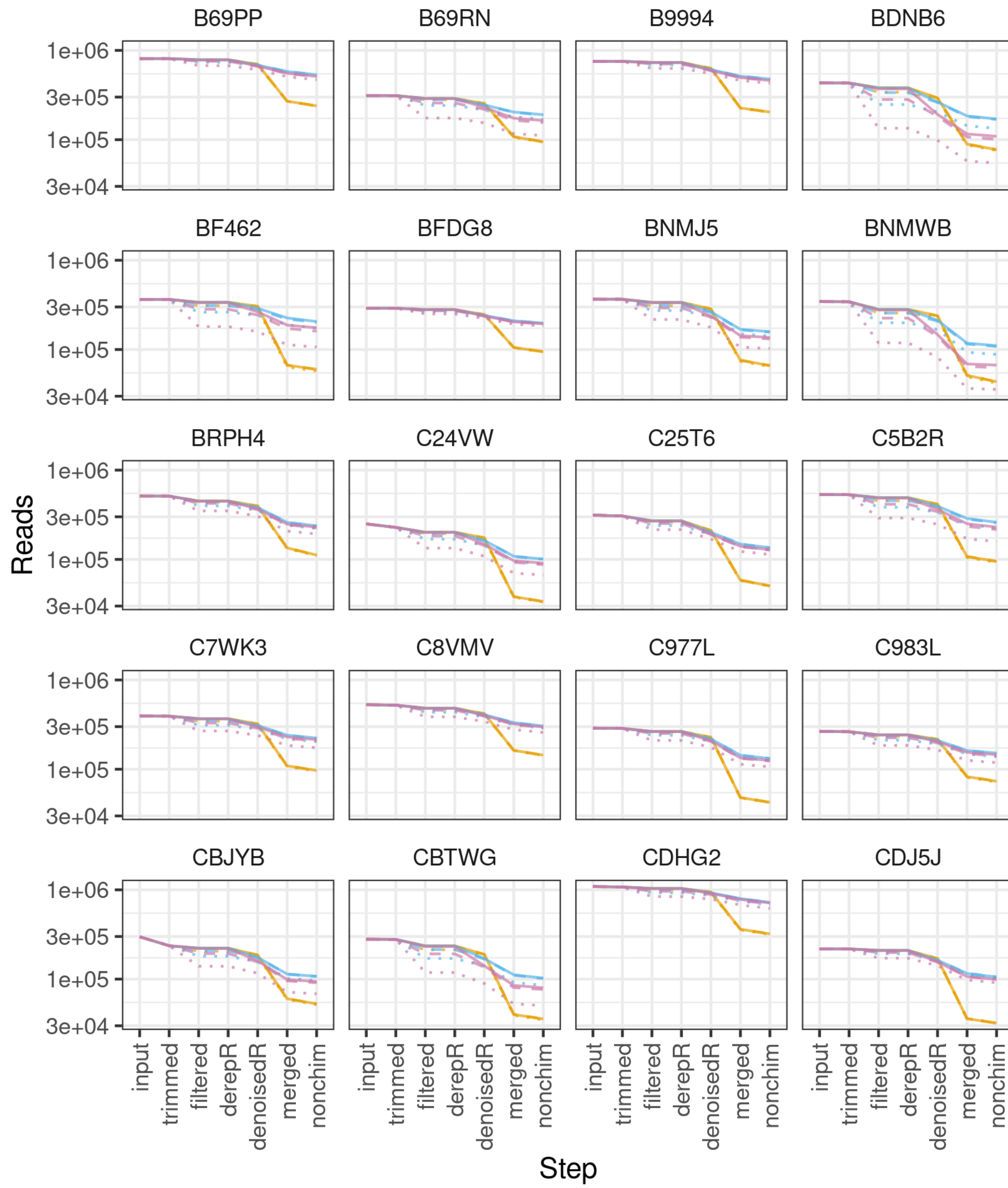

data
├── sequence_metadata
├── raw_sequence
├── soil
└── tax_ref
(a)
(b)
(c)
(d)
NEON DP1.10108.001
Soil microbe marker gene sequences
("sequence metadata")
Soil sample collection (e.g. location, date)
DNA extraction (e.g. pooling, DNA conc.)
PCR amplification (e.g. primers, amplicon conc.)
Amplicon sequencing (e.g. method, primers, sequencer run ID)
Raw NEON microbe marker gene sequence data (fastq files)
NEON DP1.10086.001
Soil physical and chemical properties, periodic
Soil sample collection (e.g. location, date)
Soil moisture
Soil pH
Soil chemistry (e.g. % nitrogen, % organic carbon)
Other data types not currently used by neonMicrobe
Taxonomic reference datasets (not retrieved from NEON)
downloadSequenceMetadata()
downloadRawSequenceData()
downloadSoilData()

Figure 1.3. Simple plots produced using the outputs of the *neonMicrobe* pipeline for analyses of 16S sequences in soil samples collected from NEON grassland sites. (a) The three sites used for this analysis (Central Plains Experimental Range, CPER; Konza Prairie Biological Station, KONZ; Northern Great Plains Research Laboratory, NOGP) span the Great Plains region and belong to different NEON eco-climatic domains (delineated with borders). Samples were rarefied to a common sequencing depth of 10,000 reads. The sites differ by observed ASV richness (b) and ASV Shannon index (c). (d) Pairwise Bray-Curtis dissimilarities were used for NMDS ordination analysis, and environmental covariates were fitted to the ordination using the envfit function in the vegan R package (Oksanen et al. 2020). Vectors with solid lines denote significant drivers (PERMANOVA, $P \leq 0.05$); the vector with the dashed line denotes an insignificant driver ($P > 0.05$). Abbreviations: MAT, mean annual temperature; MAP, mean annual precipitation; soilMoisture, soil moisture; soilInCaClpH, soil pH measured in calcium chloride solution.



(a)
NOGP
CPER
KONZ
(b)
Observed ASV richness
1500
1000
500
NOGP CPER KONZ
Site
(c)
ASV Shannon index
7.0
6.5
6.0
5.5
5.0
NOGP CPER KONZ
Site
(d)
NMDS2
NMDS1
soilMoisture
MAP
MAT
soilInCaClpH
Site
NOGP
CPER
KONZ

Figure 1.4. Retention of NEON 16S reads throughout the processing pipeline across different values of quality filtering parameters $truncLen_R$ and $maxEE_R$. Read counts are tracked for each available NEON 16S sequencing run at each major step in the processing pipeline: *input*, raw marker gene sequence data from NEON; *trimmed*, after primer trimming; *filtered*, after quality filtering; *derepR*, reverse reads after dereplication; *denoisedR*, after inferring sequences for reverse reads; *merged*, after merging forward and reverse sequences; *nonchim*, after removing chimeras. Within each sequencing run, the read counts for $truncLen_R$ = 170 across different values of $maxEE_R$ largely overlap.



truncLenR 170 220 250
maxEER 4 8 16
B69PP
B69RN
B9994
BDNB6
BF462
BFDG8
BNMJ5
BNMWB
BRPH4
C24VW
C25T6
C5B2R
C7WK3
C8VMV
C977L
C983L
CBJYB
CBTWG
CDHG2
CDJ5J
1e+06
3e+05
1e+05
3e+04
Reads
input
trimmed
filtered
derepR
denoisedR
merged
nonchim
Step

Figure 1.5. Shannon diversity of ASV tables produced from NEON 16S marker gene sequences across different values of quality filtering parameters $truncLen_R$ and $maxEE_R$.

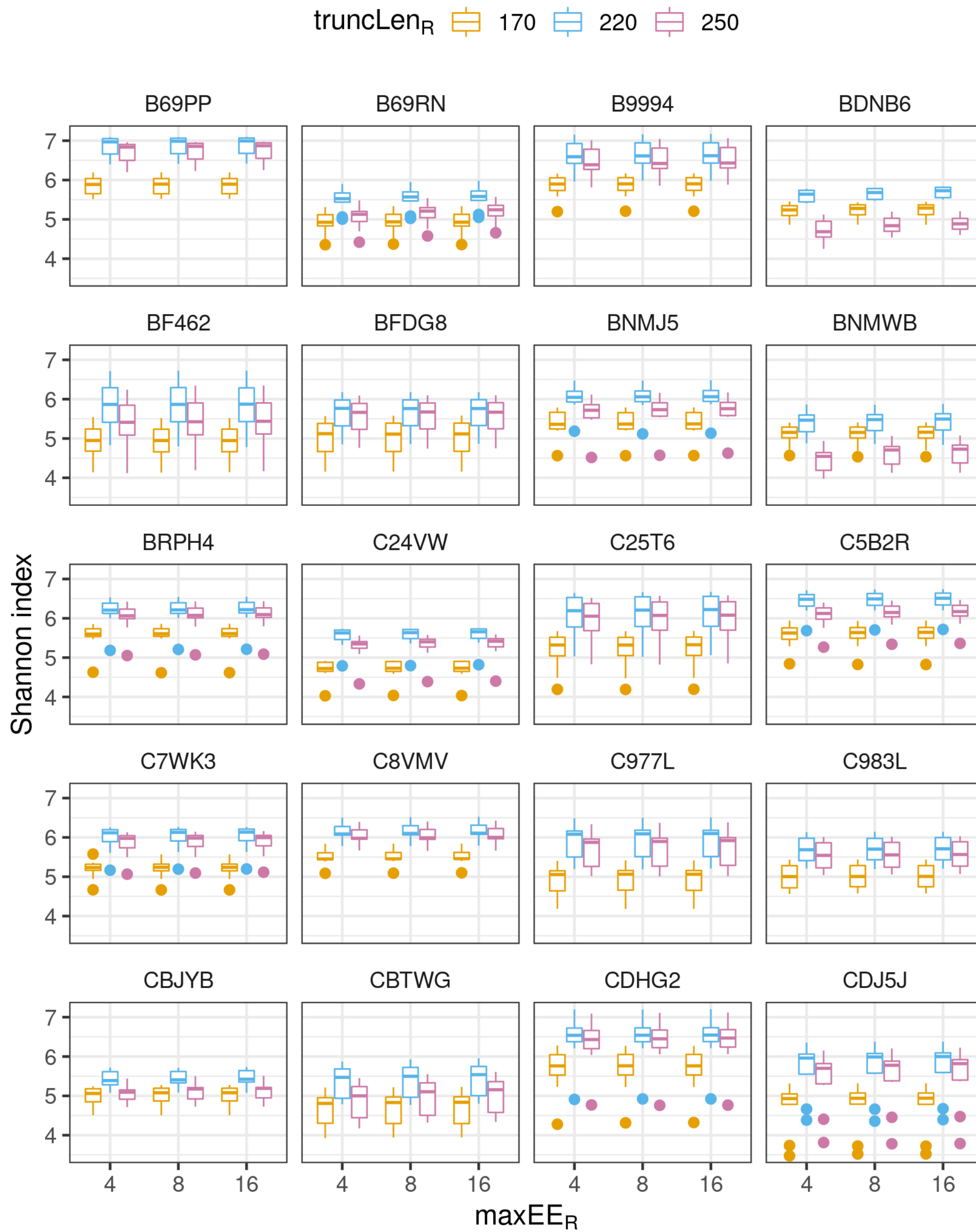


Figure 1.6. Community composition of ASV tables produced by NEON 16S marker gene sequences across different values of quality filtering parameters $truncLen_R$ and $maxEE_R$. Each plot corresponds to an NMDS ordination for a specified sequencing run (x-axis, NMDS1; y-axis, NMDS2). Note that for a given soil sample and value of $truncLen_R$, points corresponding to outcomes across different values of $maxEE_R$ largely overlap. Ellipses are multivariate normal 95% data ellipses for ASV tables produced from the same soil sample. Ordinations were based on Bray-Curtis dissimilarities and plotted using the metaMDS algorithm in the *vegan* R package (Oksanen et al. 2020).

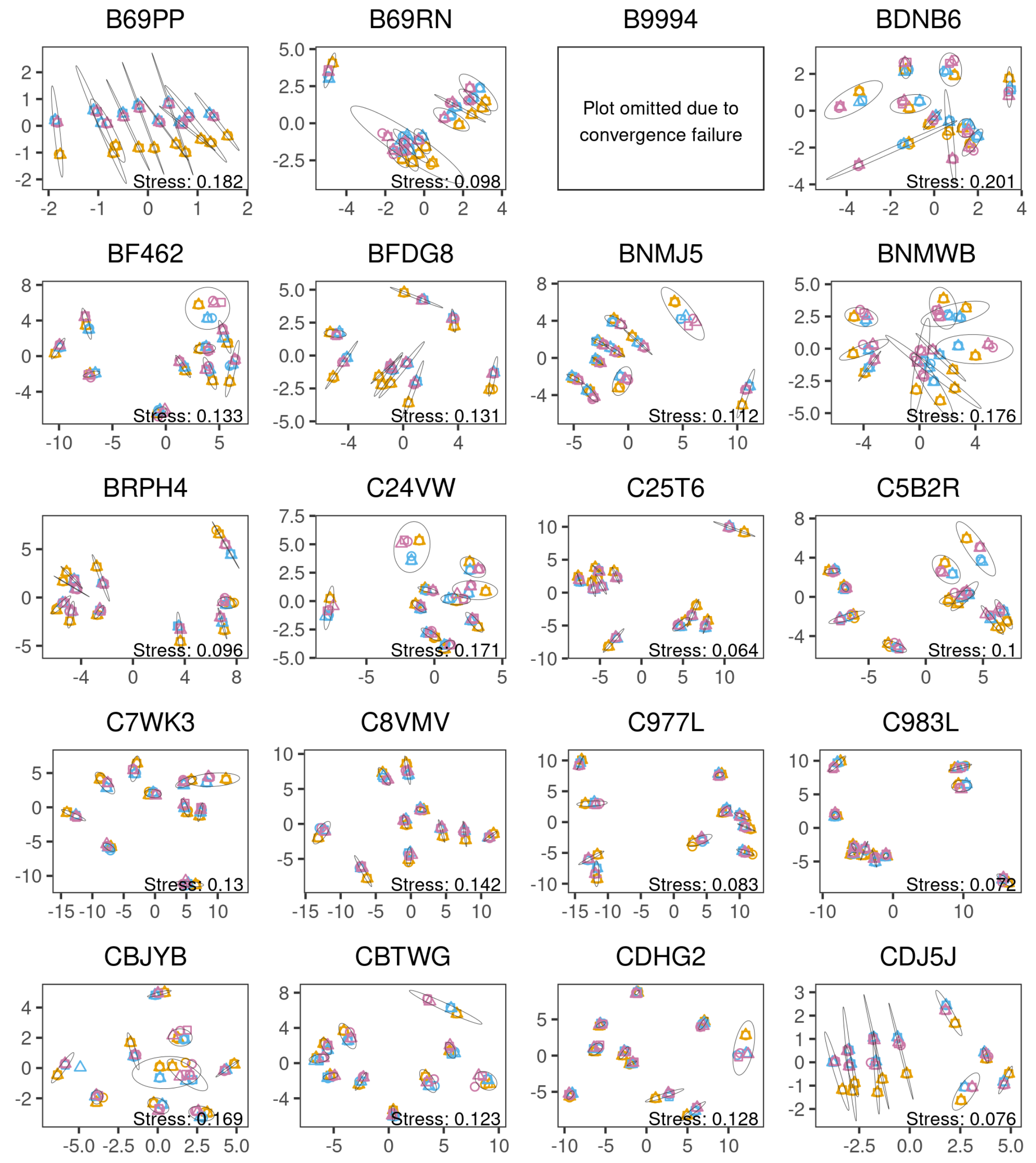

truncLenR 170 220 250
maxEER 4 8 16
B69PP
Stress: 0.182
B69RN
Stress: 0.098
B9994
Plot omitted due to convergence failure
BDNB6
Stress: 0.201
BF462
Stress: 0.133
BFDG8
Stress: 0.131
BNMJ5
Stress: 0.112
BNMWB
Stress: 0.176
BRPH4
Stress: 0.096
C24VW
Stress: 0.171
C25T6
Stress: 0.064
C5B2R
Stress: 0.1
C7WK3
Stress: 0.13
C8VMV
Stress: 0.142
C977L
Stress: 0.083
C983L
Stress: 0.072
CBJYB
Stress: 0.169
CBTWG
Stress: 0.123
CDHG2
Stress: 0.128
CDJ5J
Stress: 0.076

Table 1.S1. Permutational analysis of variance (PERMANOVA) results for the sensitivity of Bray-Curtis dissimilarity to quality filtering parameters $truncLen_R$ and $maxEE_R$, after removal of communities produced with $truncLen_R$ = 170. Samples were permuted 999 times within sampleIDs. PERMANOVA was conducted using the adonis2 function in the *vegan* R package (Oksanen et al. 2020).

	df	Sum Sq	R^2	*F*	P(>*F*)
$truncLen_R$	1	0.37	0.00068	0.8004	**<0.001**
$maxEE_R$	1	0.03	0.00005	0.0604	0.697
Residuals	1175	537.00	0.99927		

Table 1.S2. Permutational analysis of variance (PERMANOVA) results for the sensitivity of Bray-Curtis dissimilarity to quality filtering parameters $truncLen_R$, $maxEE_R$, and *sampleID*. Samples were permuted 999 times without blocking by any grouping variable. PERMANOVA was conducted using the adonis2 function in the *vegan* R package (Oksanen et al. 2020).

	df	Sum Sq	R^2	*F*	P(>*F*)
$truncLen_R$	1	9.61	0.01235	86.0821	**<0.001**
$maxEE_R$	1	0.02	0.00003	0.2013	1.000
sampleID	196	599.77	0.77053	27.3985	**<0.001**
Residuals	1711	168.98	0.21709		

Figure 1.S1. Generalized approach to analyzing NEON microbial sequence data for ecological analyses using the DADA2 pipeline. Begin by determining the scope of the study, in terms of spatial and temporal extent, and determine whether additional studies will be incorporated. When combining studies, the methods and the data itself may differ significantly, and ensuring that data from different studies follow similar methods and are formatted similarly is a critical step that should be taken before combining studies: otherwise, results may not be valid. The first decision point is whether to merge forward and reverse reads when using bidirectional sequencing data. The 16S rRNA gene fragment presents few drawbacks to merging, since this region of the rRNA cistron is highly conserved and the read length does not vary significantly; both of these factors make it beneficial to merge reads and use as much information as possible. The ITS region, on the other hand, is highly variable, which makes it more challenging to correctly merge reads, and some research suggests that using just the forward reads can more accurately reflect fungal taxonomic composition in mock communities (Pauvert et al. 2019). Therefore, users should consider the research goals and objectives of their specific study and determine whether merging reads for ITS data is optimal. Next, low quality data are removed. For NEON 16S data, avoiding excessively permissive or restrictive parameters for maximum error rates (e.g. *maxEE*) and removing short reads using the *truncLen* parameter produced high quality results and retained an acceptable number of reads. For ITS data, we recommend a similar range for the maximum error rate ($2 \leq maxEE \leq 10$), but recommend against using *truncLen*, in order to preserve natural sequence

length variation in ITS sequences. After quality filtering, the reads are denoised using DADA2. The denoised data are then merged if desired, and chimeras are removed. The output ASV data from this pipeline would then be used for taxon classification and downstream ecological analyses.

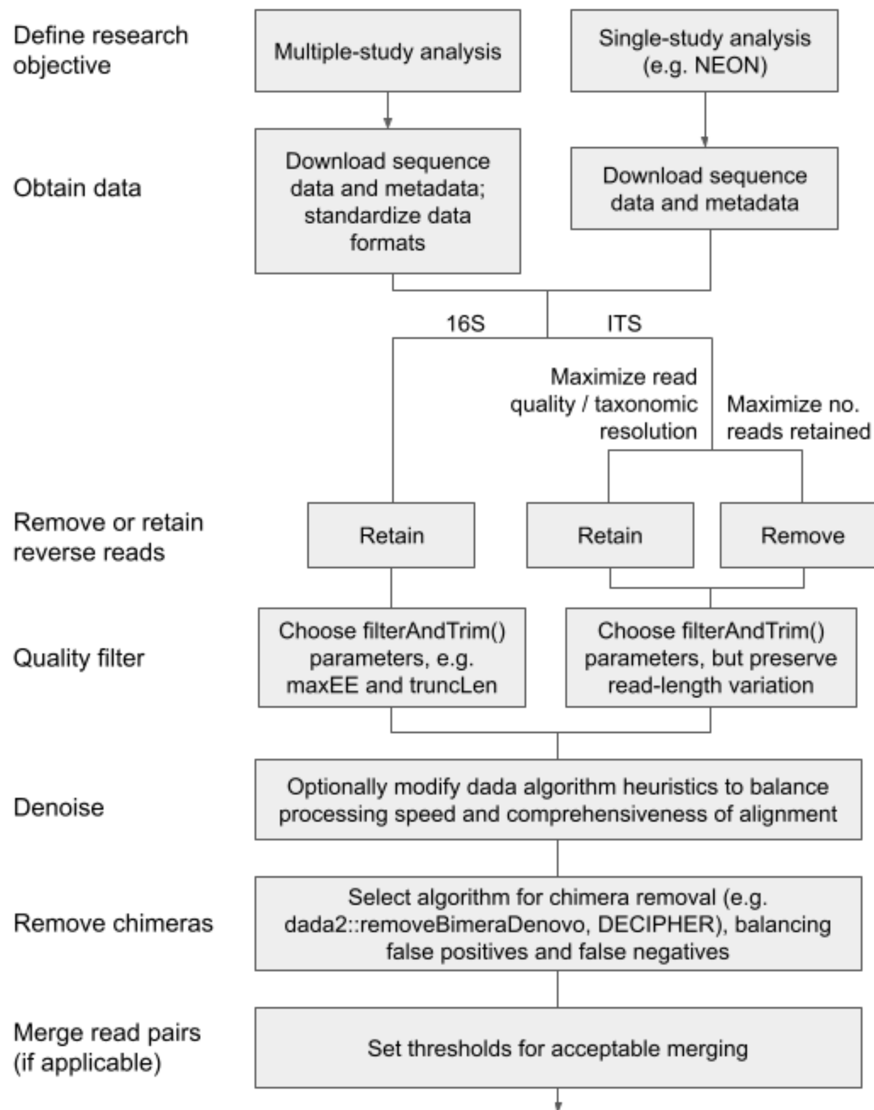

Define research objective
Multiple-study analysis
Single-study analysis (e.g. NEON)
Obtain data
Download sequence data and metadata; standardize data formats
Download sequence data and metadata
16S
ITS
Maximize read quality / taxonomic resolution
Maximize no. reads retained
Remove or retain reverse reads
Retain
Retain
Remove
Quality filter
Choose filterAndTrim() parameters, e.g. maxEE and truncLen
Choose filterAndTrim() parameters, but preserve read-length variation
Denoise
Optionally modify dada algorithm heuristics to balance processing speed and comprehensiveness of alignment
Remove chimeras
Select algorithm for chimera removal (e.g. dada2::removeBimeraDenovo, DECIPHER), balancing false positives and false negatives
Merge read pairs (if applicable)
Set thresholds for acceptable merging

Figure 1.S2. Cumulative expected error profiles for NEON 16S sequencing runs BDNB6 and BFDG8, based on the first 10 samples from each run. Curves represent the number of expected errors for the mean (solid green), median (solid orange), and first and third quartiles (dotted orange) of forward and reverse reads in each sequencing run. Horizontal gray lines indicate possible cut-off points for the maximum number of expected errors (*maxEE*) allowed in each read after truncation to a given length (*truncLen*) at the quality filtering step; these lines do not appear when the cut-off point for *maxEE* = 1 exceeds the maximum range of the y-axis. Note the differing y-axis scales. Expected error profiles were generated using the novel R function plotEEProfile. See Supplementary File 1.3 for cumulative expected error profiles for each sequencing run, and the code used to plot them.

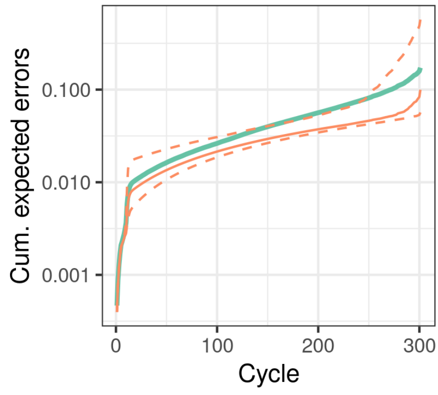
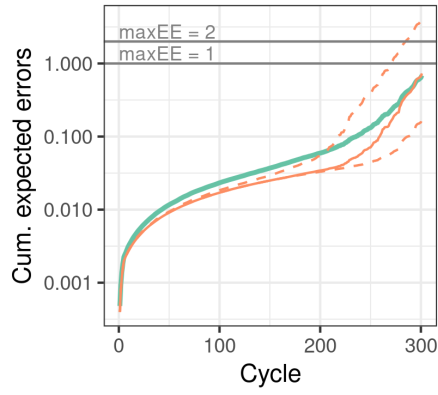
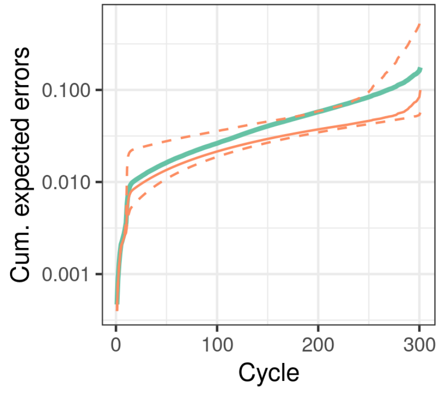
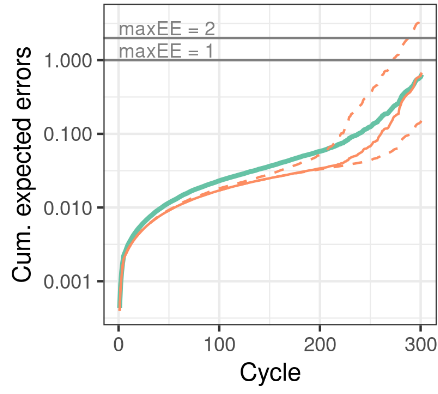

BDNB6, R1
Cum. expected errors
0.100
0.010
0.001
0
100
200
300
Cycle

BDNB6, R2
maxEE = 2
maxEE = 1
Cum. expected errors
1.000
0.100
0.010
0.001
0
100
200
300
Cycle

BFDG8, R1
Cum. expected errors
0.100
0.010
0.001
0
100
200
300
Cycle

BFDG8, R2
maxEE = 2
maxEE = 1
Cum. expected errors
1.000
0.100
0.010
0.001
0
100
200
300
Cycle

Figure 1.S3. Standard ANOVA diagnostic plots for the sensitivity of ASV Shannon diversity to quality filtering parameters. See Table 1.1 for ANOVA results.

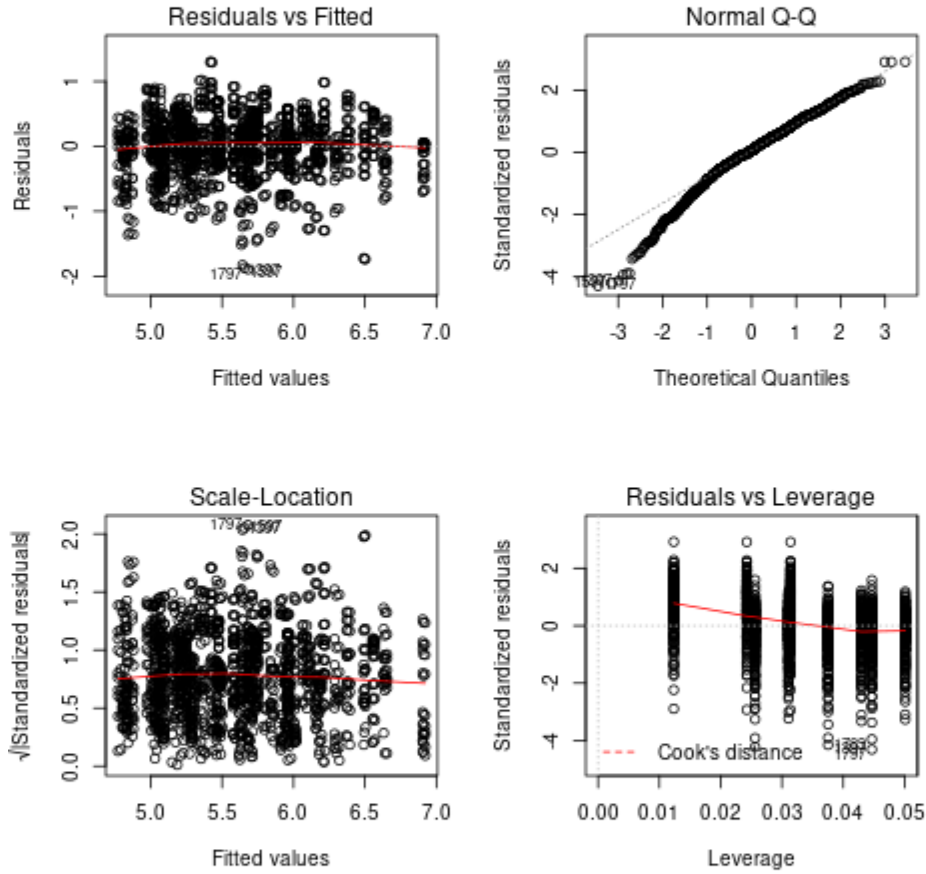


Figure 1.S4. Standard ANOVA diagnostic plots for the sensitivity of observed ASV richness to quality filtering parameters. See Table 1.2 for ANOVA results.

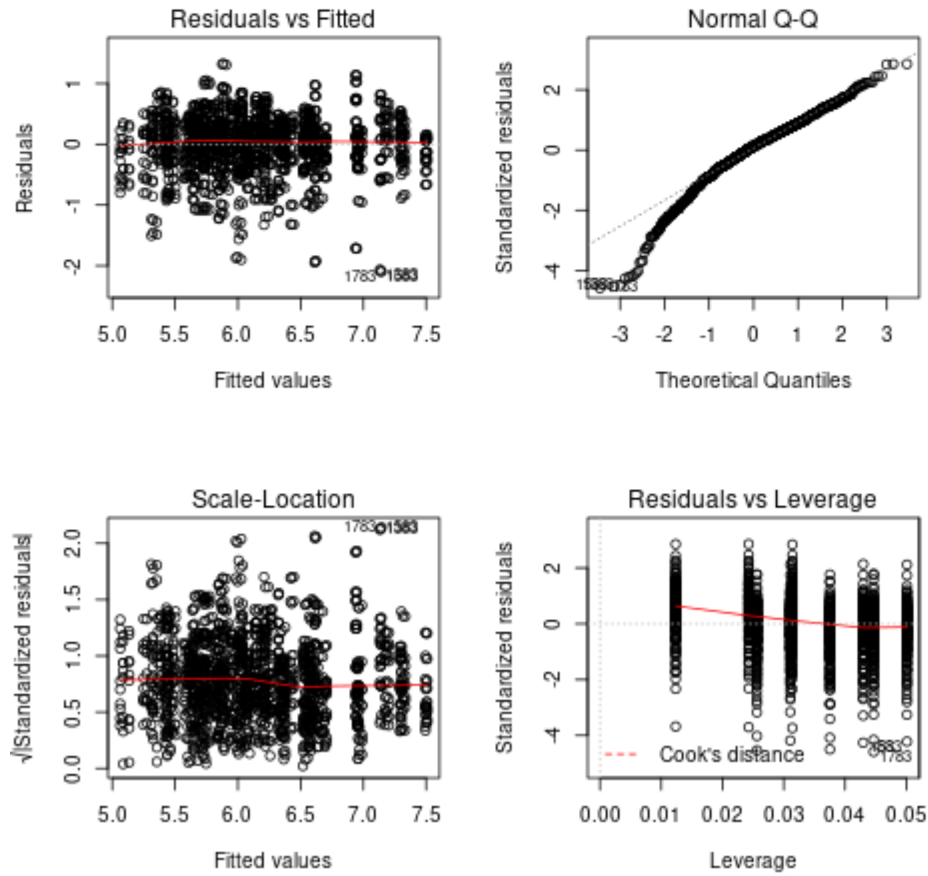


Figure 1.S5. Observed richness of ASV tables produced from NEON 16S marker gene sequences across different values of quality filtering parameters $maxEE_R$ and $truncLen_R$.

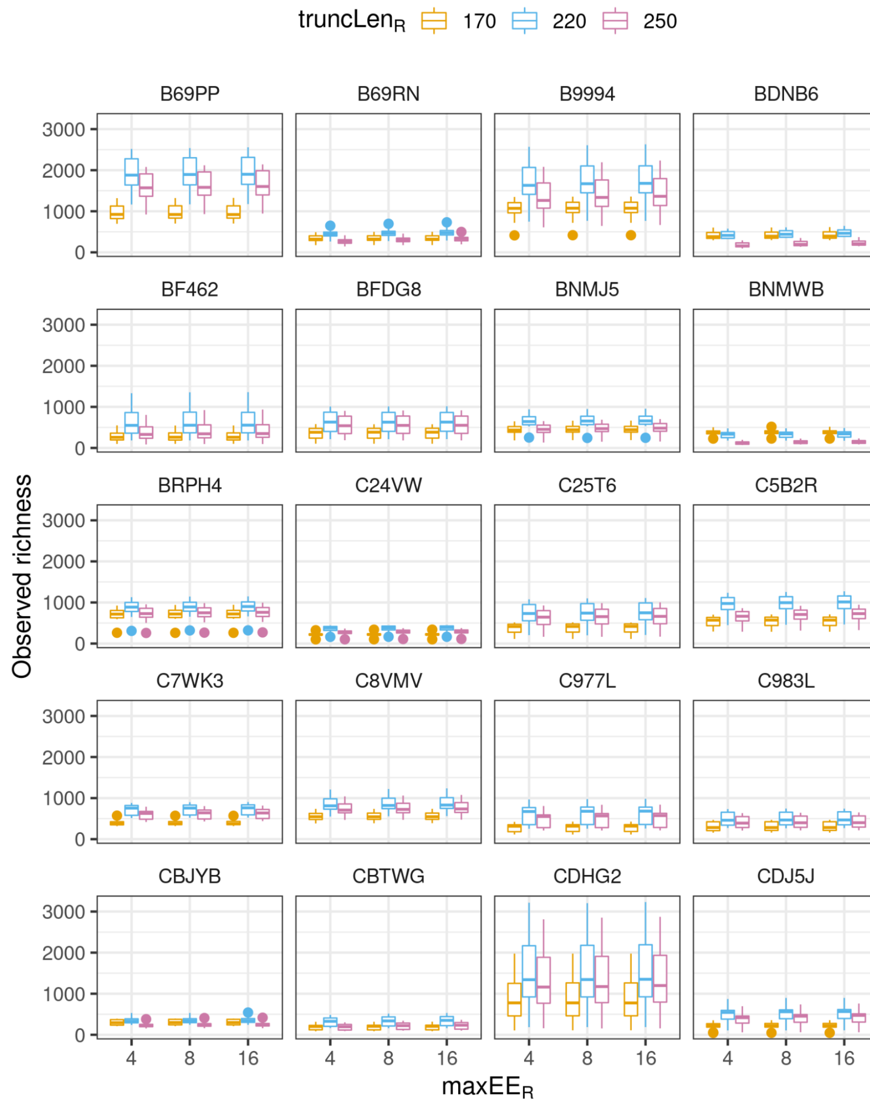
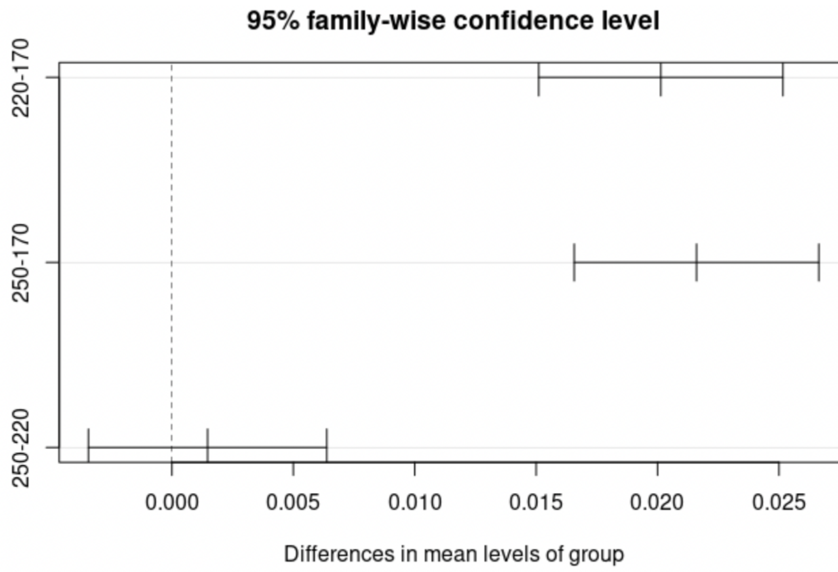


Figure 1.S6. Tukey's HSD test for the homogeneity of group dispersions (variances) for different values of *truncLen*$_R$ (170, 220, 250).



CHAPTER 2

Niche modeling predicts that soil fungi occupy a precarious climate in boreal forests

Clara Qin, Peter Pellitier, Michael Van Nuland, Kabir Peay & Kai Zhu

Published in *Global Ecology and Biogeograpy*, 2023

Introduction

Because soil fungi both regulate and respond to global biogeochemical cycles, elucidating their roles in the trajectory of earth system dynamics requires a precise understanding of their sensitivity to climate change (McGuire & Treseder, 2010; Terrer, Vicca, Hungate, Phillips, & Prentice, 2016; Cavicchioli *et al*., 2019). Climate has been shown to shape soil fungal communities from small-scale climate change experiments (Fernandez *et al*., 2017; Qin *et al*., 2020; Van Nuland *et al*., 2020) to continental-scale observational studies (Talbot *et al*., 2014; Tedersoo *et al*., 2014; Větrovský *et al*., 2019; Steidinger *et al*., 2020). However, research displays considerable variability in the influence of climatic drivers on fungal community assembly (Hendershot, Read, Henning, Sanders, Classen, 2017), especially in comparison to other factors such as host distribution, edaphic conditions, and

stochastic mechanisms (Peay, Kennedy, & Talbot, 2016; Zhou & Ning, 2017).

Some of this variability may be ascribed to differential responses to climate change across environmental space. For example, previous studies have shown that the effect of warming on ectomycorrhizal (EcM) fungal species richness depends on initial temperatures due to the existence of multiple diversity optima along temperature gradients (Steidinger *et al*., 2020), and the relative abundance and activity of soil saprotrophic fungi change abruptly across particular temperature thresholds (Feng *et al.*, 2022). In the absence of this research, one might initially expect that these community properties change in a smooth, linear fashion across climate gradients. Indeed, this reasoning provides the basis for linking novel climates to no-analog communities (Williams & Jackson, 2007), as their novelty from the perspective of organismal life histories suggests that climatic factors should dominate predictions of biotic response. The prevalence of differential responses to climate change, on the other hand, demonstrates that climate-driven changes in fungal assemblages depend critically on the *sensitivity* of fungal assemblages to the extent of climate change under study.

In conservation biology, the term "sensitivity" was introduced in a conceptual framework to guide the conservation of species and populations in a changing climate (Williams et al., 2008). It refers to the biotic factors that shape organisms' responses to climate change, including physiological limits, plasticity, and genetic diversity (Williams *et al*., 2008; Dawson, Jackson, House, Prentice, & Mace, 2011). We extend this concept to the community level to describe variation in the rate of compositional

turnover with respect to climate. While there already exist community-level modeling approaches for representing possibly nonlinear compositional turnover (e.g., Ferrier, Manion, Elith, & Richardson, 2007), an alternative approach is to model turnover as the aggregate of individual species' presence or absence through ecological niche modeling (ENM; Murphy & Smith, 2021; Calabrese, Certain, Kraan, & Dormann, 2014). In this approach, the overlay of species niches predicts the composition under a set of environmental conditions, and the distribution of niche edges gives the compositional turnover predicted with an environmental change (Figure 2.1). Stacking ENMs has been shown to reproduce patterns in plant community composition (Peppler-Lisbach & Schröder, 2004). Furthermore, niche modeling opens new avenues to research: it not only captures a finer scale of biological organization which aids in the inference of community assembly processes, but it also enables the testing of hypotheses that require community data to be disaggregated to individual taxa or functional guilds.

For instance, an ongoing area of research in fungal biogeography concerns the drivers of the distributions of saprotrophic fungi and mycorrhizal fungi (Peay, 2016; Dickey *et al*., 2021) – two functional guilds of major importance to carbon and nutrient cycling in terrestrial ecosystems (McGuire & Treseder, 2010; Fernandez & Kennedy, 2016). Mycorrhizal fungal distributions are strongly influenced by plant host ranges (Sato, Tsujino, Kurita, Yokoyama, & Agata, 2012; Kivlin, Muscarella, Hawkes, & Treseder, 2017; van der Linde, *et al.*, 2018), and some evidence suggests that this creates stronger climatic controls on EcM fungi relative to saprotrophic fungi

(Sato *et al*., 2012). Among mycorrhizal fungi, EcM fungi are a particularly useful point of comparison for saprotrophic fungi because EcM lineages evolved multiple times from free-living saprotrophic ancestors (Tedersoo, May & Smith, 2010). The ratio of EcM fungi to saprotrophic fungi was decreased by in situ warming in an experimental forest site (Van Nuland *et al*., 2020), suggesting that saprotrophic fungi may tolerate higher temperatures than EcM fungi. In addition, there is evidence that EcM fungi can acquire water from their plant hosts during extreme drought (Querejeta, Egerton-Warburton, & Allen, 2003), providing a potential mechanism for greater drought tolerance in EcM fungi relative to saprotrophic fungi. Applying niche models to observational data at large spatial scales provides a means for validating these hypotheses.

In this study, we unify the climate sensitivity concept with a niche modeling approach to investigate the climatic drivers of soil fungal community assembly. We combine fungal sequence data from two continental-scale sampling networks in North America spanning all major biomes except tropical rainforest (Figure 2.2), and use ecological niche modeling (ENM) to model the composition of climate-constrained fungal assemblages across present North American climates. By constraining our analysis to the continental scale, we avoid the intercontinental dispersal barriers which would have otherwise violated the equilibrium assumptions of niche modeling, while maintaining an unprecedented breadth of climatic gradients over which to detect fungal niches. Finally, we operationalize the climate sensitivity of fungal assemblage composition using a novel statistical index that is a function of the

distribution of niches within climate space (Figure 2.1). We hypothesize that (1) the compositional turnover of soil fungal assemblages exhibits differential responses to climate change across environmental space, as captured by the clumped distribution of climate sensitivity values, and (2) ectomycorrhizal fungi differ from saprotrophic fungi in the distribution of their climate niche edges. Furthermore, we identify the biomes in which soil fungal assemblages are most sensitive to climate change.

Materials and Methods

Sequence data and bioinformatics

Soil samples were derived from two sampling networks: the Dimensions of Biodiversity of Ectomycorrhizal Fungi (DoB-Fun) network, which includes 68 sites distributed across North American forests dominated by Pinaceae, and the National Ecological Observatory Network (NEON), which includes 47 terrestrial sites distributed across all major eco-climatic zones in the USA. DoB-Fun samples were collected in 2011 through 2013, and we compiled data from all NEON soil samples collected in 2016 through 2018 using the 'neonMicrobe' R package (Qin *et al*., 2021). Across DoB-Fun sites, sampling and DNA extraction was conducted using methods previously described by Talbot *et al*. (2014). Across NEON sites, soils from multiple 40 m-by 40 m plots were cored to a maximum depth of 30 cm and split into mineral and organic horizons; DNA was then extracted from individual soil horizons. The

molecular methods for NEON soil samples are described in NEON's standard operating procedures (Battelle Memorial Institute, 2018, 2022). Combining these networks yielded 7,749 soil samples from across 113 sites (Figure 2.2; not all NEON sites yielded soil samples in this time frame).

Across all soil samples, DNA was extracted and the ITS1 locus of the internal transcribed spacer (ITS) region was amplified (Schoch *et al*., 2012) using the ITS1F–ITS2 primer pair (Smith & Peay, 2014). All NEON samples and a majority of DoB-Fun samples were sequenced on the Illumina MiSeq platform, while the remaining DoB-Fun samples were sequenced using 454 pyrosequencing technology. Due to the generally low quality of reverse reads, only forward reads were retained for Illumina sequences. Single read approaches have been found to be highly accurate in recovering mock fungal communities (Pauvert *et al.*, 2019). Illumina-sequenced amplicon reads were processed using the DADA2 pipeline (Callahan *et al*., 2016) with the following quality filtering parameters: *maxN* = 0, *maxEE* = 8, *truncQ* = 2, *minLen* = 50. 454 pyrosequencing reads were processed using a separate pipeline due to challenges in using DADA2 to generate reasonable estimates of sample richness. Briefly, QIIME and USEARCH (Edgar, 2010) were used to process and denoise 454 pyrosequencing reads, with a minimum (maximum) sequence length cutoff of 350 (1,200) bp, maximum homopolymer run length of 10 bp, and maximum barcode error number of 1.5. Due to computational constraints, denoised sequences could not be directly merged across sequencing runs. To facilitate cross sequencing platform comparisons, denoised sequences were clustered into species-level operational

taxonomic units (OTUs) at 97% similarity using VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016); only OTUs comprising more than 25 sequences were retained post clustering.

Representative sequences for each OTU were assigned taxonomy using the IDTAXA classifier via the 'DECIPHER' R package, which has been shown to outperform the naïve Bayesian classifier (Wright, 2016; Murali, Bhargava, & Wright, 2018). The UNITE v9 dynamic database (Nilsson *et al*., 2018) served as the reference for the IDTAXA classifier. In total, we identified over 68,000 fungal OTUs. In order to reduce the variance associated with conducting logistic regression on sparse datasets (Šinkovec, Heinze, Blagus, & Geroldinger, 2021), we retained only those OTUs present in at least 10 "operational sites," as defined below. Finally, OTUs were assigned to functional guilds using the FungalTraits database (Põlme *et al*., 2020).

Ecological niche modeling

Mean annual temperature (MAT), annual precipitation (MAP), temperature seasonality (TSEA), and precipitation seasonality (PSEA) were retrieved for the years 1970 to 2000 from WorldClim v2 (Fick & Hijmans, 2017). Climate data was structured as raster objects with a spatial resolution of 10 minutes of a degree. Soil samples were spatially aggregated into the nearest climate raster cells, resulting in 128 "operational sites." Soil pH in water (pH), total soil carbon content (%C), and total soil nitrogen content (%N) were retrieved from our soil samples or, when missing (8 missing pH, 5 missing %C and %N), were filled in using SoilGrids 2.0

(Poggio *et al.*, 2021).

ENM was conducted independently using regularized logistic regression for each species present in at least 10 operational sites. We initially fit species presence–absence against all climate variables and soil physicochemical variables, in addition to the quadratic transforms of MAT and MAP (MAT2 and MAP2). However, to reduce the computational complexity of the proceeding climate sensitivity analysis while minimizing the impact on model performance, we excluded PSEA and %N from our final models, as these were the two least important covariates not counting the quadratic transforms of more important variables (Table 2.1; see Appendix 2.1).

To improve model performance and reduce overfitting, we apply the ridge penalty to our logistic regression models using the 'glmnet' R package (Friedman, Hastie, & Tibshirani, 2010). Ridge regression is a form of regularization that penalizes model complexity by shrinking coefficients towards zero, thereby yielding more stable coefficients and predictions. Ridge regression and other forms of model regularization (e.g. the lasso) are already standard practices in maximum entropy modeling of species distributions (Warren & Seifert, 2011). The maximum likelihood estimator for the coefficients of the logistic ridge regression model for the *k*th OTU is given by the following set of equations:

$$\hat{\beta}_k = argmin_{\beta_k}[(\mu_k - X \times \beta_k)^2 + \lambda\|\beta_k\|^2] \text{ (Equation 1)}$$

$$\mu_k = ln(Y_k / (1 - Y_k))$$

The regularization parameter λ was chosen to minimize model deviance across 10 cross-validation folds using the *cv.glmnet* function in the 'glmnet' R package. ENMs predict habitat suitability, a continuous metric. To convert these predictions into species presence–absence Z_k, a dynamic threshold q_k was selected for each ENM such that $Z_k = \mathrm{I}(Y_k > q_k)$ maximizes the true skill statistic (Allouche, Tsoar, & Kadmon, 2006). The logistic ridge regression model was validated alongside an alternative maximum entropy model (Table 2.S1), implemented with the 'maxnet' (Phillips, 2021) and 'enmSdm' (Smith, 2021) R packages. These two independent approaches yield similar results; thus, we focus on reporting results from logistic regression in the main text. For all proceeding steps of our analysis, we remove OTUs with TSS less than or equal to zero.

Single-axis climate niche edges

We define the climate niche edges of an OTU by its predicted presence–absence transition points across MAT and MAP gradients. We construct MAT and MAP gradients spanning the range of observed values in our dataset (-13 to 25.1 ºC; 11.6 to 2556 mm), holding all other covariates constant at their median observed values (TSEA = 8 ºC; pH, 5.11; %C, 7.3%). For each OTU, we find the minimum and maximum values along each gradient for which an OTU is predicted to be present, and aggregate these distributions across all OTUs and by guild. We test for differences in the distributions of niche edges between guilds using the Kolmogorov–Smirnov test. In addition, we count the number of OTUs that are predicted to be

present at either extreme of each gradient, and consider these to be representative of niche truncation in our dataset, i.e. the degree to which the characterization of niches is limited by the extent of the environmental gradients sampled.

Climate sensitivity of fungal assemblage composition

To examine patterns in the compositional turnover of fungal assemblages with respect to climate, ENMs were used to predict species presence or absence across an environmental space that is discretized into a regular grid to form five-dimensional environmental grid cells spanning our predictor axes: MAT, MAP, TSEA, pH, and %C. The niche count in cell j, n_j, is operationalized as the number of species predicted to be present somewhere in the enclosed set of environmental conditions, i.e. $\Sigma_k Z_{jk}$. A cell j is considered to be a climate niche edge for species k if the following conditions are met: (1) species k is predicted to be present in cell j, i.e. Z_{jk} = 1; (2) cell j is edge-adjacent to another cell j' in which species k is predicted to be absent, i.e. $Z_{j'k} = 0$; and (3) cell j is adjacent to cell j' along the MAT, MAP, or TSEA axes. The latter condition ensures that only niche edges related to changes in climate are counted. The niche edge count in cell j is described by $\delta_j = \Sigma_k \varepsilon_{jk}$, where ε_{jk} is a dummy variable that equals 1 if cell j is a niche edge for species k, and equals 0 otherwise.

We propose an index for the climate sensitivity of the composition of biological communities. The *Sørensen climate sensitivity index* (abbreviated as Sørensen sensitivity or SS) is defined for a given climate cell j by the following

equation:

$$SS_j = \delta_j / n_j \text{ (Equation 2)}$$

The Sørensen climate sensitivity index is analytically related to the Sørensen dissimilarity index. Let *a* equal the number of species shared between two communities; let *b* and *c* be the number of species unique to the respective communities. Then the Sørensen dissimilarity index is defined by the following equation:

$$S = (b + c) / (2a + b + c) \text{ (Equation 3)}$$

The numerator in Equation 3, $b + c$, corresponds to the number of species gains plus the number of species losses associated with a transition from either community to the other; analogously, the numerator in Equation 2, δ_j, corresponds to the number of species that transition between presence and absence within cell *j*. The denominator in Equation 3, $2a + b + c$, corresponds to the number of species in either community ($a + b$) plus the number of species in the other ($a + c$); the denominator in Equation 2, n_j, corresponds to the number of species in the quasi-community consisting of species whose niches overlap with cell *j*. Our method for counting niches and niche edges within a cell ensures that the niche edge count does not exceed the niche count (i.e. $\delta_j \leq n_j$), so Sørensen sensitivity is bounded between 0

and 1.

Niches and niche edges were counted and Sørensen sensitivity was calculated across a five-dimensional grid formed by splitting each predictor axis into 9 regularly spaced intervals spanning the range of observed environmental conditions. For display purposes, these niche metrics were separately calculated across a 2-dimensional grid formed by splitting MAT and MAP into 20 regularly spaced intervals while all other predictors were held constant at their median observed values. Because all metrics are dependent on cell size, we generalize niche count and niche edge count as measures of *niche overlap* and *niche edge density*, respectively, and only compare metrics derived from the same grid. Metrics for cells that fell outside a 15% buffer of the convex hull enclosing observed environmental conditions were removed to limit model extrapolation. Niche overlap, niche edge density, and Sørensen sensitivity were mapped onto geographic space using the climate and soil rasters, generating maps of the spatial distribution of the metrics across North America. Sørensen sensitivity was summarized across all North American biomes compiled from the World Wildlife Fund (WWF) Terrestrial Ecoregions (Olson *et al*., 2001).

We used bootstrapping to produce uncertainty estimates for all our proposed metrics. Specifically, we used random resampling with replacement to generate a bootstrap sample of operational sites, then re-fitted ENMs across all OTUs and re-calculated all metrics. We repeated this procedure 50 times and calculated the mean and coefficient of variation across bootstrap iterations.

Metrics standardization and permutation analysis

To account for potentially spurious patterns generated in niche overlap, niche edge density, and Sørensen sensitivity during the modeling procedure, we generated null model distributions for each niche metric by randomly reshuffling the environmental data across sites, re-fitting ENMs, and re-calculating niche metrics; this was repeated 150 times. We then compare the distribution of each metric to its respective null distribution in two ways. First, we calculate a standardized form of each metric by dividing it by its mean null-model value on a cell-by-cell basis. Second, we use the null model to construct a permutation test for the hypothesis that each metric exhibits a more clumped distribution across environmental space than expected by random chance. We chose the variance-to-mean ratio as the test statistic for our clumping hypothesis: higher variance-to-mean ratios are indicative of more clumped distributions, whereas lower ratios are indicative of more uniform distributions. For a given metric, an empirical distribution of P-values was generated by calculating the proportion of null distributions with a variance-to-mean ratio greater than that of each bootstrap iteration. Because the expected distribution of P-values under the null model is uniform, a one-sided one-sample Kolmogorov-Smirnov test of the P-value distribution against the uniform distribution can be used to test the clumping hypothesis.

Results

Ecological niche models

After filtering to OTUs present in at least 10 out of 128 operational sites, 8,597 OTUs remained, including 1,015 EcM OTUs and 2,887 saprotrophic OTUs. ENMs based on logistic ridge regression have an out-of-sample predictive accuracy of 0.738 ± 0.101 (mean ± 1 SD) across OTUs and a true skill statistic (TSS) of 0.398 ± 0.295, outperforming maximum entropy models (Table 2.S1). TSS does not vary substantially across guilds or with the minimum number of presence or absence records (Figure 2.S6). After filtering to OTUs for which the model TSS was greater than zero, 7,274 OTUs remained, including 848 EcM OTUs and 2,426 saprotrophic OTUs.

Across ENMs, temperature covariates (MAT, TSEA, MAT2) have a higher median variable importance than precipitation covariates (MAP, PSEA, MAP2; Table 2.1, Figure 2.S5), suggesting that temperature is a stronger driver of fungal distributions in North America than precipitation. Among soil covariates, pH and %C are more important than %N, and all soil covariates are generally less important than temperature covariates (Table 2.1, Figure 2.S5).

Climate niche edges across guilds

Across the MAT gradient, 13.6% of all OTUs (19.6% of EcM, 14.1% of saprotrophs) of have cold niche edges that extend to the coldest sampled climates (-13 ºC), while 49.7% of all OTUs (47.3% of EcM, 51.2% of saprotrophs) have warm

niche edges that extend to the warmest sampled climates in our dataset (25.1 ºC; Figure 2.3a). Across the MAP gradient, 8.96% of all OTUs (12.0% of EcM, 9.8% of saprotrophs) have dry niche edges that extend to the driest sampled climates (116 mm), while 58.5% of OTUs (56.1% of EcM, 58.6% of saprotrophs) have wet niche edges that extend to the wettest sampled climates (2556 mm; Figure 2.3b). EcM and saprotrophic fungi differ in their distributions of cold and warm niche edges (two-sample Kolmogorov–Smirnov test, $P_{cold} = 0.005$, $P_{warm} = 0.004$; Figure 2.3a) but not in their distributions of dry and wet niche edges ($P_{dry} = 0.413$, $P_{wet} = 079$; Figure 2.3b). Notably, the cold niche edges of ectomycorrhizal fungi tend to occur at cooler temperatures than those of saprotrophic fungi (Figure 2.3a), though there exists considerable intra-guild variation in temperature niche edge distributions across taxa (Figure 2.S1). The null model partially accounts for the unimodal shape of the niche edge distributions, but not the differences between upper and lower niche edges (Figure 2.S2).

Clumped distributions of niche metrics

The permutation test provides strong support for the hypothesis that there are clumped patterns in the distributions of niche overlap (one-sample Kolmogorov-Smirnov test, $P < 0.0001$), niche edge density ($P = 0.0001$), and Sørensen climate sensitivity ($P < 0.0001$) across North American environments (Figure 2.4g-i). Niche overlap and niche edge density are highest in the warmest and wettest regions of the sampled environmental space (Figure 2.4b; Figure 2.S3). Sørensen sensitivity exhibits

a multimodal distribution across an environmental space defined across MAT and MAP, such that its response to precipitation depends on temperature (Figure 2.4c). These metrics exhibit different patterns prior to standardization, underscoring the importance of the null model for producing interpretable results (Figure 2.S3, Figure 2.S4).

Niche metrics over geographic space

Mapping niche metrics onto geographic space reveals high standardized niche edge density in the boreal forest, the Washington Cascades, and the southeastern U.S. (Figure 2.5a). Standardized niche overlap is high across southern temperate grasslands and eastern temperate forests and low across cooler and drier regions (Figure 2.5b). Consequently, standardized Sørensen sensitivity peaks in boreal forests and the small portions of tundra and Neotropical environments represented in our dataset (Figure 2.5c, Figure 2.6a). However, standardized Sørensen sensitivity in the tundra and the Neotropics displays high bootstrap uncertainty (Figure 2.5f, Figure 2.6b).

Discussion

Niche distributions suggest abrupt transitions and inter-guild differences

We find that the distribution of fungal niches displays nonuniform patterns across contemporary climate space, suggesting that despite the local-scale relevance of niche partitioning in maintaining the high α-diversity of soil fungal communities, broad groups of fungi share physiological limitations that create clusters in the distribution of their climate niches. Specifically, niche overlap and niche edge density are each more clumped in their distributions than expected by random chance (Figure 2.4g-i). This manifests in the unimodal distribution of cold and dry niche edges across the observed climate gradients (Figure 2.3). The peak in cold niche edges may reflect biophysical constraints on the reaction rate of extracellular enzymes involved in decomposition (Razavi, Blagodatskaya, & Kuzyakov, 2015), whereas the peak in dry niche edges may reflect biophysical limitations to substrate transport at low water potentials (Lennon, Aanderud, Lehmkuhl, & Schoolmaster, 2012; Schimel, 2018). By contrast, the distributions of warm and wet niche edges are not well characterized because approximately half lie beyond the climatic extent of our dataset (Figure 2.3). This indicates that in climates typical to North America, soil fungal assemblages are more likely to be constrained by cold and dry environments than by warm and mesic environments. Indeed, fungal isolates have been known to tolerate temperatures up to 40 ºC (Pietikäinen, Pettersson, & Bååth, 2005; Maynard *et al.*, 2019). Because warming is expected under climate change, our results suggest that fungal ranges in North America will tend to expand in the future. It has been theorized that the hypoxic conditions associated with especially wet soils may inhibit fungal activity

(Lennon *et al*., 2012), but as a local-scale control on fungal distributions, the effects of soil moisture may not be well represented by precipitation gradients.

The distributions of warm and wet niche edges within our sampling extent nonetheless provide novel insights into fungal compositional turnover at continental scales. Previous research utilizing the DoB-Fun network found that EcM fungal species richness displays two local optima across a temperature gradient, separated by a dip in diversity and a rapid shift in composition at mean annual temperatures around 10 ºC (Steidinger *et al*., 2020). In our analysis, this temperature separates a cluster of warm niche edges from a cluster of cold niche edges, such that the effects of warming between 0 and 20 ºC are characterized by species loss followed shortly by species gain (Figure 2.3a). In other words, we find evidence that non-EcM fungi also experience rapid compositional turnover and abrupt changes in species richness that are consistent with the existence of multiple α-diversity optima across this temperature range.

EcM fungi and saprotrophic fungi display subtle differences in their climatic controls, with EcM fungi spanning cooler but not necessarily drier or wetter climatic ranges than saprotrophic fungi (Figure 2.3). Sato *et al*. (2012) hypothesize that these inter-guild differences primarily reflect the host specificity of EcM fungi and the limited availability of EcM plant hosts across climatic gradients, rather than the physiological tolerance limits of EcM fungi per se. The tendency for EcM fungi to inhabit cooler climates is consistent with the observation that EcM plants tend to dominate in seasonally cold climates and diminish in abundance in warm, aseasonal

climates (Steidinger *et al*., 2019).

The compositional precarity of soil fungi in boreal forests

Considering patterns in standardized Sørensen climate sensitivity, we identify North American boreal forests as a segment of environmental space where soil fungal assemblage composition is likely to be especially sensitive to climate change (Figure 2.6a). We believe that this biome may have been overlooked in previous analyses due to their focus on warming-induced species losses and not species gains (e.g. Tedersoo *et al.*, 2022). As the boreal forest is experiencing especially rapid warming relative to lower-latitude biomes (Serreze & Barry, 2011), we predict that the effects of this exposure will be compounded by the compositional changes in soil fungi expected across these climatic gradients. Our findings are consistent with experimental studies that identify the boreal–temperate forest ecotone as a rapid transition zone for soil microbial communities. Fernandez *et al*. (2017) find that rapid shifts in the ectomycorrhizal fungal community at the boreal–temperate forest ecotone can be linked to warming-induced changes in the carbon allocation strategies of ectomycorrhizal hosts. We predict similarly rapid shifts across EcM and saprotrophic fungi, suggesting the effect of temperature on soil fungal community assembly is not entirely mediated by the response of plant hosts. Furthermore, our findings suggest that the increased dominance of saprotrophic fungi relative to EcM fungi under a warming experiment at the boreal–temperate forest ecotone (Van Nuland *et al*., 2020) is not solely driven by abiotic factors, but rather involves environment–species

interactions (Clark, Scher, & Swift, 2020) in which saprotrophs gain a competitive advantage over EcM fungi at higher temperatures.

Our findings complement previous studies that report rapid biotic changes among tree communities in the boreal–temperate forest ecotone, where warming may boost the competitive advantage of temperate tree species near their cold range limits relative to boreal species near their warm range limits (Reich & Oleksyn, 2008; Reich *et al*., 2015). Consequently, the communities that co-occur in this ecotone, including trees (Reich *et al*., 2015) and EcM fungi (Steidinger *et al*., 2020), are expected to have regionally divergent responses to warming depending on their association with temperate or boreal forests. This divergent response may account for the appearance of a latitudinal gradient in climate sensitivity (Figure 2.5c). Under equilibrium assumptions of community assembly, the northward migration of temperature isolines projected under most climate change scenarios could bring sweeping changes to the composition of soil fungal communities throughout Canada and Alaska within the twenty-first century, characterized largely by the influx of warm-adapted species.

The composition of soil fungal communities is closely tied to the rates of ecosystem processes such as decomposition (McGuire & Treseder, 2010) and the CO_2 fertilization effect (Terrer *et al*., 2016). These links exist independently of climate, such that the decomposition rate of recalcitrant organic matter such as wood is better predicted by fungal community structure than by climate (Bradford *et al*., 2014; Maynard *et al*., 2018; Smith & Peay, 2021). Thus, the multimodal pattern in the climate sensitivity of fungal assemblage composition points to potential climate

tipping points. In our study, the clearest peak in climate sensitivity is located spatially and environmentally within the boreal forest, a biome distinguished by its massive soil carbon stocks and slow carbon turnover (Crowther *et al*., 2019). Ultimately, the ecosystem implications of this high climate sensitivity depend on how fungal species are reshuffled into communities, how these communities are stratified across functional traits and guilds, and what kinds of disturbance regimes and successional trajectories are underway.

Limitations and future directions

Although our dataset offers an unprecedented coverage of soil fungal diversity across the North American continent and global biomes (Figure 2.2), it nevertheless has limitations to estimating fungal species niches and modeling community-level properties. Most niche models require tens to hundreds of presence or presence–absence records in order to fit a species niche; however, the high degree of regional endemism across fungal species (Talbot *et al*., 2014) forces a steep trade-off between taxonomic coverage and niche modeling performance. Consequently, our results primarily reflect the climatic responses of common taxa, which in our dataset tend to be niche generalists and pine specialists. Due to the prevalence of niche truncation in our dataset (Figure 2.3), an analysis of niche breadth is also not possible. In addition, our modeling approach seeks to represent deterministic community assembly via climate filtering and its associated biotic interactions. It does not account for resource limitation which can be important in tropical regions (Kivlin *et al*., 2017), nor does it

account for community assembly mechanisms with a stochastic component such as dispersal, diversification, drift, or priority effects (Zhou & Ning, 2017). Finally, our modeling approach does not predict abundance, which is a limitation of ecological niche modeling (Lee-Yaw *et al.*, 2022). Changes in the abundance of guilds, functional groups, or important taxa characterize the mycorrhizal fungal community shifts in some climate change experiments, sometimes to a greater extent than changes in diversity (Antoninka, Reich, & Johnson, 2011; Fernandez *et al.*, 2017). In addition, recent studies find that the relative abundance of saprotrophs decreases under warming (Feng *et al.*, 2022) while the relative abundance of pathogens increases (Delgado-Baquerizo *et al.*, 2020); in conjunction with our finding of greater niche overlap at higher temperatures, this suggests that competition among soil fungal taxa could increase at higher temperatures with consequences on ecosystem function.

Future studies may investigate the influence of environment–species interactions in fungal community assembly (Clark *et al.*, 2020), and the role of stochastic community assembly mechanisms such as dispersal limitation, historical contingencies, and drift in driving the high diversity of soil fungi, rather than niche partitioning across climate alone. Another potential area of research concerns the scale-dependency of the dispersion of fungal traits. Our study examines soil fungal species pools across sites and identifies under-dispersed trait distributions in the form of clustered niche edges, whereas at a finer-scale unit of analysis, over-dispersed trait distributions among fungi may point to the importance of competition in structuring communities (Crowther *et al.*, 2014). Finally, the Sørensen sensitivity index may

warrant more analytical attention, particularly as a tool for not only quantifying the rate of compositional turnover across environmental gradients but also predicting the total extent of turnover across spatiotemporal gradients.

Conclusions

While climate is not the sole driver of fungal community assembly, we demonstrate that its relationship with the distributional patterns of North American soil fungi is strong enough to drive complex, multimodal patterns in the rate of compositional turnover of soil fungal assemblages with respect to environmental gradients. We identify distinct patterns in the distribution of soil fungal niches and consequently the climate sensitivity of the composition of soil fungal assemblages, which peaks in boreal forests. Boreal forests have previously been shown to occupy a precarious climate space with respect to carbon storage and plant community turnover, and our findings demonstrate that this precarity extends to the composition of soil fungal assemblages. We propose that by modeling fungal assemblages in environmental hyperspace, niche modeling captures variation in compositional turnover that can be generalized to regions occupying similar climates around the globe. Thus, while the ecosystem implications of this compositional precarity are uncertain, the scale of its potential impacts is massive, as the geographic space occupied by the peak in high climate sensitivity is projected to sweep throughout subarctic latitudes under climate change.

Appendix 2.1: Variable importance and model validation

Variable importance and selection are coupled under the elastic net, a regularized regression method that penalizes unimportant variables in a model by shrinking their coefficients. In our analysis, we use a form of the elastic net known as ridge regression, which enforces an L_2 penalty that shrinks coefficients toward zero. Consequently, the absolute values of scaled coefficients in a ridge regression model can be used as a metric of variable importance (Kuhn, 2008).

To select variables for use in modeling the presence and absence of fungal operational taxonomic units (OTUs), we begin with the dataset containing only OTUs present in at least 10 operational sites. We split this dataset into training and testing subsets using an approximate 70-30 split, leaving 88 operational sites in the training set and 40 operational sites in the testing set. First, we fit an initial suite of logistic ridge regression models (n = 8,597) using the full set of available predictors: mean annual temperature (MAT), annual precipitation (MAP), temperature seasonality (TSEA), precipitation seasonality (PSEA), soil pH in water (pH), total soil carbon content (%C), and total soil nitrogen content (%N). This produces a distribution of variable importance for each predictor (Figure 2.S5). Second, we fit a suite of logistic ridge regression models consisting of only the climate predictors. Third, we fit a suite of logistic ridge regression models consisting of the 5 most important predictors, including any applicable quadratic transforms; this excludes PSEA and %N (Figure 2.S5).

We select from among the candidate predictors by validating their models'

performance against the testing subset. The models using only climate predictors are consistently outperformed by the alternatives; the models using the full set of variables perform marginally better than the models using the 5 most important variables in terms of the true skill statistic, while the opposite is true for total accuracy (Table 2.S1). Because the computational complexity of our climate sensitivity analyses increases exponentially with the addition of new predictor axes, we decide to use only the 5 most important variables rather than the full set of predictors in our climate sensitivity analyses.

Finally, we compare the logistic ridge regression model using the 5 most important variables against a maximum entropy model using the same predictors.

The code used for this appendix is available at https://github.com/claraqin/fungal-climate-niche.

Table 2.1. Variable importance of the full set of covariates across ridge regression models for all fungal operational taxonomic units (OTUs; n = 8,597), in order of median variable importance. The right-most column indicates whether each variable was ultimately included in the models used to quantify Sørensen climate sensitivity.

Covariate	Variable importance, median	Variable importance, mean	Included in final models
Mean annual temperature (MAT)	0.200	0.392	Yes
Temperature seasonality (TSEA)	0.196	0.408	Yes
Soil pH (pH)	0.193	0.296	Yes
Total soil carbon content (%C)	0.176	0.329	Yes
MAT^2 (MAT2)	0.169	0.288	Yes
Mean annual precipitation (MAP)	0.155	0.334	Yes
Precipitation seasonality (PSEA)	0.155	0.254	No
Total soil nitrogen content (%N)	0.0969	0.177	No
MAP^2 (MAP2)	0.0899	0.175	Yes

Figure 2.1. Conceptual diagram of the relationship between niche distributions and the climate sensitivity of soil fungal assemblage composition, i.e. Sørensen climate sensitivity. (a) If niches (gray ellipsoids) are randomly or uniformly distributed across environmental space, then the numbers of niches and niche edges within a standard unit of climate change (unfilled square) are approximately regularly distributed across environmental space. (b) On the other hand, if niches are clustered together, then the numbers of niches and niche edges have clumped distributions, resulting in pockets of environmental space in which fungal assemblage composition is especially sensitive to climate change. Although a 2-dimensional climate space is used here, this conceptual framework may be extended to any *n*-dimensional environmental hyperspace.

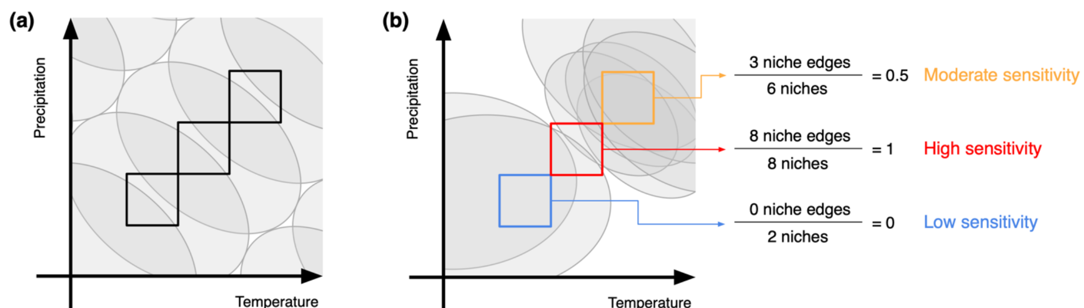


Figure 2.2. Distribution of soil sampling sites (n = 113) in (a) geographic space and (b) climate space. Points represent site locations: blue, NEON site; red, DoB-Fun site. The study sites cover all major biomes except tropical rain forest. The biome plot is constructed using the 'plotbiomes' R package (Stefan & Levin, 2022).

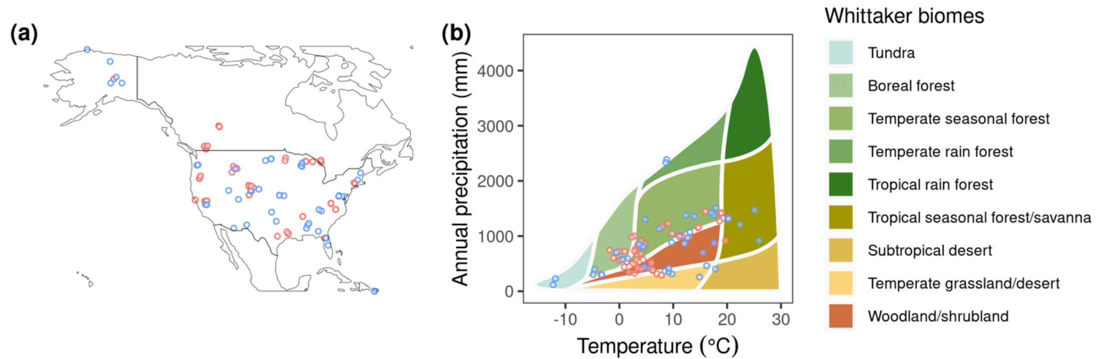
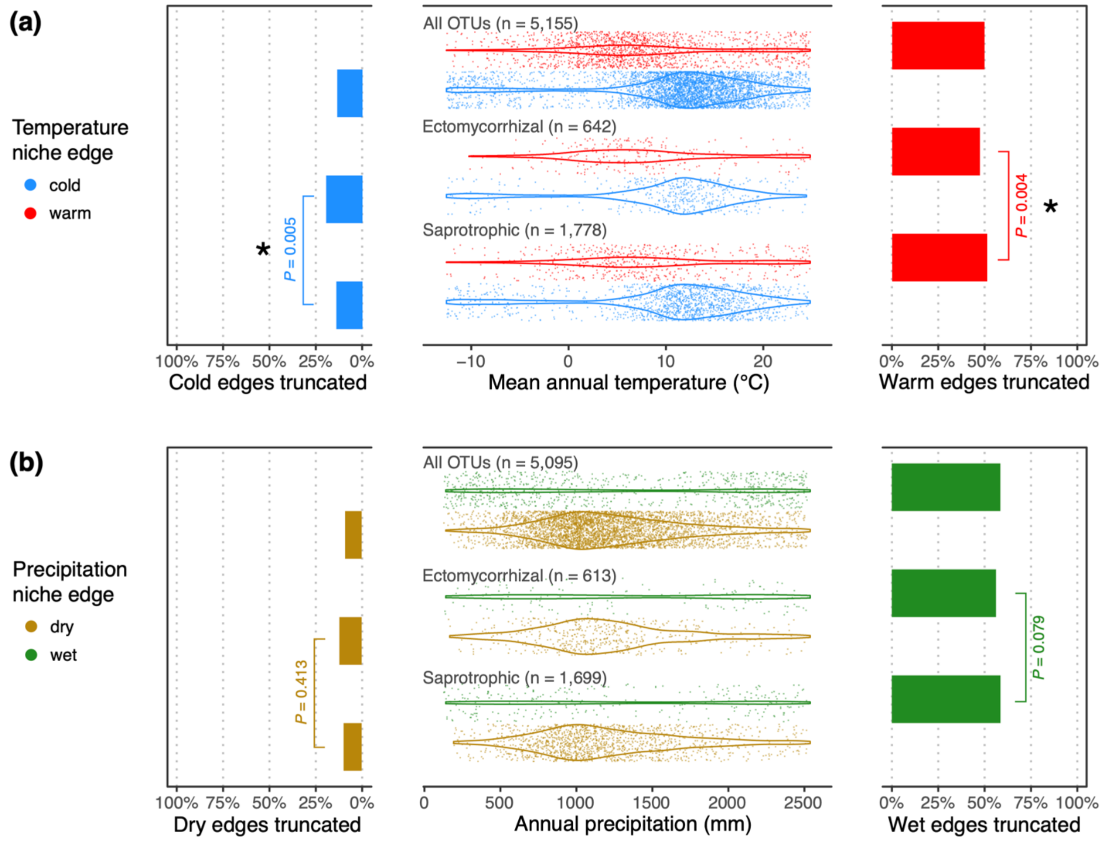
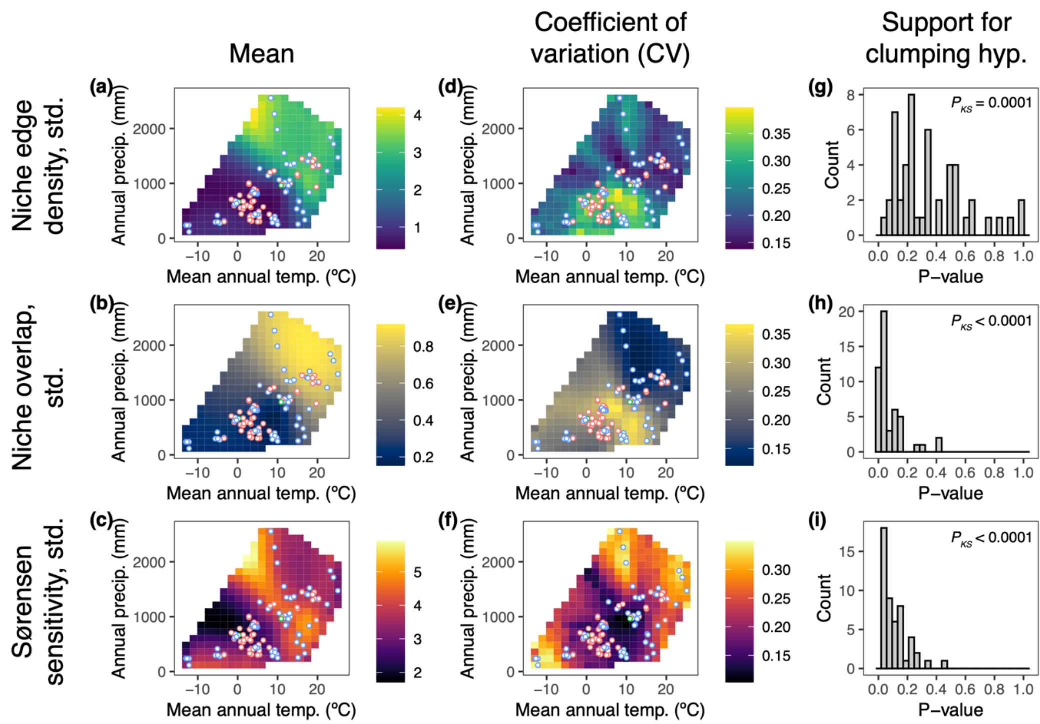


Figure 2.3. Fungal species niche edges across temperature and precipitation gradients in North America. (a) The distributions of cold and warm niche edges for all fungal operational taxonomic units (OTUs) and across guilds, and the proportions of cold (or warm) niche edges that are truncated at the minimum (or maximum) sampled mean annual temperature (MAT). (b) The distributions of dry and wet niche edges for all OTUs and across guilds, and the proportions of dry (or wet) niche edges that are truncated at the minimum (or maximum) sampled annual precipitation (MAP). P-values correspond to the two-sample Kolmogorov-Smirnov test comparing the niche edge distributions between ectomycorrhizal and saprotrophic fungi. All covariates not displayed in a plot are held constant at their median observed values: MAT, 8 ºC; MAP, 810 mm; temperature seasonality, 8 ºC; soil pH, 5.11; soil carbon content, 7.3%. If an OTU is absent across an entire displayed gradient (MAT: 29.1% absent; MAP: 30.0% absent), it is not included in the OTU counts for the gradient.



(a)
Temperature niche edge
cold
warm
All OTUs (n = 5,155)
Ectomycorrhizal (n = 642)
Saprotrophic (n = 1,778)
P = 0.005
P = 0.004
*
100% 75% 50% 25% 0%
Cold edges truncated
−10 0 10 20
Mean annual temperature (°C)
0% 25% 50% 75% 100%
Warm edges truncated
(b)
Precipitation niche edge
dry
wet
All OTUs (n = 5,095)
Ectomycorrhizal (n = 613)
Saprotrophic (n = 1,699)
P = 0.413
P = 0.079
100% 75% 50% 25% 0%
Dry edges truncated
0 500 1000 1500 2000 2500
Annual precipitation (mm)
0% 25% 50% 75% 100%
Wet edges truncated

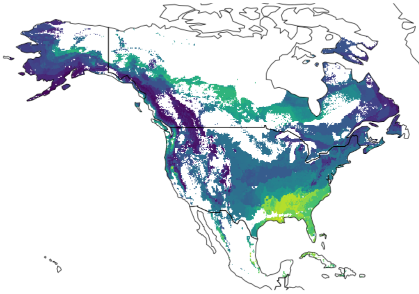
Figure 2.4. Standardized niche metrics and climate sensitivity of soil fungal assemblage composition across North American climates. Mean standardized values (a-c) and coefficients of variation of standardized values (d-f) are calculated across bootstrap iterations and displayed across a two-dimensional slice of climate space for each metric: standardized niche edge density (a, d), standardized niche overlap (b, e), and standardized Sørensen sensitivity (c, f). Points in (a-f) represent the positions of operational sites (n = 128) in climate space: blue, NEON site; red, DoB-Fun site; green, operational site composed of both NEON and DoB-Fun sites. Metrics displayed in (a-f) are calculated over a 20×20-cell regular grid spanning observed mean annual temperature and annual precipitation values constrained by a convex hull, while all other covariates are held constant at their median observed values: temperature seasonality, 8 ºC; soil pH, 5.11; soil carbon content, 7.3%. (g-i) Support for the hypothesis that a metric follows a clumped rather than a random distribution across the five-dimensional environmental hyperspace is based on the empirical distribution of P-values derived from the permutation analysis. A randomly distributed metric would produce a uniform distribution of P-values, whereas a clumped metric would produce a distribution of P-values that tends toward zero. The deviation from a uniform distribution is evaluated using the one-sided one-sample Kolmogorov-Smirnov test, producing one measure of significance (P_{KS}) for each metric.



Mean
Coefficient of variation (CV)
Support for clumping hyp.
Niche edge density, std.
Niche overlap, std.
Sørensen sensitivity, std.
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
Annual precip. (mm)
Mean annual temp. (°C)
Count
P−value
P_{KS} = 0.0001
P_{KS} < 0.0001
P_{KS} < 0.0001

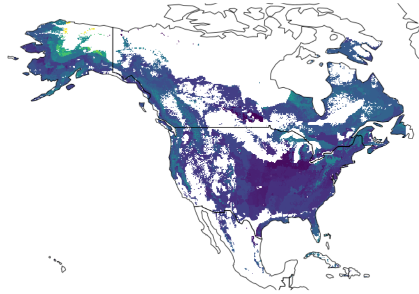
Figure 2.5. Niche metrics and climate sensitivity of soil fungi across North America. For each metric, mean standardized values (a-c) and coefficients of variation of standardized values (d-f) are calculated across bootstrap iterations and mapped to North American geographies according to climatic and soil physicochemical rasters: standardized niche edge density (a, d), standardized niche overlap (b, e), and standardized Sørensen sensitivity (c, f). Metrics are calculated over a $9 \times 9 \times 9 \times 9 \times 9$-cell regular grid spanning the observed range of each environmental axis, constrained by a five-dimensional convex hull with a 15% buffer.

(a)

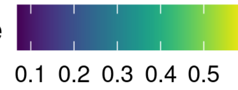


Mean niche edge density, std.

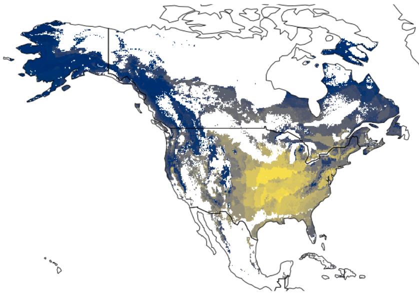
(d)



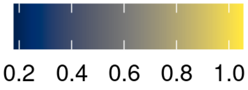
CV of niche edge density, std.



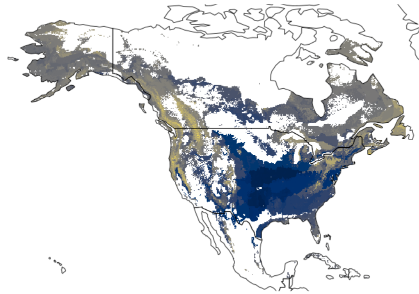
(b)



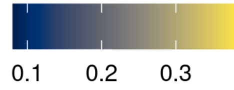
Mean niche overlap, std.



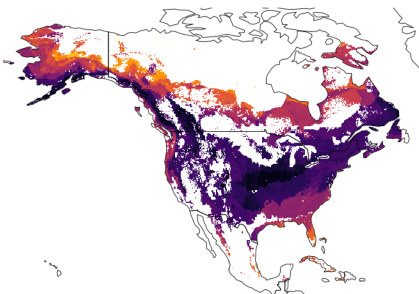
(e)



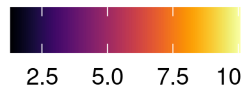
CV of niche overlap, std.



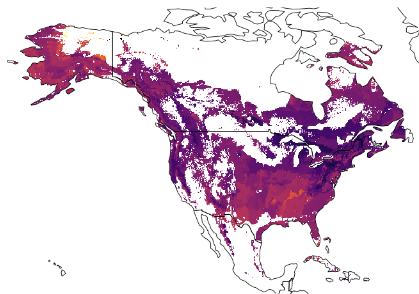
(c)



Mean Sørensen sensitivity, std.



(f)



CV of Sørensen sensitivity, std.

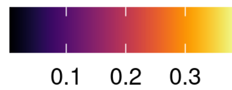
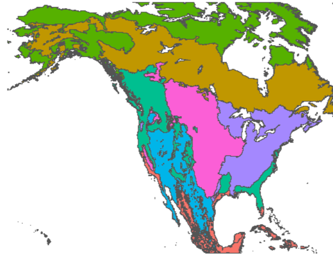
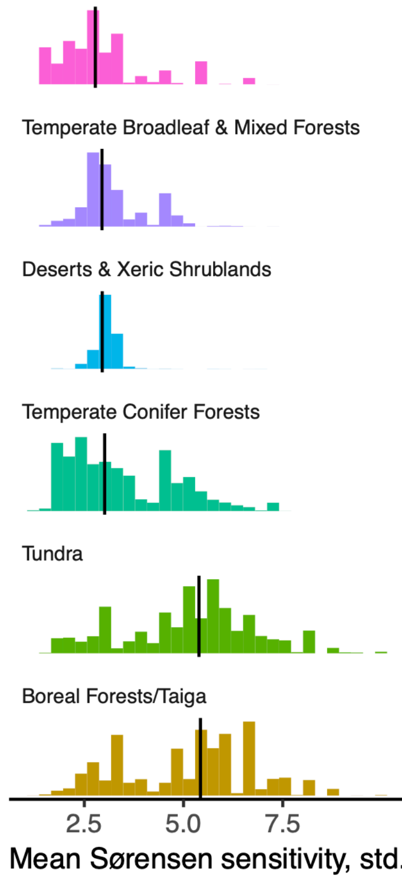


Figure 2.6. Standardized Sørensen climate sensitivity of soil fungi across major North American biomes. (a) Mean standardized Sørensen sensitivity across bootstrap iterations and (b) the coefficient of variation of standardized Sørensen sensitivity across bootstrap iterations are grouped by biome boundaries to generate within-biome distributions. Black vertical lines mark the median of each distribution. Biomes are displayed from top to bottom in order of increasing median value of the mean standardized Sørensen sensitivity. Only North American biomes occupying more than 500,000 km^2 of the climates represented in our dataset are displayed.




(a)
Temperate Grasslands, Savannas & Shrub
Temperate Broadleaf & Mixed Forests
Deserts & Xeric Shrublands
Temperate Conifer Forests
Tundra
Boreal Forests/Taiga
2.5
5.0
7.5
Mean Sørensen sensitivity, std.

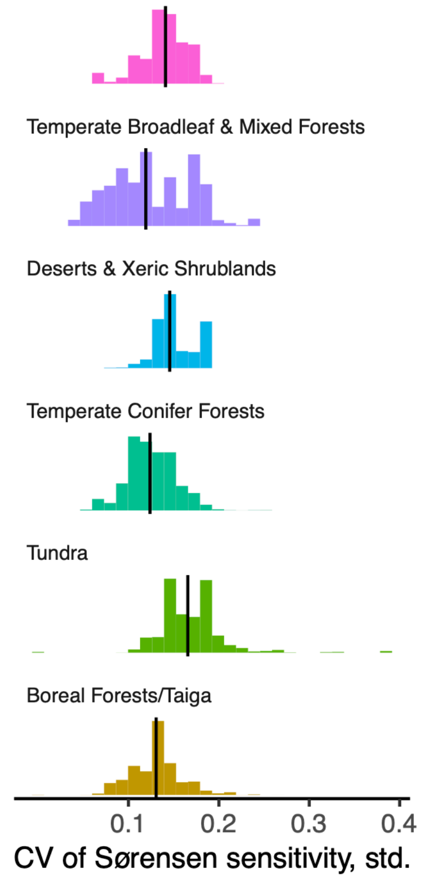

(b)
Temperate Grasslands, Savannas & Shrub
Temperate Broadleaf & Mixed Forests
Deserts & Xeric Shrublands
Temperate Conifer Forests
Tundra
Boreal Forests/Taiga
0.1
0.2
0.3
0.4
CV of Sørensen sensitivity, std.

Table 2.S1. Out-of-sample predictive performance of logistic ridge regression models and maximum entropy models for the presence and absence of 8,597 fungal operational taxonomic units (OTUs) present in at least 10 operational sites. Three sets of predictors are compared based on total accuracy, sensitivity, specificity, and true skill statistic. The full model includes mean annual temperature (MAT), annual precipitation MAP), temperature seasonality (TSEA), precipitation seasonality (PSEA), the quadratic transforms of MAT (MAT2) and MAP (MAP2), soil pH in water (pH), total soil carbon content (%C), and total soil nitrogen content (%N). MAT, MAP, TSEA, PSEA, MAT2, and MAP2 comprise the full set of climate variables. Values shown are the mean (and standard deviation) across all OTUs.

Model class	**Predictors**	**Accuracy**	**Sensitivity**	**Specificity**	**True skill statistic**
Logistic ridge regression	Full set	0.737 (0.102)	0.650 (0.302)	0.750 (0.118)	0.400 (0.299)
Logistic ridge regression	Climate variables only	0.724 (0.104)	0.640 (0.313)	0.738 (0.124)	0.378 (0.303)
Logistic ridge regression	MAT + MAP + TSEA + MAT2 + MAP2 + pH + %C	0.738 (0.101)	0.647 (0.304)	0.751 (0.118)	0.398 (0.295)
Maximum entropy	Climate variables only	0.621 (0.197)	0.839 (0.206)	0.545 (0.252)	0.384 (0.288)
Maximum entropy	MAT + MAP + TSEA + MAT2 + MAP2 + pH + %C	0.473 (0.127)	0.783 (0.174)	0.377 (0.165)	0.160 (0.216)

Figure 2.S1. Fungal species presence–absence transitions across a temperature gradient, by genus. Only the top 10 most abundant genera among ectomycorrhizal and saprotrophic fungi in our dataset are displayed, from top to bottom. The distribution of cold and warm niche edges is plotted across a mean annual temperature gradient, with all other climate variables held constant at their median observed values: annual precipitation, 81.0 cm; temperature seasonality, 8 ºC; soil pH, 5.11; soil carbon content, 7.3%. Fungal species-level taxa are approximated by operational taxonomic units (OTUs) at 97% similarity. If an OTU is absent across an entire displayed gradient, it is not included in the OTU counts for the gradient.

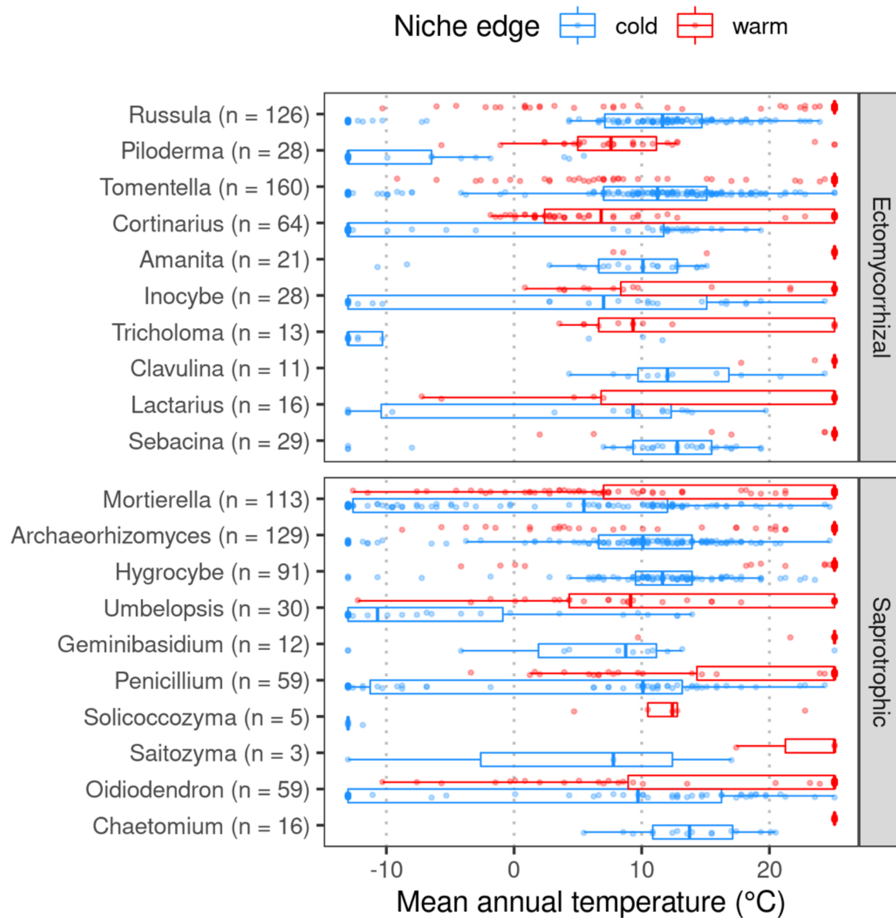


Figure 2.S2. Fungal species niche edges across temperature and precipitation gradients in North America under the null model. The null model is constructed using 10 random permutations of the environmental data. The distributions of cold and warm niche edges (a) and dry and wet niche edges (c) for all fungal operational taxonomic units (OTUs). Thin lines represent distributions across null-model permutations, and bold lines represent the average distributions across permutations. (b) The proportions of cold (or warm) niche edges that are truncated at the minimum (or maximum) sampled mean annual temperature (MAT), and (d) the proportions of dry (or wet) niche edges that are truncated at the minimum (or maximum) sampled annual precipitation (MAP). Bars represent the mean across permutations and error bars represent the mean ± 1 standard deviation across permutations. All covariates not displayed in a plot are held constant at their median observed values: MAT, 8 ºC; MAP, 810 mm; temperature seasonality, 8 ºC; soil pH, 5.11; soil carbon content, 7.3%.

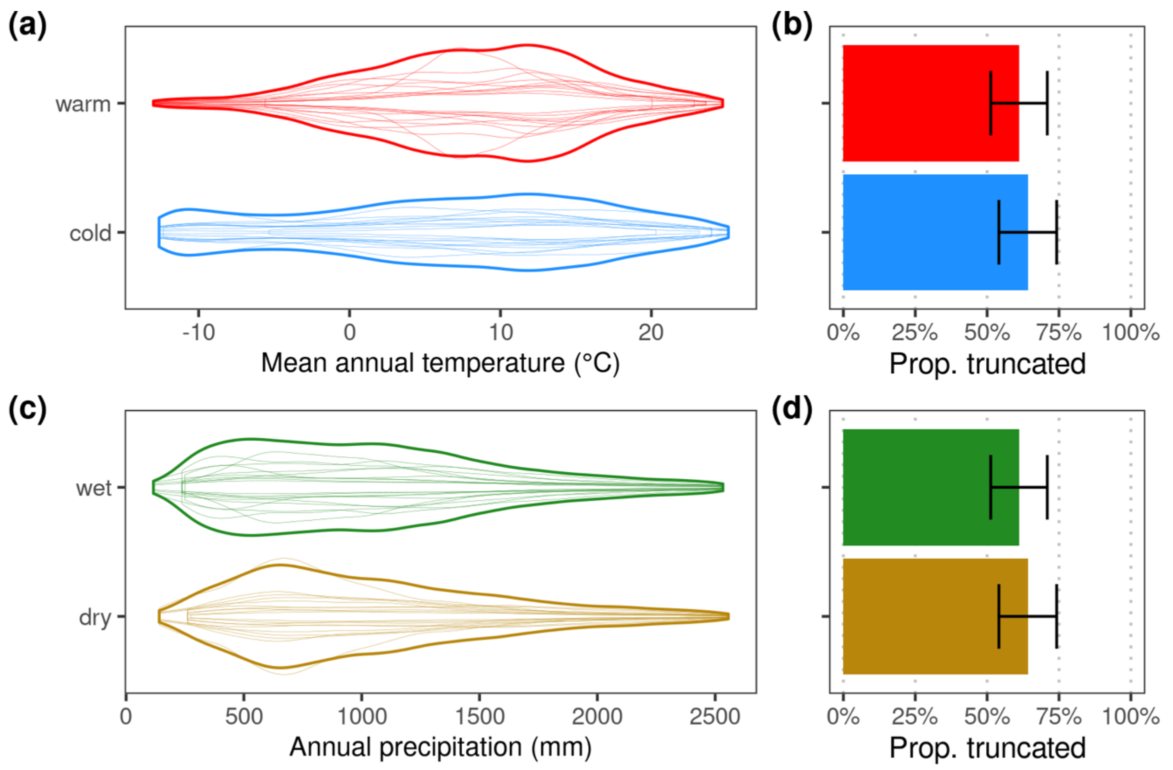

(a)
warm
cold
-10
0
10
20
Mean annual temperature (°C)
(b)
0%
25%
50%
75%
100%
Prop. truncated
(c)
wet
dry
0
500
1000
1500
2000
2500
Annual precipitation (mm)
(d)
0%
25%
50%
75%
100%
Prop. truncated

Figure 2.S3. Niche metrics and climate sensitivity of soil fungi across a two-dimensional slice of North American climate space, including null and standardized distributions. All metrics reported in this figure are the mean values across 200 iterations (of bootstrap samples or null-model permutations). (a) Niche edge density, reported as the mean number of niche edges cell across bootstrap iterations in each climate cell. (b) Niche overlap, reported as the mean number niches across bootstrap iterations in each climate cell. (c) Sørensen climate sensitivity is calculated by dividing niche edge density by niche overlap and is reported as its mean value across bootstrap iterations. The metrics are also reported as mean values in the null model (d, e, f), and mean values in standardized form (g, h, i); the latter is obtained by dividing the mean boostrap values by the mean null-model values on a cell-by-cell basis. Points represent the positions of operational sites in climate space: blue, NEON site; red, DoB-Fun site; green, climate cell containing both NEON and DoB sites. Climate cells displayed in this plot are defined by a 20 cell-by-20 cell regular grid spanning observed mean annual temperature and mean annual precipitation values, constrained by a convex hull. Across climate cells displayed in this figure, all other environmental covariates are held constant at their median observed values: temperature seasonality, 8 ºC; soil pH, 5.11; soil carbon content, 7.3%.

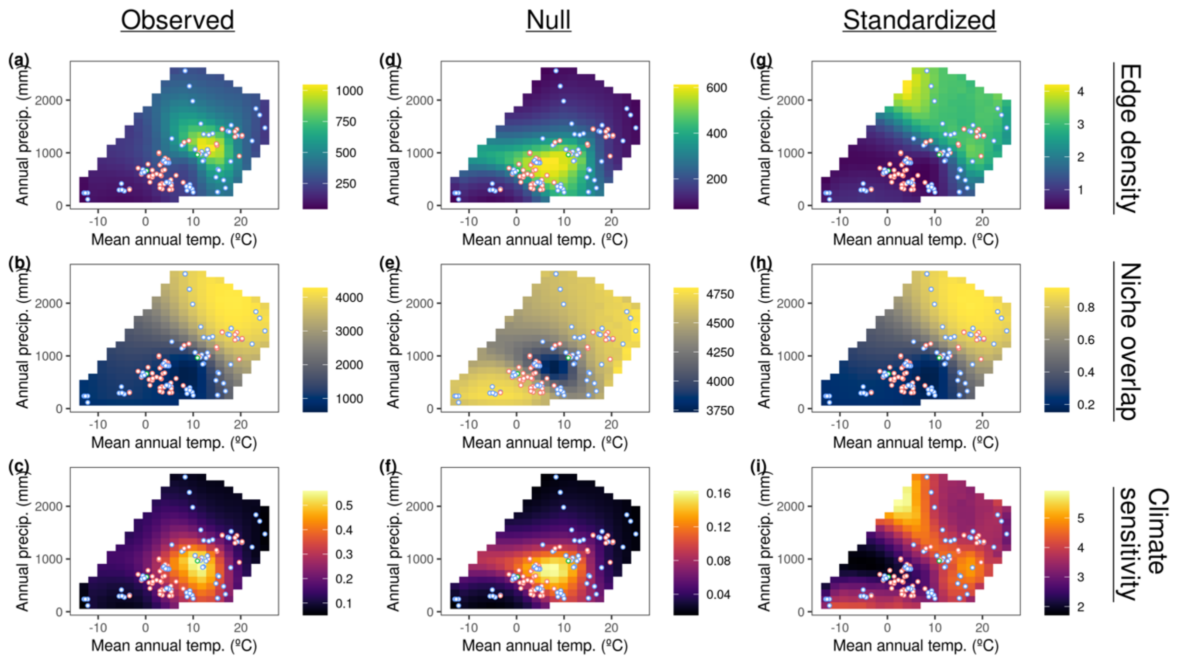
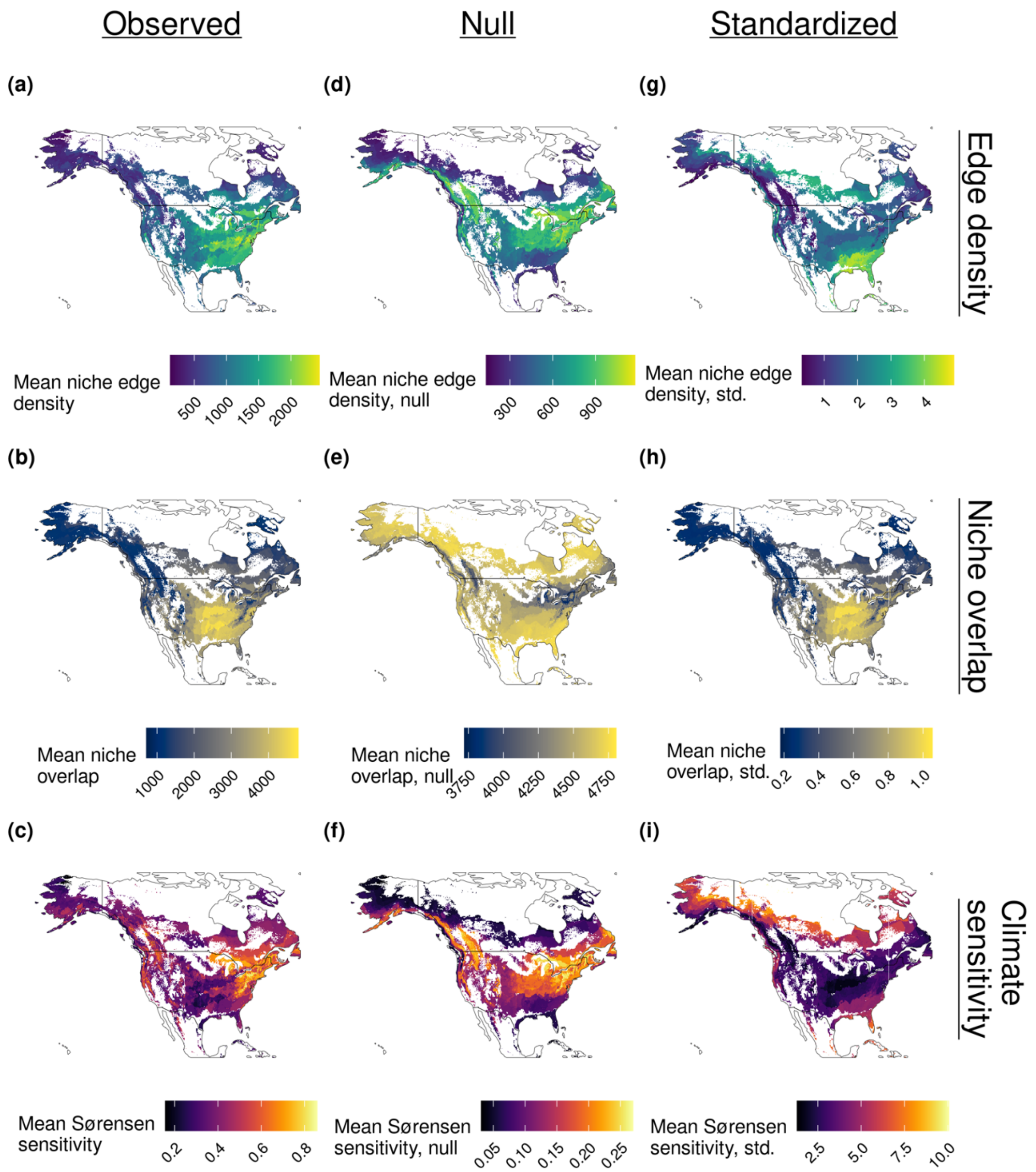

Observed
Null
Standardized
Edge density
Niche overlap
Climate sensitivity
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
Annual precip. (mm)
Mean annual temp. (ºC)

Figure 2.S4. Niche metrics and climate sensitivity of soil fungi across North America. For each metric, mean values across bootstrap iterations (a-c), mean values across null model permutations (d-f), and mean standardized values (g-i) are mapped to North American geographies according to climatic and soil physicochemical rasters: niche edge density (a, d, g), niche overlap (b, e, h), and Sørensen sensitivity (c, f, i). Metrics are calculated over a 9×9×9×9×9-cell regular grid spanning the observed range of each environmental axis, constrained by a 5-dimensional convex hull with a 15% buffer.



Observed
Null
Standardized
(a)
(d)
(g)
Edge density
Mean niche edge density
500 1000 1500 2000
Mean niche edge density, null
300 600 900
Mean niche edge density, std.
1 2 3 4
(b)
(e)
(h)
Niche overlap
Mean niche overlap
1000 2000 3000 4000
Mean niche overlap, null
3750 4000 4250 4500 4750
Mean niche overlap, std.
0.2 0.4 0.6 0.8 1.0
(c)
(f)
(i)
Climate sensitivity
Mean Sørensen sensitivity
0.2 0.4 0.6 0.8
Mean Sørensen sensitivity, null
0.05 0.10 0.15 0.20 0.25
Mean Sørensen sensitivity, std.
2.5 5.0 7.5 10.0

Figure 2.S5. Variable importance of the full set of predictors across all ridge regression models (n = 8,597), in order of median variable importance. Each red point represents a variable importance estimate for the ridge regression model for a given OTU, and a violin plot summarizes the distribution of variable importance values for each predictor. Abbreviations for variable names: MAT, mean annual temperature; TSEA, temperature seasonality; MAP, mean annual precipitation; PSEA, precipitation seasonality; MAT2, quadratic transform of mean annual temperature; MAP2, quadratic transform of mean annual precipitation; pH, soil pH in water; %C, total soil carbon content; %N, total soil nitrogen content.

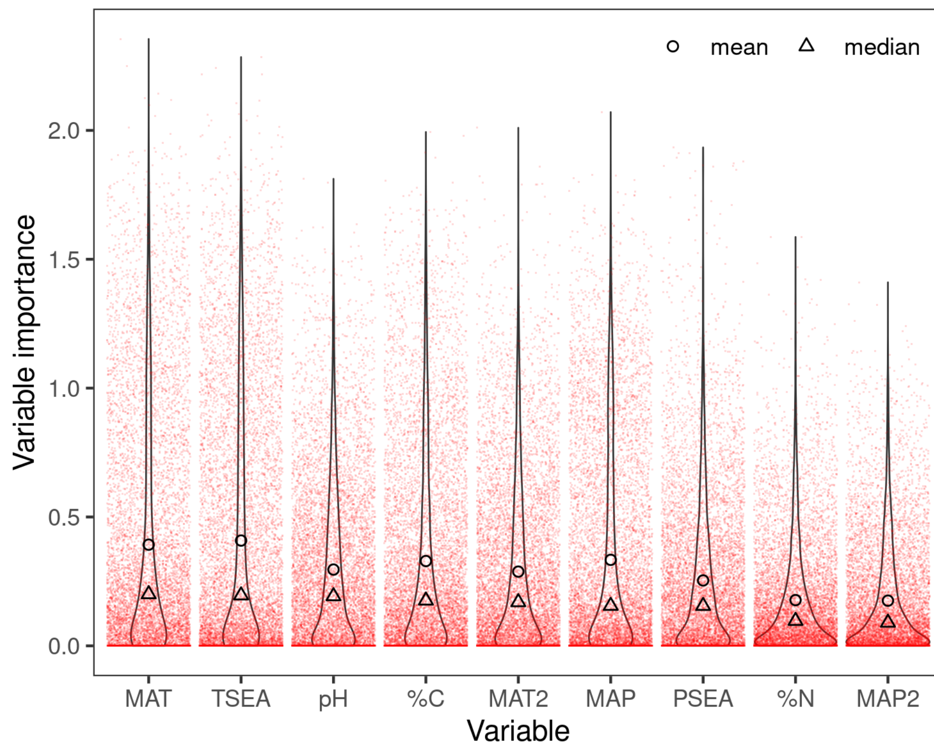
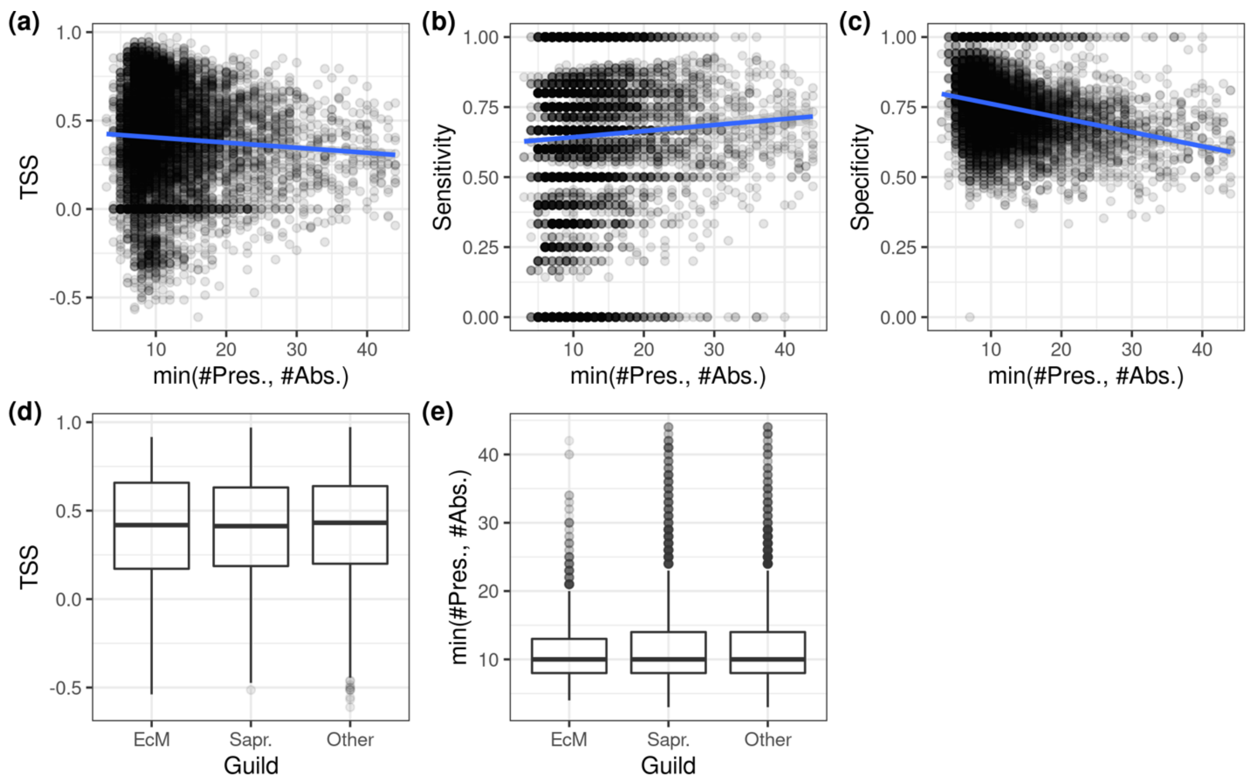


Figure 2.S6. Model validation metrics for the logistic ridge regression models including all predictors except soil pH and total soil nitrogen content (n = 8,597). True skill statistic (TSS) (a), sensitivity (b), and specificity (c) versus the lower value between the number of presences and the number of absences (min(#Pres., #Abs.)) of the corresponding fungal operational taxonomic unit (OTU). The blue line represents the line of best fit. TSS (d) and min(#Pres., #Abs.) (e) across the major guilds analyzed in this study (n_{EcM} = 1,015, n_{Sapr} = 2,887, n_{Other} = 4,695). Guild abbrevations: EcM, ectomycorrhizal fungi; Sapr., saprotrophic fungi; Other, fungi from another guild or with an unknown guild assignment.



CHAPTER 3

"It could be anyway": Ambivalence to the knowing of soilborne plant disease in post-fumigant agriculture

Introduction

The plant pathology lab at the U.S. Department of Agriculture (USDA) research station in Salinas, California is made up of several sites. Greenhouses are lined with potted plants awaiting or undergoing a randomized controlled trial. Agricultural fields are available for production-scale trials operating on similar experimental principles. Aptly named "clean rooms" are designated for cell culturing and DNA-based molecular work. In theory, all sites are united in the study of soilborne strawberry disease and methods for controlling it. However, in practice, disease looks and behaves differently at each site. In the fields, disease is confirmed by the sight of wilted leaves. In the greenhouse, disease is discerned not by symptoms but by the density of a pathogen in the soil. In the analysis of DNA from a soil microbiome, disease is marked by the presence of a pathogen amidst the absence of suppressive microbiota. All of these methods were used by the same research team—a group of agricultural scientists whom I was fortunate to call my colleagues—and were not considered to be particularly unusual. Our methods had, in fact, been vetted and funded by a research grant program administered by the USDA specifically to rescue the specialty crop industry from the existential threat posed by soilborne plant

disease.

California is a major producer of specialty crops, including 90 percent of strawberries in the U.S. (CDFA, 2022). The adoption of chemical fumigants into its production routines in the mid-twentieth century was key to its success (Guthman, 2019), as it allowed the industry to control soilborne diseases without practicing crop rotation (Willhelm & Paulus, 1980). In particular, methyl bromide fumigants helped the California strawberry industry gain a foothold by controlling Verticillium wilt (Wilhelm et al., 1961), weeds, and nematodes (Fennimore et al., 2003). Under the Montreal Protocol of 1987, however, methyl bromide was phased out of agricultural production due to its deleterious effects on the ozone layer. Its phaseout coincided with the proliferation of novel soilborne fungal pathogens that cause widespread wilt and root rot in strawberry, motivating a search for alternatives to methyl bromide. In the short term, the reduction in methyl bromide has been offset by an increase in other fumigants (CDPR, 2019), though mounting pressure from public health advocates may lead to more general fumigant bans (Holmes et al., 2020). Popular non-fumigant methods for soil disinfestation, which include soil solarization and steam sterilization, have proven cost-prohibitive or inconsistent in their performance (Holmes et al., 2020). The absence of a perfect alternative to methyl bromide is evidenced by California strawberry growers' struggle to find economically feasible disease management solutions without it. As a result, California strawberry growers received critical use exemptions to continue fumigating with methyl bromide through 2016, over ten years later than outlined in the original mandate. Now the California

strawberry industry, along with several other specialty crop industries for which methyl bromide served as the cornerstone of disease management, is forced to consider more radical departures from chemical fumigation.

"Biological control" offers an alternative approach to disease management. Unlike chemical methods, biological control methods manage disease by managing the interactions between a disease-causing pathogen and the organisms that inhibit its growth or suppress its pathogenicity. This can be achieved through the direct inoculation of beneficial microbes like *Trichoderma* and *Bacillus*, or through modifications to the soil environment to indirectly enhance beneficial populations (Mazzola & Freilich, 2017). The phaseout of methyl bromide thus created a market opportunity for soil microbial inoculants, recently valued at over $400 million (Mazzola & Freilich, 2017), and simultaneously spurred an interest in evaluating the biocontrol potential of time-honored practices for improving soil health. These include cover cropping and a lesser-known organic method called anaerobic soil disinfestation (ASD), which combines organic soil amendments, plastic tarping, and irrigation to induce oxygen-free conditions in agricultural soils (Shennan et al., 2014). Of the over $100 million that the USDA has invested into research and extension projects through the Methyl Bromide Transitions (MBT) program and related programs (Holmes et al., 2020), a small but increasing percentage has been awarded to projects that employ biological rather than strictly chemical control methods (USDA 2023).

The push for biological alternatives to methyl bromide reveals an opening to

reconceptualize disease in the agricultural industry. Perhaps disease is not only the presence of one microbe—the pathogen—but the absence of a whole consortium of microbes that could have suppressed the development of disease. In mapping a future for the specialty crop industry, it is reasonable to expect that the question of how disease should be managed hinges on the matter of *what disease is*. If disease is the presence of a pathogen, then other broad-spectrum, biocidal solutions will do: chemicals, steam, and solarization. But if disease is anything other than the presence or absence of a pathogen, then we can no longer be confident in treatments deemed superior for their ability to exclude or eliminate pathogens. Understood as a question about objective reality, it is turned over to the scientists. And once in their laboratories, it also becomes a question about how reality is known—at what sites, with what methods, and with what explanations?

In this chapter, I build upon a body of scholarship in Science and Technology Studies (STS) that follows the sociomaterial practices through which objects are known and simultaneously made to matter (Mol, 1999, 2002; Moser, 2008; Latour, 2004). This literature stresses the ontological stakes of knowledge practices that had previously been considered merely epistemic—a move that has been called the "ontological turn" in STS (Woolgar & Lezaun, 2013). I situate myself within the ontological turn to ask, what are the limitations of knowledge practices as points of intervention, especially when those practices are being deployed in a highly instrumental research setting? I propose that multiplicity in knowledge practices does not merely create new possibilities; it may also enshrine preexisting certainties in

other practices that draw from them.

This chapter stems from my observations on how soilborne plant disease is known and how its knowing is entangled with the technical and economic realities of production agriculture. For two consecutive summers, I worked alongside scientists who sought to understand the mechanisms of biological methods for soilborne disease control. I assisted them on a research and extension project funded by the MBT program to optimize biological disease control strategies in California strawberry production. To understand disease in practice, I follow the knowledge practices of the research team, focusing on techniques used in the field and the laboratory. I supplement my observations of knowledge practices with a review of recent literature on the biological control of soilborne plant disease. Additionally, I incorporate interviews with some of my collaborators to understand why certain practices are used over others. I find that changes in how soilborne plant disease is known do not guarantee a change in how agricultural production is done. On the contrary, new ways of knowing disease can even help agriculture stay the same when they are cited as potential mechanisms by which incumbent agricultural practices take effect.

In what follows, I first outline the theoretical foundations of this chapter, which center on Annemarie Mol's (1999, 2002) method of observing objects as they are done in practice, or "enacted." Thinking with Mol, I identify three sets of practices for distinguishing disease. I call the versions of disease that they enact *symptomatic* disease, *pathogenic* disease, and *microbiomic* disease (Figure 3.1). Along the way, I describe how each version contends with the shifting realities of

specialty crop agriculture in the phaseout of methyl bromide. In the following section, I show how the multiplicity of practices for knowing soilborne plant disease has been incompletely managed, such that two versions of disease now coexist in tension without either consistently dominating the other. I then recount an instance where this multiplicity was instrumentalized in deference to an existing agricultural practice—a stance that I call "ambivalence." Finally, I theorize ambivalence as a style of ontological politics that troubles calls within STS to seek more relations always.

Enacting and interfering

In *The Body Multiple*, Mol (2002) presents a method for studying the entanglements between ontology and epistemology. By observing medical practices in a hospital setting, Mol found that what seemed at first to be a singular disease called atherosclerosis took on multiple forms depending on how it was diagnosed or treated by medical professionals. Atherosclerosis under the microscope was different from atherosclerosis in the clinical interview was different from atherosclerosis in the bypass surgery. It is tempting to resolve this tension by presuming a singular object, "atherosclerosis itself," at the center of multiple perspectives. But Mol rejects this logic, claiming that it falls back on subjectivist explanations that fail to take seriously the materiality of objects. As an alternative to what she calls "perspectivalism," Mol proposes attending to the practices through which objects are "done" or "enacted." An object enacted in a variety of practices can thus be more than singular—it can have *multiple ontologies*. In this framing, objects are not given, but rather "brought

into being, sustained, or allowed to wither away in common, day-to-day, sociomaterial practices" (Mol, 2002, p. 6). Instead of interpretations on the one hand and the "object-in-itself" on the other, Mol's ethnography observes their shared point of origin in practices. It is a "praxiography."

Against the Western, modernist logics that presume a singular reality of objects gradually "revealed" by science, the main intervention of a praxiographic approach is to demonstrate that objects are multiple, and thus, that their reality is always up for debate. Objects appear singular *in spite of* their multiplicity, and only at the resolution of ontological contestations that make realities commensurable ("coordination") or justify the local suspension of one reality for another ("distribution"). Mol is careful to point out that multiplicity does not imply plurality, however. While atherosclerosis is "more than one," it is also "less than many," because the practices that enact atherosclerosis depend upon each other (Mol, 2002, p. 55). Statistical norms in the blood analysis inform clinical diagnosis, and clinical diagnosis informs statistical norms. The reality of non-disease objects is also at stake. Because the normal ranges of hemoglobin are known to differ between men and women, statistical norms of hemoglobin levels inform sex difference, and sex difference informs statistical norms. In this way, seemingly distinct objects may *interfere* with one another (Mol, 1999). A practice for diagnosing disease becomes a practice for enacting differences in other parts of reality. If reality itself is thus open for debate, then the contestations and interferences revealed in a praxiographic analysis form the basis for a new mode of politics. It is an "ontological politics" (Mol,

1999, 2002), which does not operate between competing subjects, but between multiple, partially overlapping objects.

What gets left in the background of Mol's approach is the multiplicity of other practices that are engaged in interference with knowing disease. Whereas Mol's interferences are located in the hospital and the lives of its patients, Bruno Latour (1988) provides a framework for thinking about the interferences between scientific ideas and broader societal-scale concerns. In *The Pasteurization of France*, Latour (1988) examines the seemingly transformative power of the Pasteur Institute's scientific ideas, showing that what was commonly attributed to the goodness of the science or the "genius" of Louis Pasteur could instead be attributed to the alliances forged between Pasteurian scientists and a broad array of actors. Pasteurian science, weak on its own, became strong by "translating" the various concerns of the public hygiene movement, medical professionals, and French colonial interests into matters pertaining to the microbe—matters over which the Pasteurians, with their serums and vaccines, could then claim authority. According to Latour, the universalization of Pasteurian science is the exception to the rule. The Pasteurians' influence grew not because everybody believed in its scientific principles, but because "the hygienists believed them and forced everybody else to *put them into practice*" (Latour, 1988, p. 54, emphasis in original). There it is again: the materiality of practice. But unlike the focus of Mol's praxiography, these are not knowledge practices. They are social, technical, and managerial practices that are informed—or perhaps *interfered in*—by knowledge practices. Latour (1988) demonstrates that the scope of a praxiographic

analysis can be broadened to include an object's allies. In contrast to Mol's (2002) praxiographic approach, which starts with knowledge practices and notes their interferences, Latour's (1988) approach emphasizes the dynamic configuration of knowledge practices within heterogeneous networks of practice.

However, knowledge practices are not always so privileged. While Pasteurian science acted as an "obligatory passage point" (Callon, 1984), there are other networks wherein knowledge practices do not occupy such a central position. Many scholars in Science and Technology Studies (STS) have followed the enacted realities of disease through to their handling in the policy sphere. This tradition shows that science and policy, and a range of practices that span between them, all have a role in the ontological politics of disease. For example, Ingunn Moser (2008) considers Alzheimer's disease. Moser explains that the dominance of pharmaceutical treatments over care-based treatments is due not only to the practices through which Alzheimer's is known, but to the "performativity of parliamentary politics" that approves pills but not caregivers for public healthcare reimbursements (pp. 106-107). Later work, especially on infectious animal diseases, has examined the difficulty of sustaining multiple ontologies in policymaking sites (Law & Singleton, 2014; Mather, 2014). Policymakers tend to presume that there exists a singular reality upon which a general policy may be built. While a heterogeneous reality of disease would appear to the praxiographer as the expected outcome of multiple knowledge practices, it appears to the policymaker as a gap in scientific knowledge (Schrader, 2010). Policymakers may hold off on policy action until a "consensus" has been reached, resulting in policy

inaction (Oreskes & Conway, 2011). Alternatively, they may attempt to privilege one reality to the exclusion of others (Law & Singleton, 2014). In either case, the object multiple is made to be singular again.

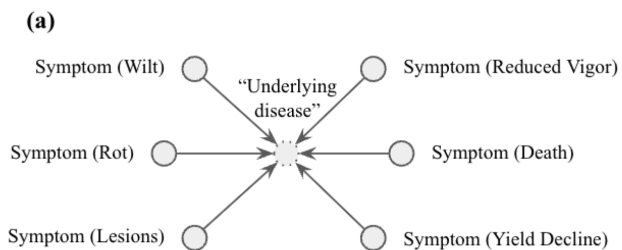
Aside from policy, there may be other kinds of practices interfering with the realities of disease. Julie Guthman (2019), writing on the California strawberry industry's attempts to repair itself in the wake of the methyl bromide phaseout, provides an interesting example. Guthman found that agricultural scientists tended to assume the emergence of novel strawberry diseases was linked primarily to the spread of pathogens and dismissed other explanations, including those invoked by biological control, as speculative and "of no practical importance." It is not that these scientists could produce evidence to settle the matter for good; the science had simply not been done. While alternative realities of strawberry disease have not been refuted *in theory*, they were discounted *in practice*—that is, by the technical routines and infrastructural commitments that comprise the engines of agricultural production. In this sense, the practical constraints of the specialty crop industry interfere with the knowing of disease, shaping its possibilities. To understand how objects are enacted, then, it may also be necessary to understand how their enactments are configured within networks that include production practices.

Knowledge practice

Configurations with(in) production agriculture

Symptomatic disease

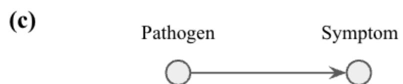
(a)



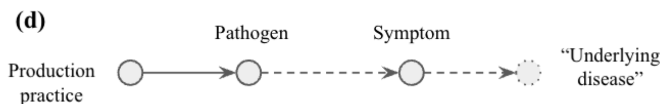
(b)

Pathogenic disease

(c)

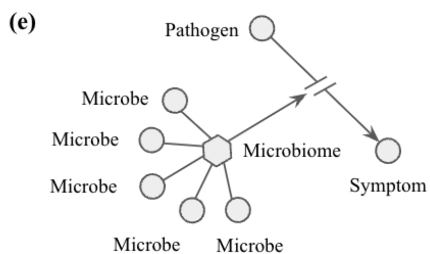


(d)

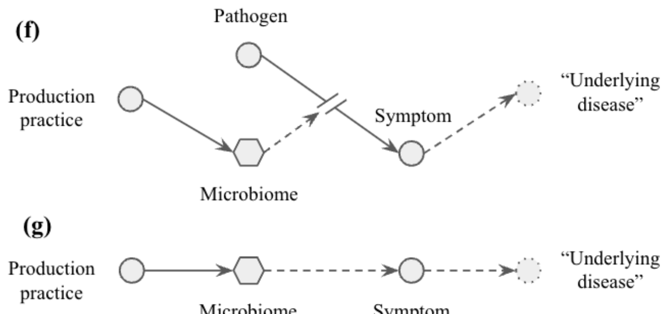


Microbiomic disease

(e)



(f)



(g)

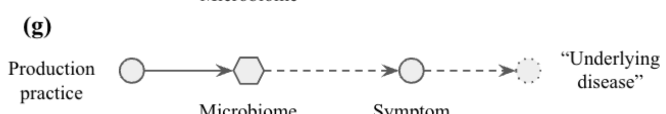


Figure 3.1. Sociomaterial knowledge practices enacting multiple ontologies of soilborne plant disease in the agricultural sciences, and their configuration with(in) production agriculture. Knowledge practices are unbracketed demonstrations of the relations that constitute the ontologies of disease. Production agriculture brackets some objects and their relations (dotted lines)—especially those already established through knowledge practices—to instead validate or legitimize the relation between a production practice and the presumed "underlying disease."

Symptomatic disease

Plant disease: you know it when you see it. Or do you? Even through visual inspection, plant disease appears multiple. For example, strawberry disease may look like wilt symptoms. It may also look like stunted plant growth or yellowing leaves. Sometimes it looks like the browning of crown tissues, but this is only apparent when the plant is cut open. While different pathogens are associated with different disease symptoms in strawberry, their symptoms can look strikingly similar. Disease may also look like a decline in crop yield. In the push for alternatives to methyl bromide, it could be argued that yield loss is the disease ontology that the specialty crop industry knows best. The 2023 request for applications for the USDA Methyl Bromide Transition Program prioritizes projects that control "economically important pests" and encourages applicants to include a "focused economic analysis" of the costs and returns of implementation. I refer to soilborne plant disease enacted through these everyday visual inspections of plants as *symptomatic disease* (Figure 3.1a).

The methods sections of scientific articles on soilborne plant disease reveal how disease is practiced. While some descriptions consider disease to be self-evident, others delve into the particularities that enact disease, a move that Mol (2002) calls “unbracketing.” To unbracket knowledge practices is to lay bare the simultaneously social and material elements that go into the doing of objects. For example, here is the unbracketing of “disease incidence”: “Disease incidence (%) was calculated using the formula: (number of dead plants/sum of plants in each replicate) × 100” (Liu et al., 2022). Disease incidence is thus enacted as the aggregate of a binary condition—diseased or healthy. But how is the diseased state enacted? In most enactments of disease incidence, the diseased state is assumed to be so obvious that it warrants no explanation. But in this article, the authors continue: “Clear pathological symptoms of *Fusarium* wilt were noted before the watermelon plant died, which mainly manifested as severe infection of the vascular bundles and wilting of the leaves” (Liu et al., 2022). Thus, disease incidence is shown to rely upon visual cues, making it a form of symptomatic disease. In other methods, disease is enacted on a scale of “disease severity.” This can be done by reference to a visual aid—for instance, photos of plants at different stages of disease development, ordered on a linear scale by the similarity of their symptoms (e.g. Mendes et al., 2011). Or it may be enacted by measuring plant biomass and calculating its deviation from a standard. For example, van Agtmaal et al. (2015) report the following method: “Roots were visually examined for root rot severity according to van Os et al. (1998) using an arbitrary disease index ranging from 0 to 5, where 0 = no root rot, 1 = 1–20%, 2 = 21–40%, 3 =

41–60%, 4 = 61–80%, and 5 = >80% root rot, i.e., relative loss of healthy root mass induced by infection, compared to the corresponding healthy root systems." The use of the word "arbitrary" in the preceding quote demonstrates that scientists themselves recognize disease scoring as a sociomaterial practice that functions primarily to make their science doable. Rather than being given to them, it requires them to make a leap of faith. The saving grace is that they make this leap together with their colleagues and predecessors, including "van Os et al. (1998)," such that it becomes standard practice. It is only through these resolutions that disease symptoms come to stand in for "disease itself," while the heterogeneous social and material elements that allow for this translation, like biomass and the trained eye, get bracketed once again.

Yet, there is more to disease than visual symptoms. It is telling that I never encountered symptomatic disease during my participation in the project, at least to my knowledge. This is to make two points. First, it underscores that disease symptoms are not self-evident to the untrained eye. The brown, wilted strawberry plants that I noticed on the lab bench may have been wilted because of disease or because they had not been watered in a long time. The two conditions share a common reality after all, as wilt diseases are known to constrict water uptake by plant vascular tissue. Second, it shows that symptomatic disease is not always worth training the eye for. The main problem with symptomatic disease is that it is difficult to study in the laboratory. As I alluded to in my introduction, it is not straightforward to reproduce disease symptoms in strawberry plants in a greenhouse setting. They simply do not develop, except when manipulated through specific techniques. One of

my collaborators informed me that it is possible to induce disease in strawberry plants by penetrating the plant tissue with a pathogen-infested toothpick. This is where a praxiographic analysis breaks most clearly from perspectivalism. Disease not only looks but *behaves* differently under different practices.

But then the challenge lies in making the case that disease-infested toothpicks or their absence have meaningful analogues in production practices (Figure 3.1b). The production setting can be better represented by a field trial. However, field trials are time- and labor-intensive. A researcher needs to be fairly confident in their experimental design before rolling it out in a full-blown field trial. And even with all the necessary precautions, field trials can still go awry for many reasons: unseasonable weather, broken irrigation tape, and miscommunications or misaligned priorities between the researcher and the grower, to name a few. In the end, it can be ambiguous whether the development of disease symptoms in treated field soils was due to ineffective treatments or due to any number of potential interferences.

In addition to these technical and biological limitations, symptomatic disease has another weakness: it is *always already too late*. My collaborator tells me, “Once you observe disease in the field, there’s basically nothing that you can do to prevent disease progression other than continuing to fertilize and irrigate to promote a healthy crop to reduce plant stress.” Growers can also slow the spread of the disease and the accumulation of the pathogen by killing symptomatic plants with herbicides—a practice known as crop termination (Holmes et al., 2020). And they can plan ahead for the next crop with the knowledge that their soils are infested—but only if their

operations can survive the losses from the current one, and only if they can expect to maintain land tenure. Thus, however "real" symptomatic disease appears to be, its conditions of possibility make it impractical for specialty crop growers and scientists alike.

Agricultural scientists would prefer to work with a version of disease that is not so unruly, so they can anticipate treatment effects before investing in field trials. And growers want practices for identifying disease before the wilting, stunting, and damping-off. Fortunately for both of these groups, a second version of soilborne plant disease can be enacted without the participation of the plant at all.

Pathogenic disease

On my second day in the research team, I learned a method for quantifying the soilborne population density of *Macrophomina phaseolina*, a strawberry pathogen that causes a disease known as charcoal rot. I blended soil samples into a slurry with a bleach solution, which acted as a chemical filter for sclerotia, the hardened and stress-tolerant vegetative bodies of *Macrophomina*. After being sieved, rinsed, and diluted in water, the resulting soil solution was plated onto Petri dishes, where a selectively antibiotic agar gel further narrowed down the remaining microbial survivors. In a few weeks, someone would count the number of "colony-forming units" visible to the naked eye as white clusters of hyphae, the thread-like structures that comprise the non-reproductive bodies of multicellular fungal species. This configuration of biological knowledge and technical know-how is called a *cell culture assay*. In my

observations and in the literature, it was used for the detection and quantification of fungal pathogens that comprised the major pathogens of strawberries in the California specialty crop industry. Apart from innovations like the use of bleach solution for *Macrophomina*, the cell culture assay is a time-honored practice for knowing pathogens, with its roots in the beginnings of the field of bacteriology. In the past two decades, molecular tools have been developed as an alternative for quantifying pathogens via strain-specific DNA markers, but even so, the metrics of DNA density are converted back to "colony-forming units per gram soil" using correlations that have become conventional in the field of plant pathology (e.g. Bilodeau et al., 2012).

I use *pathogenic disease* to describe the enactment of soilborne plant disease as the presence or abundance of particular pathogens. Pathogens, after all, have been considered an obligatory passage point to disease ever since Pasteur demonstrated that anthrax, a disease of cattle, developed if and only if cattle were infected with the anthrax bacillus (Latour, 1988). Whereas plant pathologists understand symptomatic disease to be built upon an axiomatic relation between a symptom and its supposed referent—that is, "disease itself"—pathogenic disease is built upon empirically derived relations to symptoms (Figure 3.1c). The associations upon which they are built are not so different, however; they are both constituted in practice. Guthman (2019) describes how the plant pathologist Harold Thomas established *Verticillium dahliae* as a strawberry pathogen in 1931. After culturing a population of *Verticillium* in the lab, Thomas inoculated half of the plants in an experiment with a serum containing live *Verticillium*—which he must have initially been isolated, purified, and

cultured—and found that a substantial proportion of those plants developed disease symptoms while none of the controls did. Through these practices, pathogen abundance comes to stand in for disease symptoms (Figure 3.1d).

In the contemporary literature on ASD and other soilborne disease control methods, pathogenic disease is used more frequently than symptomatic disease as the variable for judging the efficacy of treatments. Some studies acknowledge that this is the result of a translation, and that as such, it may not always line up with symptomatic disease. Even so, many of these same studies slip up in ways that reveal the solidity of pathogenic disease. Some equate "disease suppression" with a decline in soilborne pathogen abundance (e.g. Trivedi et al., 2017), and some even use pathogen presence to define "diseased soils" (e.g. Liu et al., 2022), as if plant disease did not require plants.

Growers would prefer practices for identifying disease ideally before their plants even enter the soil, and pathogenic disease offers that. Perhaps the popularity of this plant-free disease enactment across the specialty crop industry can also be understood as the legacy of fumigation having been the primary method of disease control for so long. After methyl bromide, the specialty crop industry is not looking for just any solution to its disease problems; it is specifically looking for methyl bromide alternatives. This framing reveals an ideological and infrastructural commitment to solutions that operate in much the same way as methyl bromide fumigation did—that is, applied to soil before planting and used in conjunction with disease-free nursery stock, thereby closing all passage points for soilborne strawberry

pathogens. The prevalence of pathogenic disease in scientific studies says less about agricultural scientists' abilities to do holistic research and more about how they contend with "economic realities," especially those of the growers for whom they produce solutions and the federal research programs that fund them to do so.

Microbiomic disease

In 1959, the plant pathologist James Menzies became the first to document the phenomenon of disease-suppressive soils. He described a soil that slowed the development of common scab of potatoes even when inoculated with its causative pathogen, *Streptomyces scabies*. He also found that these soils lost their suppressiveness after being sterilized. Noting that little changed in the process of sterilization except for the inactivation of its biota, Menzies (1959) concluded that the cause of suppressive soils must be biological. And since the pathogen was not introduced until after sterilization, suppressiveness must have specifically been linked to soil biota *not including the pathogen*—in other words, it was attributed to the soil microbiome (Figure 3.1e).

I use *microbiomic disease* to describe the enactment of soilborne plant disease as a soil microbiome that fails to suppress disease in the presence of a pathogen. While a suppressive soil microbiome cannot definitively prevent disease symptoms from developing, it can reduce disease incidence across the field to manageable levels. Suppressive soil microbiomes not only reduce the soil pathogen population; they may also outcompete the pathogens for space on the root surface, which is

thought to be necessary for soilborne plant disease to develop (Trivedi et al., 2020). Whereas pathogenic disease made it possible to study soilborne plant disease without plants, microbiomic disease makes it impossible—or at least, less convincing—to study pathogenic disease without also considering the microbiome.

A recent variation of microbiomic disease stems from the adoption of so-called "omics" methods into plant pathology. This variation relies on notions of the microbiome that began with the introduction of "molecular tools," DNA- and RNA-based methods introduced in the 1990s to characterize the genetic makeup of biological entities. Molecular tools are often described as a revolution for microbiology because they allowed researchers to survey a wider diversity of microbiota from the environment. A major limitation of the more classical culture-based tools was that the vast majority of soil microbes could not be reliably grown in cell culture. While sequencing was initially expensive and could only be used to sequence one DNA strand at a time, technological improvements at the turn of the century collectively known as "high-throughput sequencing" accelerated the pace and lowered the cost of sequencing. This made it feasible to sequence whole communities of organisms all at once, creating a class of methods known as metagenomics. Metagenomics joins other "post-genomic" methods—including transcriptomics, proteomics, and metabolomics—as a standardized package (Fujimura, 1988) that grew out of the Human Genome Project and the data-driven science that it propelled (Falkenberg et al., 2023). Omics technologies made it possible to know who is present in a community and what the community is doing as a whole, but they cannot

identify the functional contributions of any individual microbial taxon in the way that culture-based studies do. Thus, while omics technologies are more holistic in some ways, they are also disintegrative in others. Omics technologies have not, for example, yielded a consensus regarding *which* microbes are responsible for disease suppressiveness in soils. The "omic" in "microbiomic disease" is a nod to the omics technologies, especially metagenomics, that have contributed to this highly technical enactment of soilborne plant disease.

In the literature, Mendes et al. (2011) were among the first to use metagenomics to confirm that soils suppressive to *R. solani* were associated with distinct microbial communities that declined after gamma irradiation—a thorough method for killing microbes (Weller, 2002). While previous iterations of molecular methods also found traces of distinct microbial communities in suppressive soils, metagenomic methods allowed Mendes et al. (2011) to identify specific bacterial taxa associated with disease suppression. Studies like these helped to define a research program for translating disease suppressiveness to specific microbial taxa or properties of the microbiome, like its biodiversity, overall composition, or emergent qualities of biotic interactions. Recent studies even describe changes to the soil microbiome under ASD without measuring disease symptoms or specific soilborne pathogens at all (e.g., Achmon et al., 2020; Vincent et al., 2022).

Omics technologies may have helped operationalize microbiomic disease as a way of knowing soilborne plant disease, but microbiomic disease is a relative newcomer, still finding its place within agricultural science. Crucially, microbiomic

disease in its current state cannot guide new interventions in production settings. While experiments have shown that the disease-suppressive microbiomes are transferable between soils through the "inoculation" of conducive soils with suppressive soils in a ten-to-one ratio (Weller et al., 2002), this nonetheless requires too much soil to be economically feasible. On the other hand, experiments on isolated microbial populations have proven more fruitful. Numerous studies have linked disease suppression to individual microbial taxa, such as *Bacillus* and *Trichoderma* (Weller et al., 2002). Yet, the performance of individual taxa in field settings has proven inconsistent (Holmes et al., 2020; Mazzola & Freilich, 2017), suggesting that perhaps a consortium of microbiota that mutually sustain each other may be necessary for longer-lasting disease suppression. Thus, the strength of microbiomic disease as an ontology of knowing in agricultural science and management today should not be overstated.

Notwithstanding its current limitations, microbiomic disease occupies a similar niche as pathogenic disease, in that both can be enacted without the participation of the plant. Many of the advantages that pathogenic disease displays over symptomatic disease therefore reappear as promises that microbiomic disease makes today. Yet, as I will show in the next section, microbiomic disease does not replace pathogenic disease. In the phaseout of methyl bromide, they instead coexist in tension with each other.

A tense coexistence

Symptomatic disease, pathogenic disease, and microbiomic disease make up the cast of this praxiography. Because they are enacted through distinct practices, it is not a given that they will agree with one another—by what measure should they be said to agree? In this section, I use Mol's (2002) concepts of "coordination" and "distribution" to explore how clashes between different versions of soilborne plant disease are managed, either by making realities commensurable or by assigning them to non-overlapping settings. Within this framing, the coherence of disease realities is neither given nor guaranteed. It is *achieved* by scientists as the resolution to an ontological contestation, and yet, not every contestation resolves or stays resolved. In the phaseout of methyl bromide, I find that microbiomic disease has not been made commensurable to pathogenic disease, as if they measure the same thing, nor has microbiomic disease fully replaced pathogenic disease, as if in a paradigm shift. Rather, the work of coordination and distribution remains incomplete. As a consequence, where pathogenic disease once dominated, microbiomic disease now sits alongside it in a tense coexistence.

Coordination is one of two primary means by which the multiplicity of objects is managed. It relies on the notion that while objects-in-practice may be different, they are not necessarily incommensurable—that is, they can be brought together through what Mol (2002) calls "coordination work" (p. 66). In the preceding sections, coordination is what allows me to name only three versions of disease but more than three knowledge practices. Naming three versions of disease is meant to illustrate the ease with which agricultural scientists coordinate across some diseases-in-practice,

which are cohered into "versions," but not across others.

Sometimes coordination is straightforward. For instance, every knowledge practice includes a coordination technique at its core. In the biomass-based disease index from descriptions of symptomatic disease, a root biomass loss of 21% is made equivalent to a root biomass loss of 40%, and a 40% root biomass loss is the same regardless of whether the loss is due to fewer roots or shallower roots. Coordination also occurs when researchers translate diseases-in-practice to answer to a common measure. For example, while my engagement with the research team began with cell culture assays, my primary responsibilities changed to focus on quantifying pathogen DNA in soil samples using molecular tools. I would then follow up by converting these DNA quantities into counts of colony-forming units using pathogen-specific calibration curves that had been derived from prior experiments (e.g. Bilodeau et al., 2012). This allowed us to enact pathogenic disease in some experiments as if it had been done in a cell culture, but using a different practice entirely.

Coordination techniques make difference out of differences, sustaining difference when it is deemed important for the knowing of an object and flattening difference when it is not. The singular noun in each of "symptomatic disease," "pathogenic disease," and "microbiomic disease" is meant to reflect this flattening of difference. Yet Mol (2002) reminds us that while disease is "less than many," it is also "more than one" (p. 55). It is not obvious how one should coordinate across differences between these three versions of soilborne plant disease. It does not make sense to create calibration curves to translate between them. If anything, correlations

between them are treated as a welcome surprise. Liu et al. (2022), for example, report that "the abundance and proliferation of *F. oxysporum* [a soilborne fungal pathogen] after planting significantly and positively ($p < 0.05$) correlated with the disease incidence (Fig. 2c and d)." In the literature, these correlations are surprising enough to warrant statistical support and multiple subfigures. Even after coordination has done all the work that it can do, three versions of disease remain.

Distribution is the other primary means for managing multiplicity. It relies on the notion that when multiple versions of an object exist, they are not necessarily put into practice in the same settings. Mol (2002) uses a spatial metaphor to convey this: realities are "distributed over different sites" such that they no longer contradict each other (p. 115). In the literature on soilborne plant disease control, distribution indeed occurs across *spatial compartments*. For instance, a symptomatic plant does not necessarily contradict a suppressive soil microbiome, as the soil microbiome merely serves as the "first line of defence" (Trivedi et al., 2020). Similarly, distribution may occur across *chains of probabilistic cause and effect*. Studies on disease control posit, for example, that "the quantity of the pathogen in root tissue may be predictive of spread to crown tissue, and subsequent wilt development, later in the growing season" (Muramoto et al., 2016). Here, pathogen density is a risk factor for disease symptoms but does not decide them. Distribution may also occur across *types of disease*. Some types of disease are difficult to diagnose using visual methods. Instead, their pathogens must be cultured in selective media, or—for pathogens that cannot be cultured—identified using molecular tools. In these cases, pathogenic disease is

enacted over symptomatic disease. Other types of soilborne disease, like apple replant disease (Mazzola et al., 2012) and Prunus replant disease (Browne et al., 2018), are thought to be caused by a consortium of microbial taxa under the banner of a "disease complex." Lacking a single known pathogen, these types of diseases can only be studied through an analysis of the whole microbiome (e.g., Khan et al., 2022; Dundore-Arias, 2023). Finally, as highlighted in the preceding sections, distribution may occur across different *conditions of possibility*. Plants are required for symptomatic disease, but not for pathogenic disease. This makes pathogenic disease particularly useful for pre-plant soil tests, which can help a grower decide what kind of disease management strategy to use before the growing season begins.

In practice, these distributive criteria appear as rules and routines that are shared across the community of practitioners. In the hospital, for instance, Mol (2002) observed well-established "indication criteria" that determine which patients are given which treatments (p. 101). Clashes are thereby avoided in practice. In the agricultural sciences, however, the rules appear to be less established. Thus, contestations arise in the form of research questions such as: Can this pathogen cause *asymptomatic* infections (Henry et al., 2019)? (If so, symptomatic disease loses some ground to pathogenic disease.) Can the endosphere microbiome, which resides inside plant tissue, act as a *second* line of defense against disease (Carrión et al., 2019)? (If so, pathogenic disease loses some ground to microbiomic disease.) Can this disease complex be narrowed down to a manageable number of taxa, which is thereafter treated in practice as if they *were* the causative agents (Khan et al., 2022)? (If so, the

lines between pathogenic disease and microbiomic disease are blurred.) These ongoing contestations demonstrate that the realities of soilborne plant disease are incompletely distributed. Their borders push, pull, and overlap.

The multiplicity of objects is not apparent when different versions are made to answer to a common measure, nor when only one version exists at a time. But in the ongoing contestation between pathogenic disease and microbiomic disease, neither of these conditions has been met. The phaseout of methyl bromide has changed the conditions of possibility over which the multiplicity of disease had once been neatly distributed. Pathogen *presence* is no longer a meaningful way to measure or manage the risk of soilborne plant disease because pathogens have become endemic in soils seemingly everywhere and methyl bromide fumigation—growers' go-to method for keeping pathogens at bay throughout the latter half of the twentieth century—has been banned. For growers today, it is no longer a question of whether to exclude pathogens. Like it or not, pathogens are here to stay. Growers have to search for other ways to close off the possibility of disease. One approach is to accept that pathogens are likely to colonize the plant tissue, which informs efforts to breed new crop varieties for disease resistance. Another approach, which is more specific to this chapter's focus on soilborne disease, is to shore up the plant's defenses in the surrounding environment through modifications to the soil microbiome. This raises an important question: what happens when an abundant pathogen meets a suppressive soil microbiome?

Until these questions are resolved, the tense coexistence of pathogenic disease and microbiomic disease will be difficult to ignore. For example, a recent study on the population dynamics of *Fusarium oxysporum*, the pathogen linked to Fusarium wilt in strawberry, acknowledges in closing that "if general suppression by the microbial community also occurs, then increased inoculum density may not correlate with disease severity" (Henry et al., 2020). Microbiomic disease has become just relevant enough to agricultural production that it is able to undercut pathogenic disease in places where it once stood unchallenged. In the following section, I explore the implications of this tense coexistence for how soilborne plant disease gets to be managed.

The ontological politics of ambivalence

During my engagement with the research team, I worked with one of my collaborators on a project to investigate the effects of cover cropping treatments on charcoal rot of strawberry, a disease linked to the pathogen *Macrophomina phaseolina*. He suspected that charcoal rot would be reduced by all of the treatments except one, and he included that treatment with the specific intent to dispel a rumor. According to my collaborator, the treatment in question had been presented to growers after a sparsely replicated greenhouse trial showed that it led to small improvements in disease outcomes. Since then, it has spread widely among growers through word-of-mouth. While the cover crop treatment was neither proprietary nor costly, it did deprive growers of the opportunity to plant other crops that could control

disease more effectively or simply bring in revenue, which this treatment did not. Using a field trial, my collaborator found that the treatment in question was not associated with any significant difference in the abundance of the pathogen in soil. So far, things were not looking good for the treatment. But before making a definitive statement, my collaborator suggested that we conduct an additional analysis to examine the composition of the soil microbiomes across these treatment groups. Through a metagenomic analysis, we found that the cover cropping treatment led to a subtle shift in the soil microbiome. It is hard to say what this means for soilborne plant disease. As I alluded to earlier, scientists do not yet have a predictive understanding of the connections between soil microbiomes and disease suppression. Thus, my collaborator and I had no reason to suspect that the pathogen would behave any differently within the treated microbiome. But my collaborator, as a scientist who must also confront the needs and desires of growers, understood that a shift in the microbiome whatsoever could be read as conflicting evidence. That is, a shift in the microbiome could be a shift towards disease suppressiveness—never mind the fact that it could also be a shift towards disease conduciveness (Figure 3.1g). If the microbiome had been unchanged, then perhaps we could confidently refute this cover crop treatment. Or if we had inoculated the soils with a pathogen, planted the soils with strawberries, and surveyed their disease symptoms, then perhaps that could provide more conclusive evidence as well (Figure 3.1f). As it stood, however, our study would not be the one to settle the debate.

When I mentioned this outcome to other scientists at a plant pathology

conference, they were not surprised. In response to the claim that the cover crop treatment suppresses charcoal rot, they agreed that "that doesn't sound right." But they were also understanding of my collaborator's abundance of caution. Another one of my collaborators who attended the conference added that conflicting recommendations from scientists may ultimately lead to a loss of growers' trust. Before making a claim that contradicted popular belief, then, my collaborator on the cover cropping project would need evidence capable of either resolving the contestations between pathogenic disease and microbiomic disease or getting around them by measuring disease symptoms (Figure 3.2).

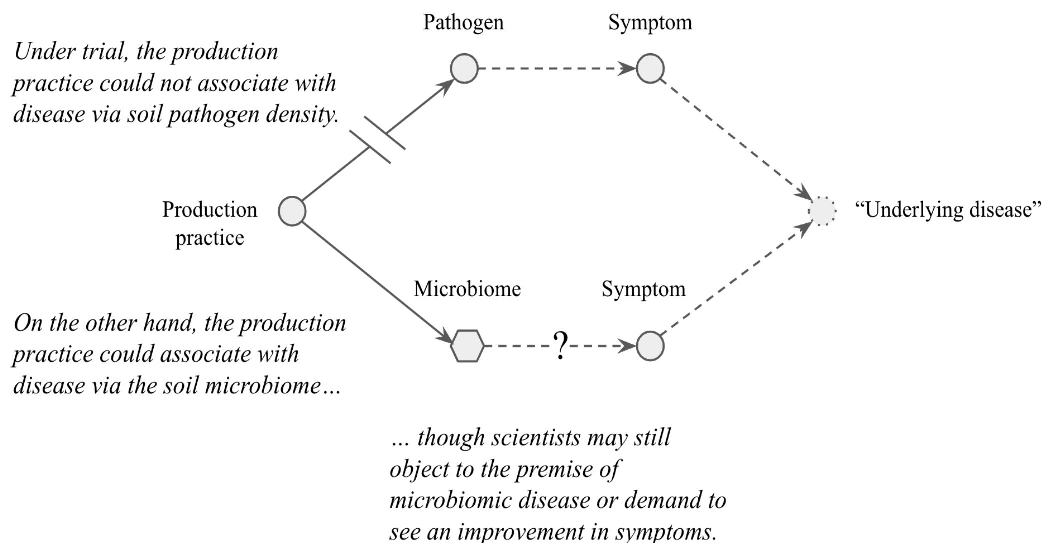


Figure 3.2. An ontological contestation between pathogenic disease and microbiomic disease and its temporary resolution in deference to an existing disease management practice. The practice in question could have been connected to "underlying disease" through either pathogenic disease or microbiomic disease.

What this anecdote shows is that introducing more ontologies of disease does not merely open up new possibilities for managing disease. It may also enshrine existing practices by giving them a new plausible mechanism of action. What is at stake in ontological clashes about mechanisms? More than nothing, but less than everything. Researching potential mechanisms may help to tailor disease management practices to specific conditions, but it usually does not refute them altogether. The configuration of a mechanism within production practices makes it neither the beginning nor the end of a chain of associations. Inserted into the middle, it becomes instrumentalized to reinforce or provide a backup plan to what may otherwise be only tentative connections between a production practice and its intended effect. "The consistency of an alliance is revealed by the number of actors that must be brought together to separate it" (Latour, 1988, p. 206), and it does not matter where, exactly, those separations happen.

Ontological multiplicity implies difference, and difference implies the possibility of change in the constitution of objects—this is ontological politics. Yet even when multiple ontologies are enacted, as in the knowing of soilborne plant disease, little may change about the applications. Prior work on ontological politics has linked this stagnation to the assertion that an object "can only be what it obviously is" (Woolgar & Lezaun, 2013). Here I want to make a slightly different point. It is not that a belief in any particular reality-as-we-know-it underlies a commitment to a particular kind of application. Rather, a particular kind of

application may incorporate multiple ontologies into its logics, ambivalent to their possibilities as long as they pose no threat.

Ambivalence is not the same as ignorance. Ignorance, as theorized by the historian Frank Uekötter (2014), entails the dominance of one form of knowledge over another. In the history of soil fertility management, chemical knowledge came to dominate over microbiological knowledge, tying the rise of industrial agriculture to a post-war boom in agrochemical manufacturing (Uekötter, 2014). The coexistence of pathogenic disease and microbiomic disease cannot be analyzed using this same framework, as neither consistently dominates over the other. Instead, their dynamic more closely resembles that depicted by Anna Krzywoszynska (2019) when she describes the role of soil scientists in a community of farmers who were experimenting with sustainable soil management practices. Scientists there did not provide recommendations so much as they legitimized practices that farmers were already using. Ambivalence and ignorance have something in common: they are both made possible by ontological contestations, known outside of a praxiographic analysis as *uncertainty*. This enables them to dismiss a particular ontology at the crucial moment—but unlike ignorance, which confronts that moment early and often, ambivalence saves that moment for the very end. This makes ambivalence a relatively easy stance to adopt. If growers and agrochemical manufacturers were able to get behind a particular solution even when it consistently ignored a particular reality, then what is preventing them from tentatively enlisting multiple realities and cherry-picking the one that is most compatible with their proposed solutions at any given

moment? Ambivalence, after all, is less of a feat than ignorance.

Admittedly, the ontological stakes of ambivalence in the cover cropping example are relatively low. Cover cropping is done in sequence with fumigation, not instead of it, so it cannot challenge major presuppositions about how disease should be managed. To fully appreciate the stakes of ambivalence, then, one can look to the literature on disease control by fumigants and their biological alternatives. This time, it requires reading beyond the methods sections where knowledge practices are foregrounded, and turning one's attention to the discussion sections where specific interferences are proposed. For example, the premise of Shi et al. (2022) reads at first like a paradox. The authors set out to examine the impacts of triple fumigation on the soil microbiome in order to "provide theoretical guidance on its application" (Shi et al., 2022). How is it that chemical fumigants, brought into agriculture specifically to create conditions inhospitable to life, were now being considered as potential allies to biological control? In the discussion section, the authors' rationale becomes clear: "the low dosage of three fumigants used at one time effectively improved the control effect on pathogenic fungi such as *Fusarium oxysporum*, saving the control time [sic], significantly increasing the abundance of beneficial species such as *Bacillus* and *Trichoderma* and improving the potential disease resistance of soil" (Shi et al., 2022). (Lest one mistakes "potential disease resistance of soil" as a separate enactment involving disease symptoms, the authors clarify elsewhere that this is merely a restatement of the observed increase in *Bacillus* and *Trichoderma*.) One could attempt to pit this study against contrasting studies that find deleterious effects of fumigation

on the soil microbiome (e.g. Dangi et al., 2017), but that is not the point. It is, in fact, *plausible* that triple fumigation might nudge the soil microbiome toward that of a suppressive soil because all of the fumigants tested by Shi et al. (2022) are known to be less-deadly alternatives to methyl bromide (Rosskopf et al., 2016). This crucial fact allows the authors to cite both pathogenic disease and microbiomic disease as potential mechanisms by which an already-established production practice may be legitimized. Fumigation may work because it reduces pathogen density, or it may work because it makes soils more disease-suppressive. Either way, the reader is reminded that *fumigation works*.

By contrast, the challengers of well-established production practices are not in a strong position to be ambivalent, as claims about the potential mechanisms of effect are less convincing when the effect itself is still in question. Considerable research on ASD, for instance, is dedicated to understanding the conditions where it *does* work—including specific combinations of target pathogens, treatment durations, organic amendments, and soil textures (Shrestha et al., 2016). Many of these studies employ methods that characterize the soil microbiome, enacting microbiomic disease in addition to pathogenic disease. However, microbiomic disease does not tip the scales against fumigants—except on the rare occasions when specific microbiomes are first linked to disease symptoms and are then shown to be more closely associated with the alternative practice than with fumigation (e.g. Mazzola et al., 2015). Yet these are the exceptions that prove the rule. Microbiomic disease can uplift less-established practices only when it is first translated to symptomatic disease. Ambivalence

between microbiomic disease and pathogenic disease does not provide the same leverage here.

Ironically, even as ambivalence appears to level the playing field between multiple realities of disease, it also widens the gap between the production practices that the reality of disease is thought to inform. Moser (2008) recognized that even in a praxiographic analysis, the ontological is not necessarily privileged. In the mattering of Alzheimer's disease, the sphere of formal politics is a site where ontological multiplicity becomes singularized (Moser, 2008). In production agriculture, on the other hand, production practices and their infrastructural commitments play the role of "politics by other means" (Latour, 1988). The object, soilborne plant disease, is political, but so are the economic realities that it is enacted within. I find that microbiomic disease, despite offering an alternative way of knowing soilborne plant disease, does not turn the ratchet of progress. It does not create a paradigm shift or consistently dominate over any other ontologies. Despite its status as a reconfiguration of disease, it is subsequently configured into the same position—that of a mechanism—that has housed a pathogenic ontology of disease. There, microbiomic disease and pathogenic disease coexist in tension, to the ambivalence of incumbent practices.

Against this backdrop, ambivalence emerges as a style of ontological politics. While ontological politics requires contestations about the reality of objects, it does not have to begin and end with objects. Similarly, ambivalence concerns the ontological multiplicity of objects, but it does not preoccupy itself with them. It

instrumentalizes them.

Conclusion: "It could be anyway"

In *The Body Multiple*, Mol (2002) lays the groundwork for what would eventually be called the ontological turn in STS. Instead of privileging the social, as had been implied by the term "social constructivism," ontological approaches consider how technoscientific objects are done in practice (Sismondo, 2015). Its political stakes, however, were less obvious. Mol speculates thus: "In stressing ontological multiplicity this book lays bare the permanent possibility of alternative configurations. […] Medical practice is never so certain that it might not be different; reality is never so solid that it is singular" (Mol, 2002, p. 164). This foreshadows the slogan now used to convey the significance of the ontological turn, that "it could be otherwise" (Woolgar & Lezaun, 2013).

Contrary to this slogan, however, I find that an agricultural industry deeply committed to particular sociomaterial practices can use ontological multiplicity to its own benefit. The fact that agricultural science more commonly positions microbiomic disease as a launch pad for "precision farming" technologies that introduce specific beneficial microbes (e.g., Berg et al., 2020), rather than a justification for biodynamic farming principles that seek to enrich native biodiversity (e.g., Mazzola & Freilich, 2017), may have less to do with its methodology and more to do with its deference to industrial agriculture. Perhaps agriculture can be different, but it is conditional on more than simply the reality of soilborne plant disease. In the highly instrumental

research programs created in the wake of methyl bromide, I find that the slogan "it could be otherwise" is just one degree of separation away from its ambivalent twin: "it could be *anyway*."

To the extent that ontological multiplicity draws upon the STS tradition of relational materialism (Law & Singleton, 2014), this chapter troubles the heuristic in STS that relationality is inherently good. For one, some relations are toxic, and doing well becomes a matter of entangling oneself with certain toxicities to exclude even worse relations (Roberts, 2017). Similarly, the reconfiguration of soilborne plant disease to include microbiomic enactments reflects a more relational understanding of disease than was voiced by Guthman's (2019) interlocutors, yet it does not change much about the way production works. I propose that attending to knowledge practices and their configuration within production practices is one way to study whether ontological multiplicity, and the new relations that they entail, will lead to new possibilities or simply perpetuate existing certainties.

SUPPLEMENTARY FILES

Supplementary Files 1–3 are available as Data S1–S3 at https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecs2.3842.

REFERENCES

Achmon, Y., J. Claypool, J. Fernandez-Bayo, K. Hernandez, D. McCurry, D. Harrold, J. Su, B. Simmons, S. Singer, R. Dahlquist-Willard, J. Stapleton, J. VanderGheynst, and C. Simmons. 2020. Structural changes in bacterial and fungal soil microbiome components during biosolarization as related to volatile fatty acid accumulation. Applied Soil Ecology 153.

Allison, S. D., and J. B. H. Martiny. 2008. Resistance, resilience, and redundancy in microbial communities. Pages 149–166 *in* J. C. Avise, S. P. Hubbell, F. J. Ayala, editors. In the light of evolution: volume II: biodiversity and extinction. National Academies Press, Washington, D.C., USA.

Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of applied ecology 43:1223–1232.

Anthony, M. A., T. W. Crowther, S. Van Der Linde, L. M. Suz, M. I. Bidartondo, F. Cox, M. Schaub, P. Rautio, M. Ferretti, and L. Vesterdal. 2022. Forest tree growth is linked to mycorrhizal fungal composition and function across Europe. The ISME journal 16:1327–1336.

Antoninka, A., P. B. Reich, and N. C. Johnson. 2011. Seven years of carbon dioxide enrichment, nitrogen fertilization and plant diversity influence arbuscular mycorrhizal fungi in a grassland ecosystem. New Phytologist 192:200–214.

Averill, C., L. L. Cates, M. C. Dietze, and J. M. Bhatnagar. 2019. Spatial vs. temporal controls over soil fungal community similarity at continental and global scales. The ISME journal 13:2082–2093.

Bahram, M., F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Mundra, P. A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, and P. Bork. 2018. Structure and function of the global topsoil microbiome. Nature 560:233–237.

Barnett, D., P. Duffy, D. Schimel, E. Azuaje, R. Krauss, A. Thorpe, D. Gudex-Cross, and J. McCorkel. 2017. TOS science design for spatial sampling. National Ecological Observatory Network. https://data.neonscience.org/documents/10179/1723439/NEON.DOC.000913vB/81c29201-ac53-4c0a-afba-20f35de74ce4

Battelle Memorial Institute. 2018. NEON 16S and ITS sequencing standard operating procedure, v.2. Battelle Memorial Institute.

https://data.neonscience.org/documents/10179/2431540/BMI_markerGeneSequencingSOP_v2/f2f1228f-5c9b-a4dd-1bb6-e0f0dad4e8a8

Battelle Memorial Institute. 2018. NEON 16S and ITS sequencing standard operating procedure, v.2. https://data.neonscience.org/documents/10179/2431540/BMI_markerGeneSequencingSOP_v2/f2f1228f-5c9b-a4dd-1bb6-e0f0dad4e8a8

Battelle Memorial Institute. 2019. NEON 16S/ITS PCR amplification standard operating procedure, v.2. Battelle Memorial Institute. https://data.neonscience.org/documents/10179/2431540/BMI_PcrSOP_v2/e08612c4-54d7-4331-06f2-7268e8656d41

Battelle Memorial Institute. 2022. NEON DNA extraction standard operating procedure v.7. Retrieved https://data.neonscience.org/documents/10179/2431540/BMI_dnaExtractionSOP_v7/61204962-bb01-a0b9-3354-ccdaab5132c3

Berg, G., D. Rybakova, D. Fischer, T. Cernava, M.-C. C. Vergès, T. Charles, X. Chen, L. Cocolin, K. Eversole, G. H. Corral, M. Kazou, L. Kinkel, L. Lange, N. Lima, A. Loy, J. A. Macklin, E. Maguin, T. Mauchline, R. McClure, B. Mitter, M. Ryan, I. Sarand, H. Smidt, B. Schelkle, H. Roume, G. S. Kiran, J. Selvin, R. S. C. D. Souza, L. Van Overbeek, B. K. Singh, M. Wagner, A. Walsh, A. Sessitsch, and M. Schloter. 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome 8:103.

Bilodeau, G. J., S. T. Koike, P. Uribe, and F. N. Martin. 2012. Development of an assay for rapid detection and quantification of Verticillium dahliae in soil. Phytopathology 102:331–343.

Boettiger, C. 2015. An introduction to Docker for reproducible research. ACM SIGOPS Operating Systems Review 49:71–79.

Bradford, M. A., R. J. Warren II, P. Baldrian, T. W. Crowther, D. S. Maynard, E. E. Oldfield, W. R. Wieder, S. A. Wood, and J. R. King. 2014. Climate fails to predict wood decomposition at regional scales. Nature Climate Change 4:625–630.

Browne, G., N. Ott, A. Poret-Peterson, H. Gouran, and B. Lampinen. 2018. Efficacy of Anaerobic Soil Disinfestation for Control of Prunus Replant Disease. Plant Disease 102:209–219.

Busby, P. E., C. Soman, M. R. Wagner, M. L. Friesen, J. Kremer, A. Bennett, M. Morsy, J. A. Eisen, J. E. Leach, and J. L. Dangl. 2017. Research priorities for

harnessing plant microbiomes in sustainable agriculture. PLoS biology 15:e2001793.

Calabrese, J. M., G. Certain, C. Kraan, and C. F. Dormann. 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models: Stacking species distribution models. Global Ecology and Biogeography 23:99–112.

California Department of Food and Agriculture (CDFA). 2022. California Agricultural Exports 2021-2022.

California Department of Pesticide Regulation (CDPR). 2019. California pesticide information portal (CALPIP) application. https://calpip.cdpr.ca.gov/product.cfm

Callahan, B. J., P. J. McMurdie, and S. P. Holmes. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11:2639–2643.

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581–583.

Callon, M. 1984. Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. The sociological review 32:196–233.

Carrión, V. J., J. Perez-Jaramillo, V. Cordovez, V. Tracanna, M. De Hollander, D. Ruiz-Buck, L. W. Mendes, W. F. van Ijcken, R. Gomez-Exposito, and S. S. Elsayed. 2019. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science 366:606–612.

Cavicchioli, R., W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd, A. T. Classen, T. W. Crowther, R. Danovaro, C. M. Foreman, J. Huisman, D. A. Hutchins, J. K. Jansson, D. M. Karl, B. Koskella, D. B. Mark Welch, J. B. H. Martiny, M. A. Moran, V. J. Orphan, D. S. Reay, J. V. Remais, V. I. Rich, B. K. Singh, L. Y. Stein, F. J. Stewart, M. B. Sullivan, M. J. H. van Oppen, S. C. Weaver, E. A. Webb, and N. S. Webster. 2019. Scientists' warning to humanity: microorganisms and climate change. Nature Reviews Microbiology 17:569–586.

Chu, H., G. F. Gao, Y. Ma, K. Fan, and M. Delgado-Baquerizo. 2020. Soil microbial biogeography in a changing world: recent advances and future perspectives. mSystems 5:e00803-19, /msystems/5/2/msys.00803-19.atom.

Clark, J. S., L. C. Scher, and M. Swift. 2020. The emergent interactions that govern biodiversity change. Proceedings of the National Academy of Sciences of the United States of America 117:17074–17083.

Collins, H. M. 1974. The TEA set: tacit knowledge and scientific networks. Science Studies 4:165–185.

Cordovez, V., F. Dini-Andreote, V. J. Carrión, and J. M. Raaijmakers. 2019. Ecology and Evolution of Plant Microbiomes. Annual Review of Microbiology 73:69–88.

Crowther, T. W., D. S. Maynard, T. R. Crowther, J. Peccia, J. R. Smith, and M. A. Bradford. 2014. Untangling the fungal niche: The trait-based approach. Frontiers in Microbiology 5:1–12.

Crowther, T. W., J. van den Hoogen, J. Wan, M. A. Mayes, A. D. Keiser, L. Mo, C. Averill, and D. S. Maynard. 2019. The global soil community and its influence on biogeochemistry. Science 365:eaav0550.

da Veiga Leprevost, F., B. A. Grüning, S. Alves Aflitos, H. L. Röst, J. Uszkoreit, H. Barsnes, M. Vaudel, P. Moreno, L. Gatto, J. Weber, M. Bai, R. C. Jimenez, T. Sachsenberg, J. Pfeuffer, R. Vera Alvarez, J. Griss, A. I. Nesvizhskii, Y. Perez-Riverol. 2017. BioContainers: an open-source and community-driven framework for software standardization. Bioinformatics 33:2580-2582.

Dangi, S., R. Tirado-Corbalá, J. Gerik, and B. Hanson. 2017. Effect of long-term continuous fumigation on soil microbial communities. Agronomy 7:37.

Dawson, T. P., S. T. Jackson, J. I. House, I. C. Prentice, and G. M. Mace. 2011. Beyond predictions: Biodiversity conservation in a changing climate. Science 332:664.

Delgado-Baquerizo, M., C. A. Guerra, C. Cano-Díaz, E. Egidi, J.-T. Wang, N. Eisenhauer, B. K. Singh, and F. T. Maestre. 2020. The proportion of soil-borne pathogens increases with warming at the global scale. Nature Climate Change 10:550–554.

Delgado-Baquerizo, M., F. T. Maestre, P. B. Reich, T. C. Jeffries, J. J. Gaitan, D. Encinar, M. Berdugo, C. D. Campbell, and B. K. Singh. 2016. Microbial diversity drives multifunctionality in terrestrial ecosystems. Nature Communications 7:10541.

Dickey, J. R., R. A. Swenie, S. C. Turner, C. C. Winfrey, D. Yaffar, A. Padukone, K. K. Beals, K. S. Sheldon, and S. N. Kivlin. (2021). The utility of

macroecological rules for microbial biogeography. Frontiers in Ecology and Evolution 9.

Dickey, J. R., R. A. Swenie, S. C. Turner, C. C. Winfrey, D. Yaffar, A. Padukone, K. K. Beals, K. S. Sheldon, and S. N. Kivlin. 2021. The Utility of Macroecological Rules for Microbial Biogeography. Frontiers in Ecology and Evolution 9.

Docherty, K. M., H. M. Borton, N. Espinosa, M. Gebhardt, J. Gil-Loaiza, J. L. M. Gutknecht, P. W. Maes, B. M. Mott, J. J. Parnell, G. Purdy, P. A. P. Rodrigues, L. F. Stanish, O. N. Walser, and R. E. Gallery. 2015. Key edaphic properties largely explain temporal and geographic variation in soil microbial communities across four biomes. PLOS One 10:e0135352.

Dundore-Arias, J. P., M. Michalska-Smith, M. Millican, and L. L. Kinkel. 2023. More Than the Sum of Its Parts: Unlocking the Power of Network Structure for Understanding Organization and Function in Microbiomes. Annual Review of Phytopathology 61:annurev-phyto-021021-041457.

Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.

Elgersma, K. J., J. G. Ehrenfeld, S. Yu, and T. Vor. 2011. Legacy effects overwhelm the short-term effects of exotic plant invasion and restoration on soil microbial community structure, enzyme activities, and nitrogen cycling. Oecologia 167:733–745.

Escalas, A., L. Hale, J. W. Voordeckers, Y. Yang, M. K. Firestone, L. Alvarez-Cohen, and J. Zhou. 2019. Microbial functional diversity: from concepts to applications. Ecology and Evolution 9:12000–12016.

Evans, S. E., and M. D. Wallenstein. 2012. Soil microbial community response to drying and rewetting stress: does historical precipitation regime matter? Biogeochemistry 109:101–116.

Falkenberg, R., L. Sigl, and M. Fochler. 2023. From ‘making lists’ to conducting ‘well-rounded’ studies: Epistemic re-orientations in soil microbial ecology. Social Studies of Science:03063127231179700.

Feng, Y., J. Zhang, M. Berdugo, E. Guirado, C. A. Guerra, E. Egidi, J. Wang, B. K. Singh, and M. Delgado-Baquerizo. 2022. Temperature thresholds drive the global distribution of soil fungal decomposers. Global Change Biology 28:2779–2789.

Fennimore, S. A., M. J. Haar, and H. A. Ajwa. 2003. Weed control in strawberry provided by shank-and drip-applied methyl bromide alternative fumigants. HortScience 38:55–61.

Fernandez, C. W., and P. G. Kennedy. 2016. Revisiting the 'Gadgil effect': do interguild fungal interactions control carbon cycling in forest soils? New Phytologist 209:1382–1394.

Fernandez, C. W., N. H. Nguyen, A. Stefanski, Y. Han, S. E. Hobbie, R. A. Montgomery, P. B. Reich, and P. G. Kennedy. 2017. Ectomycorrhizal fungal response to warming is linked to poor host performance at the boreal-temperate ecotone. Global Change Biology 23:1598–1609.

Ferrenberg, S., S. P. O'Neill, J. E. Knelman, B. Todd, S. Duggan, D. Bradley, T. Robinson, S. K. Schmidt, A. R. Townsend, M. W. Williams, C. C. Cleveland, B. A. Melbourne, L. Jiang, and D. R. Nemergut. 2013. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. The ISME Journal 7:1102–1111.

Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. Diversity and Distributions 13:252–264.

Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International journal of climatology 37:4302–4315.

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.

Franklin, R. B., and A. L. Mills. 2003. Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. FEMS Microbiology Ecology 44:335–346.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software 33:1.

Fry, E. L., J. R. D. Long, L. Á. Garrido, N. Alvarez, Y. Carrillo, L. Castañeda-Gómez, M. Chomel, M. Dondini, J. E. Drake, S. Hasegawa, S. Hortal, B. G. Jackson, M. Jiang, J. M. Lavallee, B. E. Medlyn, J. Rhymes, B. K. Singh, P. Smith, I. C. Anderson, R. D. Bardgett, E. M. Baggs, and D. Johnson. 2019.

Using plant, microbe, and soil fauna traits to improve the predictive power of biogeochemical models. Methods in Ecology and Evolution 10:146–157.

Fujimura, J. H. 1988. The molecular biological bandwagon in cancer research: Where social worlds meet. Social Problems 35:261–283.

Giorgio, O. del, M. A. Crowley, L. X. Lu, and K. Schreiber. 2020. Building capacity through interdisciplinary graduate collaboration. Frontiers in Ecology and the Environment 18:479–479.

Guo, X., X. Zhou, L. Hale, M. Yuan, D. Ning, J. Feng, Z. Shi, Z. Li, B. Feng, Q. Gao, L. Wu, W. Shi, A. Zhou, Y. Fu, L. Wu, Z. He, J. D. Van Nostrand, G. Qiu, X. Liu, Y. Luo, J. M. Tiedje, Y. Yang, and J. Zhou. 2019. Climate warming accelerates temporal scaling of grassland soil microbial biodiversity. Nature Ecology & Evolution 3:612–619.

Guthman, J. 2019. Wilted: Pathogens, chemicals, and the fragile future of the strawberry industry. University of California Press, Berkeley, California, USA.

Hall, E. K., E. S. Bernhardt, R. L. Bier, M. A. Bradford, C. M. Boot, J. B. Cotner, P. A. del Giorgio, S. E. Evans, E. B. Graham, S. E. Jones, J. T. Lennon, K. J. Locey, D. Nemergut, B. B. Osborne, J. D. Rocca, J. P. Schimel, M. P. Waldrop, and M. D. Wallenstein. 2018. Understanding how microbiomes influence the systems they inhabit. Nature Microbiology 3:977–982.

Hawkes, C. V., and T. H. Keitt. 2015. Resilience vs. historical contingency in microbial responses to environmental change. Ecology Letters 18:612–625.

Hendershot, J. N., Q. D. Read, J. A. Henning, N. J. Sanders, and A. T. Classen. 2017. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. Ecology 98:1757–1763.

Henry, P. M., A. M. Pastrana, J. H. J. Leveau, and T. R. Gordon. 2019. Persistence of *Fusarium oxysporum* f. sp. *fragariae* in soil through asymptomatic colonization of rotation crops. Phytopathology 109:770–779.

Henry, P. M., M. Haugland, L. Lopez, M. Munji, D. C. Watson, and T. R. Gordon. 2020. The potential for *Fusarium oxysporum* f. sp. *fragariae*, cause of fusarium wilt of strawberry, to colonize organic matter in soil and persist through anaerobic soil disinfestation. Plant Pathology 69:1218–1226.

Holmes, G. J., S. M. Mansouripour, and S. S. Hewavitharana. 2020. Strawberries at the crossroads: management of soilborne diseases in California without methyl bromide. Phytopathology.

Hugerth, L. W., and A. F. Andersson. 2017. Analysing microbial community composition through amplicon sequencing: from sampling to hypothesis testing. Frontiers in Microbiology 8:1561.

Jones, J. A., P. M. Groffman, J. Blair, F. W. Davis, H. Dugan, E. E. Euskirchen, S. D. Frey, T. K. Harms, E. Hinckley, M. Kosmala, S. Loberg, S. Malone, K. Novick, S. Record, A. V. Rocha, B. L. Ruddell, E. H. Stanley, C. Sturtevant, A. Thorpe, T. White, W. R. Wieder, L. Zhai, and K. Zhu. 2020. Synergies among environmental science research and monitoring networks: a research agenda. Earth's Future 9:e2020EF001631.

Khan, A., W. Wicaksono, N. Ott, A. Poret-Peterson, and G. Browne. 2022. Random forest analysis reveals taxa predictive of Prunus replant disease in peach root microbiomes. PLOS ONE 17.

Kivlin, S. N., R. Muscarella, C. V. Hawkes, and K. K. Treseder. 2017. The Predictive Power of Ecological Niche Modeling for Global Arbuscular Mycorrhizal Fungal Biogeography. Pages 143–158 *in* L. Tedersoo, editor. Biogeography of Mycorrhizal Symbiosis. Springer International Publishing, Cham.

Krzywoszynska, A. 2019. Making knowledge and meaning in communities of practice: What role may science play? The case of sustainable soil management in England:160–168.

Kuhn, M. 2008. Building predictive models in R using the caret package. Journal of statistical software 28:1–26.

Kuhn, T. S. 1962. The Structure of Scientific Revolutions. University of Chicago Press, Chicago, Illinois, USA.

Ladau, J., T. J. Sharpton, M. M. Finucane, G. Jospin, S. W. Kembel, J. O'Dwyer, A. F. Koeppel, J. L. Green, and K. S. Pollard. 2013. Global marine bacterial diversity peaks at high latitudes in winter. The ISME Journal 7:1669–1677.

Latour, B. 1988. The Pasteurization of France. Harvard University Press, Cambridge, Massachusetts, USA.

Latour, B. 2004. Politics of Nature. Harvard University Press, Cambridge, Massachusetts, USA.

Lauber, C. L., K. S. Ramirez, Z. Aanderud, J. Lennon, and N. Fierer. 2013. Temporal variability in soil microbial communities across land-use types. The ISME journal 7:1641–1650.

Lauber, C. L., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied and Environmental Microbiology 75:5111–5120.

Law, J., and V. Singleton. 2014. ANT, multiplicity and policy. Critical Policy Studies 8:379–396.

Lee-Yaw, J. A., J. L. McCune, S. Pironon, and S. N. Sheth. 2022. Species distribution models rarely predict the biology of real populations. Ecography 2022.

Lennon, J. T., Z. T. Aanderud, B. K. Lehmkuhl, and D. R. Schoolmaster. 2012. Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology 93:1867–1879.

Levin, S. A. 1992. The problem of pattern and scale in Ecology: the Robert H. MacArthur Award Lecture. Ecology 73:1943–1967.

Lindahl, B. D., R. H. Nilsson, L. Tedersoo, K. Abarenkov, T. Carlsen, R. Kjøller, U. Kõljalg, T. Pennanen, S. Rosendahl, J. Stenlid, and H. Kauserud. 2013. Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. The New Phytologist 199:288–299.

Liu, L., Y. Yan, A. Ali, J. Zhao, Z. Cai, C. Dai, X. Huang, and K. Zhou. 2022. Deciphering the Fusarium-wilt control effect and succession driver of microbial communities managed under low-temperature conditions. Applied Soil Ecology 171.

Lozupone, C., and R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology 71:8228–8235.

Lunch, C., C. Laney, N. Mietkiewicz, E. Sokol, K. Cawley, and NEON [National Ecological Observatory Network]. 2021. neonUtilities: utilities for working with NEON data.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10-12.

Martiny, J. B. H., J. A. Eisen, K. Penn, S. D. Allison, and M. C. Horner-Devine. 2011. Drivers of bacterial diversity depend on spatial scale. Proceedings of the National Academy of Sciences 108:7850–7854.

Mather, C. 2014. Avian influenza multiple: Enacting realities and dealing with policies in South Africa's farmed ostrich sector. Journal of Rural Studies 33:99–106.

Maynard, D. S., K. R. Covey, T. W. Crowther, N. W. Sokol, E. W. Morrison, S. D. Frey, L. T. A. van Diepen, and M. A. Bradford. 2018. Species associations overwhelm abiotic conditions to dictate the structure and function of wood-decay fungal communities. Ecology 99:801–811.

Maynard, D. S., M. A. Bradford, K. R. Covey, D. Lindner, J. Glaeser, D. A. Talbert, P. J. Tinker, D. M. Walker, and T. W. Crowther. 2019. Consistent trade-offs in fungal trait expression across broad spatial scales. Nature Microbiology 4:846–853.

Maynard, D. S., T. W. Crowther, and M. A. Bradford. 2017. Fungal interactions reduce carbon use efficiency. Ecology Letters 20:1034–1042.

Mazzola, M., and L. M. Manici. 2012. Apple replant disease: role of microbial ecology in cause and control. Annual Review of Phytopathology 50:45–65.

Mazzola, M., and S. Freilich. 2017. Prospects for biological soilborne disease control: Application of indigenous versus synthetic microbiomes. Phytopathology 107:256–263.

Mazzola, M., S. S. Hewavitharana, and S. L. Strauss. 2015. Brassica seed meal soil amendments transform the rhizosphere microbiome and improve apple production through resistance to pathogen reinfestation. Phytopathology 105:460–469.

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.

McLaren, M. 2020. mikemc/speedyseq: speedyseq v0.2.0 (Version v0.2.0). Zenodo. http://doi.org/10.5281/zenodo.3923184.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8:e61217.

Mendes, R., M. Kruijt, I. De Bruijn, E. Dekkers, M. Van Der Voort, J. H. M. Schneider, Y. M. Piceno, T. Z. DeSantis, G. L. Andersen, P. A. H. M. Bakker, and J. M. Raaijmakers. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science 332:1097–1100.

Menzies, J. D. 1959. Occurrence and transfer of abiological factor in soil that suppresses potato scab. Phytopathology 49:648-652.

Mod, H. K., M. Chevalier, M. Luoto, and A. Guisan. 2020. Scale dependence of ecological assembly rules: Insights from empirical datasets and joint species distribution modelling. Journal of Ecology 108:1967–1977.

Mol, A. 1999. Ontological Politics. A Word and Some Questions. The Sociological Review 47:74–89.

Mol, A. 2002. The body multiple: ontology in medical practice. Duke University Press, Durham, North Carolina, USA.

Moser, I. 2008. Making Alzheimer's disease matter: Enacting, interfering and doing politics of nature. Geoforum 39:98–110.

Murali, A., A. Bhargava, and E. S. Wright. 2018. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome 6:1–14.

Muramoto, J., C. Shennan, M. Zavatta, G. Baird, L. Toyama, and M. Mazzola. 2016. Effect of anaerobic soil disinfestation and mustard seed meal for control of charcoal rot in California strawberries. International Journal of Fruit Science 16:59–70.

Murphy, S. J., and A. B. Smith. 2021. What can community ecologists learn from species distribution models? Ecosphere 12:1–19.

Nagy, R. C., J. K. Balch, E. K. Bissell, M. E. Cattau, N. F. Glenn, B. S. Halpern, N. Ilangakoon, B. Johnson, M. B. Joseph, S. Marconi, C. O'Riordan, J. Sanovia, T. L. Swetnam, W. R. Travis, L. A. Wasser, E. Woolner, P. Zarnetske, M. Abdulrahim, J. Adler, . . . K. Zhu. 2021. Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community. Ecosphere. In press.

National Ecological Observatory Network (NEON). 2021. Soil microbe marker gene sequences, RELEASE-2021 (DP1.10108.001). https://doi.org/10.48443/ybrs-zv89, https://data.neonscience.org

National Ecological Observatory Network (NEON). Soil microbe marker gene sequences, RELEASE-2021 (DP1.10108.001). https://doi.org/10.48443/ybrs-zv89. Dataset accessed from https://data.neonscience.org on January 22, 2021.

National Ecological Observatory Network (NEON). Soil physical and chemical properties, periodic, RELEASE-2021 (DP1.10086.001).

https://doi.org/10.48443/3qtw-w090. Dataset accessed from https://data.neonscience.org on January 22, 2021

Nemergut, D. R., S. K. Schmidt, T. Fukami, S. P. O'Neill, T. M. Bilinski, L. F. Stanish, J. E. Knelman, J. L. Darcy, R. C. Lynch, P. Wickey, and S. Ferrenberg. 2013. Patterns and processes of microbial community assembly. Microbiology and Molecular Biology Reviews 77:342–356.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32:268–274.

Nilsson, R. H., K.-H. Larsson, A. F. S. Taylor, J. Bengtsson-Palme, T. S. Jeppesen, D. Schigel, P. Kennedy, K. Picard, F. O. Glöckner, and L. Tedersoo. 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic acids research 47:D259–D264.

Nüst, D., D. Eddelbuettel, D. Bennett, R. Cannoodt, D. Clark, G. Daroczi, M. Edmondson, C. Fay, E. Hughes, L. Kjeldgaard, S. Lopp, B. Marwick, H. Nolis, J. Nolis, H. Ooi, K. Ram, N. Ross, L. Shepherd, P. Sólymos, T. L. Swetnam, N. Turaga, C. Van Petegem, J. Williams, C. Willis, and N. Xiao. 2020. The Rockerverse: packages and applications for containerization with R. The R Journal 12:437.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, E. Szoecs, and H. Wagner. 2020. vegan: community ecology package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan

Oliver, S. K., C. E. Fergus, N. K. Skaff, T. Wagner, P.-N. Tan, K. S. Cheruvelil, and P. A. Soranno. 2018. Strategies for effective collaborative manuscript development in interdisciplinary science teams. Ecosphere 9:e02206.

Oliverio, A. M., S. Geisen, M. Delgado-Baquerizo, F. T. Maestre, B. L. Turner, and N. Fierer. 2020. The global-scale distributions of soil protists and their contributions to belowground systems. Science Advances 6:eaax8787.

Olson, D. M., E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, and J. C. Morrison. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51:933–938.

Oreskes, N., and E. M. Conway. 2011. Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming. Bloomsbury Publishing USA.

Palmer, M. A., J. G. Kramer, J. Boyd, and D. Hawthorne. 2016. Practices for facilitating interdisciplinary synthetic research: the National Socio-Environmental Synthesis Center (SESYNC). Current Opinion in Environmental Sustainability 19:111–122.

Pauvert, C., M. Buée, V. Laval, V. Edel-Hermann, L. Fauchery, A. Gautier, I. Lesur, J. Vallance, and C. Vacher. 2019. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. Fungal Ecology 41:23–33.

Paxson, H., and S. Helmreich. 2014. The perils and promises of microbial abundance: Novel natures and model ecosystems, from artisanal cheese to alien seas. Social Studies of Science 44:165–193.

Peay, K. G. 2016. The Mutualistic Niche: Mycorrhizal Symbiosis and Community Dynamics. Annual Review of Ecology, Evolution, and Systematics 47:143–164.

Peay, K. G., P. G. Kennedy, and J. M. Talbot. 2016. Dimensions of biodiversity in the Earth mycobiome. Nature Reviews Microbiology 14:434–447.

Peppler-Lisbach, C., and B. Schröder. 2004. Predicting the species composition of Nardus stricta communities by logistic regression modelling. Journal of Vegetation Science 15:623–634.

Phillips, S. 2021. Maxnet: Fitting 'Maxent' species distribution models with 'glmnet'. R Package version, 0(1), 4. https://CRAN.R-project.org/package=maxnet

Pietikäinen, J., M. Pettersson, and E. Bååth. 2005. Comparison of temperature effects on soil respiration and bacterial and fungal growth rates. FEMS Microbiology Ecology 52:49–58.

Poggio, L., L. M. De Sousa, N. H. Batjes, G. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7:217–240.

Põlme, S., K. Abarenkov, R. Henrik Nilsson, B. D. Lindahl, K. E. Clemmensen, H. Kauserud, N. Nguyen, R. Kjøller, S. T. Bates, and P. Baldrian. 2020. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. Fungal diversity 105:1–16.

Prober, S. M., J. W. Leff, S. T. Bates, E. T. Borer, J. Firn, W. S. Harpole, E. M. Lind, E. W. Seabloom, P. B. Adler, J. D. Bakker, E. E. Cleland, N. M. DeCrappeo, E. DeLorenze, N. Hagenah, Y. Hautier, K. S. Hofmockel, K. P. Kirkman, J. M. H. Knops, K. J. L. Pierre, A. S. MacDougall, R. L. McCulley, C. E. Mitchell, A. C. Risch, M. Schuetz, C. J. Stevens, R. J. Williams, and N. Fierer. 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecology Letters 18:85–95.

Prodan, A., V. Tremaroli, H. Brolin, A. H. Zwinderman, M. Nieuwdorp, and E. Levin. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLOS One 15:e0227434.

Puig de la Bellacasa, M. 2015. Making time for soil: Technoscientific futurity and the pace of care. Social Studies of Science 45:691–716.

Qin, C., K. Zhu, N. R. Chiariello, C. B. Field, and K. G. Peay. 2020. Fire history and plant community composition outweigh decadal multi-factor global change as drivers of microbial composition in an annual grassland. Journal of Ecology 108:611–625.

Qin, C., R. Bartelme, Y. A. Chung, D. Fairbanks, Y. Lin, D. Liptzin, C. Muscarella, K. Naithani, K. Peay, P. Pellitier, A. St. Rose, L. Stanish, Z. Werbin, and K. Zhu. 2021. From DNA sequences to microbial ecology: Wrangling NEON soil microbe data with the neonMicrobe R package. Ecosphere 12.

Querejeta, J., L. M. Egerton-Warburton, and M. F. Allen. 2003. Direct nocturnal water transfer from oaks to their mycorrhizal symbionts during severe soil drying. Oecologia 134:55–64.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www. R-project. org.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Razavi, B. S., E. Blagodatskaya, and Y. Kuzyakov. 2015. Nonlinear temperature sensitivity of enzyme kinetics explains canceling effect—a case study on loamy haplic Luvisol. Frontiers in microbiology 6:1126.

Reich, P. B., and J. Oleksyn. 2008. Climate warming will reduce growth and survival of Scots pine except in the far north. Ecology Letters 11:588–597.

Rhoten, D. 2003. A muli-method analysis of the social and technical conditions for interdisciplinary collaboration. Final Report, National Science Foundation BCS-0129573.

Roberts, E. F. S. 2017. What gets inside: Violent entanglements and toxic boundaries in Mexico City. Cultural Anthropology 32:592–619.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584.

Rosskopf, E. N., N. Kokalis-Burelle, S. A. Fennimore, and C. A. Wilen. 2016. Soil/media disinfestation for management of florists' crops diseases. *In* R. J. McGovern and W. H. Elmer, editors. Handbook of Florists' Crops Disease, Handbook of Plant Disease Management. Springer International Publishing, Cham.

Sato, H., R. Tsujino, K. Kurita, K. Yokoyama, and K. Agata. 2012. Modelling the global distribution of fungal species: new insights into microbial cosmopolitanism. Molecular Ecology 21:5599–5612.

Schadt, C. W., A. P. Martin, D. A. Lipson, and S. K. Schmidt. 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. Science 301:1359–1361.

Schimel, D., and M. Keller. 2015. Big questions, big science: meeting the challenges of global ecology. Oecologia 177:925-934.

Schimel, J. P. 2018. Life in dry soils: effects of drought on soil microbial communities and processes. Annual review of ecology, evolution, and systematics 49:409–432.

Schoch, C. L., K. A. Seifert, S. Huhndorf, V. Robert, J. L. Spouge, C. A. Levesque, W. Chen, F. B. Consortium, F. B. C. A. List, and E. Bolchacova. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the national academy of Sciences 109:6241–6246.

Schrader, A. 2010. Responding to *Pfiesteria piscicida* (the fish killer): Phantomatic ontologies, indeterminacy, and responsibility in toxic microbiology. Social Studies of Science 40:275–306.

Schrodt, F., M. J. Santos, J. J. Bailey, and R. Field. 2019. Challenges and opportunities for biogeography—What can we still learn from von Humboldt? Journal of Biogeography 46:1631–1642.

Serna-Chavez, H. M., N. Fierer, and P. M. van Bodegom. 2013. Global drivers and patterns of microbial abundance in soil. Global Ecology and Biogeography 22:1162–1172.

Serreze, M. C., and R. G. Barry. 2011. Processes and impacts of Arctic amplification: A research synthesis. Global and planetary change 77:85–96.

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

Shennan, C., J. Muramoto, J. Lamers, M. Mazzola, E. N. Rosskopf, N. Kokalis-Burelle, N. Momma, D. M. Butler, and Y. Kobara. 2014. Anaerobic soil disinfestation for soil borne disease control in strawberry and vegetable systems: current knowledge and future directions. Acta Horticulturae:165–175.

Shi, Z., J. Zhu, J. Wu, A. Cao, W. Fang, D. Yan, Q. Wang, and Y. Li. 2022. Effects of chloropicrin, dimethyl disulfide and metham sodium applied simultaneously on soil-born bacteria and fungi. Agriculture 12:1982.

Shrestha, U., R. Auge, and D. Butler. 2016. A Meta-Analysis of the Impact of Anaerobic Soil Disinfestation on Pest Suppression and Yield of Horticultural Crops. Frontiers in Plant Science 7.

Šinkovec, H., G. Heinze, R. Blagus, and A. Geroldinger. 2021. To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. BMC Medical Research Methodology 21:199.

Sismondo, S. 2015. Ontological turns, turnoffs and roundabouts. Social Studies of Science 45:441–448.

Smith, A. B. 2021. enmSdm: Tools for modeling niches and distributions of species. R package version 0.5.3.7.

Smith, D. P., and K. G. Peay. 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS ONE 9.

Smith, G. R., and K. G. Peay. 2021. Multiple distinct, scale-dependent links between fungi and decomposition. Ecology letters 24:1352–1362.

Stanish, L., and S. Parker. 2019. NEON user guide to microbe marker gene sequences (NEON.DP1.10108; NEON.DP1.20280; NEON.DP1.20282). National Ecological Observatory Network.

https://data.neonscience.org/documents/10179/2838366/NEON_markerGenes_userGuide_vB/ab19bd56-07d2-1d53-9e79-3093e6340ca0

Stanish, L., J. J. Parnell, and K. Blevins. 2018. TOS science design for terrestrial microbial diversity. National Ecological Observatory Network. https://data.neonscience.org/documents/10179/1723439/NEON.DOC.000908vB/15287f2d-59bc-4475-bab8-ed05abc4d06b

Ștefan, V., & Levin, S. 2018. plotbiomes: R package for plotting Whittaker biomes with ggplot2 (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.7145245

Steidinger, B. S., J. M. Bhatnagar, R. Vilgalys, J. W. Taylor, C. Qin, K. Zhu, T. D. Bruns, and K. G. Peay. 2020. Ectomycorrhizal fungal diversity predicted to substantially decline due to climate changes in North American Pinaceae forests. Journal of Biogeography 47:772–782.

Steidinger, B. S., T. W. Crowther, J. Liang, M. E. Van Nuland, G. D. A. Werner, P. B. Reich, G. J. Nabuurs, S. de-Miguel, M. Zhou, N. Picard, B. Herault, X. Zhao, C. Zhang, D. Routh, and K. G. Peay. 2019. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. Nature 569:404–408.

Sulman, B. N., E. Shevliakova, E. R. Brzostek, S. N. Kivlin, S. Malyshev, D. N. L. Menge, and X. Zhang. 2019. Diverse mycorrhizal associations enhance terrestrial C storage in a global model. Global Biogeochemical Cycles 33:501–523.

Takahashi, S., J. Tomita, K. Nishioka, T. Hisada, and M. Nishijima. 2014. Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. PLOS One 9:e105592.

Talbot, J. M., T. D. Bruns, J. W. Taylor, D. P. Smith, S. Branco, S. I. Glassman, S. Erlandson, R. Vilgalys, H.-L. Liao, M. E. Smith, and K. G. Peay. 2014. Endemism and functional convergence across the North American soil mycobiome. Proceedings of the National Academy of Sciences 111:6341–6346.

Tedersoo, L., K. S. Ramirez, R. H. Nilsson, A. Kaljuvee, U. Kõljalg, and K. Abarenkov. 2015. Standardizing metadata and taxonomic identification in metabarcoding studies. GigaScience 4.

Tedersoo, L., M. Bahram, S. Põlme, U. Kõljalg, N. S. Yorou, R. Wijesundera, L. V. Ruiz, A. M. Vasco-Palacios, P. Q. Thu, A. Suija, M. E. Smith, C. Sharp, E. Saluveer, A. Saitta, M. Rosas, T. Riit, D. Ratkowsky, K. Pritsch, K. Põldmaa,

M. Piepenbring, C. Phosri, M. Peterson, K. Parts, K. Pärtel, E. Otsing, E. Nouhra, A. L. Njouonkou, R. H. Nilsson, L. N. Morgado, J. Mayor, T. W. May, L. Majuakim, D. J. Lodge, S. S. Lee, K.-H. Larsson, P. Kohout, K. Hosaka, I. Hiiesalu, T. W. Henkel, H. Harend, L. Guo, A. Greslebin, G. Grelet, J. Geml, G. Gates, W. Dunstan, C. Dunk, R. Drenkhan, J. Dearnaley, A. D. Kesel, T. Dang, X. Chen, F. Buegger, F. Q. Brearley, G. Bonito, S. Anslan, S. Abell, and K. Abarenkov. 2014. Global diversity and geography of soil fungi. Science 346:1052–1053.

Tedersoo, L., S. Anslan, M. Bahram, S. Põlme, T. Riit, I. Liiv, U. Kõljalg, V. Kisand, H. Nilsson, F. Hildebrand, P. Bork, and K. Abarenkov. 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys 10:1–43.

Tedersoo, L., T. W. May, and M. E. Smith. 2010. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. Mycorrhiza 20:217–263.

Tedersoo, L., V. Mikryukov, A. Zizka, M. Bahram, N. Hagh-Doust, S. Anslan, O. Prylutskyi, M. Delgado-Baquerizo, F. T. Maestre, J. Pärn, M. Öpik, M. Moora, M. Zobel, M. Espenberg, Ü. Mander, A. N. Khalid, A. Corrales, A. Agan, A. Vasco-Palacios, A. Saitta, A. C. Rinaldi, A. Verbeken, B. P. Sulistyo, B. Tamgnoue, B. Furneaux, C. D. Ritter, C. Nyamukondiwa, C. Sharp, C. Marín, D. Gohar, D. Klavina, D. Sharmah, D. Q. Dai, E. Nouhra, E. M. Biersma, E. Rähn, E. K. Cameron, E. De Crop, E. Otsing, E. A. Davydov, F. E. Albornoz, F. Q. Brearley, F. Buegger, G. Zahn, G. Bonito, I. Hiiesalu, I. C. Barrio, J. Heilmann-Clausen, J. Ankuda, J. Y. Kupagme, J. G. Maciá-Vicente, J. D. Fovo, J. Geml, J. M. Alatalo, J. Alvarez-Manjarrez, K. Põldmaa, K. Runnel, K. Adamson, K. A. Bråthen, K. Pritsch, K. I. Tchan, K. Armolaitis, K. D. Hyde, K. K. Newsham, K. Panksep, A. A. Lateef, L. Tiirmann, L. Hansson, L. J. Lamit, M. Saba, M. Tuomi, M. Gryzenhout, M. Bauters, M. Piepenbring, N. Wijayawardene, N. S. Yorou, O. Kurina, P. E. Mortimer, P. Meidl, P. Kohout, R. H. Nilsson, R. Puusepp, R. Drenkhan, R. Garibay-Orijel, R. Godoy, S. Alkahtani, S. Rahimlou, S. V. Dudov, S. Põlme, S. Ghosh, S. Mundra, T. Ahmed, T. Netherway, T. W. Henkel, T. Roslin, V. Nteziryayo, V. E. Fedosov, V. G. Onipchenko, W. A. E. Yasanthika, Y. W. Lim, N. A. Soudzilovskaia, A. Antonelli, U. Kõljalg, and K. Abarenkov. 2022. Global patterns in endemicity and vulnerability of soil fungi. Global Change Biology 28:6696–6710.

Terrer, C., S. Vicca, B. A. Hungate, R. P. Phillips, and I. C. Prentice. 2016. Mycorrhizal association as a primary control of the CO2 fertilization effect. Science 353:72–74.

Thompson, L. R., J. G. Sanders, D. McDonald, A. Amir, J. Ladau, K. J. Locey, R. J. Prill, A. Tripathi, S. M. Gibbons, G. Ackermann, J. A. Navas-Molina, S. Janssen, E. Kopylova, Y. Vázquez-Baeza, A. González, J. T. Morton, S. Mirarab, Z. Zech Xu, L. Jiang, M. F. Haroon, J. Kanbar, Q. Zhu, S. Jin Song, T. Kosciolek, N. A. Bokulich, J. Lefler, C. J. Brislawn, G. Humphrey, S. M. Owens, J. Hampton-Marcell, D. Berg-Lyons, V. McKenzie, N. Fierer, J. A. Fuhrman, A. Clauset, R. L. Stevens, A. Shade, K. S. Pollard, K. D. Goodwin, J. K. Jansson, J. A. Gilbert, and R. Knight. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463.

Trivedi, P., J. E. Leach, S. G. Tringe, T. Sa, and B. K. Singh. 2020. Plant–microbiome interactions: from community assembly to plant health. Nature Reviews Microbiology 18:607–621.

Trivedi, P., M. Delgado-Baquerizo, C. Trivedi, K. Hamonts, I. C. Anderson, and B. K. Singh. 2017. Keystone microbial taxa regulate the invasion of a fungal pathogen in agro-ecosystems. Soil Biology and Biochemistry 111:10–14.

U.S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA). 2023. Current Research Information System. https://cris.nifa.usda.gov/cgi-bin/starfinder/0?path=fastlink1.txt&id=anon&pass=&search=(GC=112.C)&format=WEBTITLESGIY

Uekötter, F. 2014. Ignorance is strength: Science-based agriculture and the merits of incomplete knowledge. Pages 122–139 *in* Uekötter, F., and U. Lübken, editors. Managing the unknown: Essays on environmental ignorance. Berghahn Books. New York, New York, USA.

van Agtmaal, M., G. van Os, W. Hol, M. Hundscheid, W. Runia, C. Hordijk, and W. de Boer. 2015. Legacy effects of anaerobic soil disinfestation on soil bacterial community composition and production of pathogen-suppressing volatiles. Frontiers in Microbiology 6.

van der Linde, S., L. M. Suz, C. D. L. Orme, F. Cox, H. Andreae, E. Asi, B. Atkinson, S. Benham, C. Carroll, N. Cools, B. De Vos, H.-P. Dietrich, J. Eichhorn, J. Gehrmann, T. Grebenc, H. S. Gweon, K. Hansen, F. Jacob, F. Kristöfel, P. Lech, M. Manninger, J. Martin, H. Meesenburg, P. Merilä, M. Nicolas, P. Pavlenda, P. Rautio, M. Schaub, H.-W. Schröck, W. Seidling, V. Šrámek, A. Thimonier, I. M. Thomsen, H. Titeux, E. Vanguelova, A. Verstraeten, L. Vesterdal, P. Waldner, S. Wijk, Y. Zhang, D. Žlindra, and M. I. Bidartondo. 2018. Environment and host as large-scale controls of ectomycorrhizal fungi. Nature 558:243–248.

Van Nuland, M. E., D. P. Smith, J. M. Bhatnagar, A. Stefanski, S. E. Hobbie, P. B. Reich, and K. G. Peay. 2020. Warming and disturbance alter soil microbiome diversity and function in a northern forest ecotone. FEMS Microbiology Ecology 96:fiaa108.

Větrovský, T., P. Kohout, M. Kopecký, A. Machac, M. Man, B. D. Bahnmann, V. Brabcová, J. Choi, L. Meszárošová, Z. R. Human, C. Lepinay, S. Lladó, R. López-Mondéjar, T. Martinović, T. Mašínová, D. Morais, D. Navrátilová, I. Odriozola, M. Štursová, K. Švec, V. Tláskal, M. Urbanová, J. Wan, L. Žifčáková, A. Howe, J. Ladau, K. G. Peay, D. Storch, J. Wild, and P. Baldrian. 2019. A meta-analysis of global fungal distribution reveals climate-driven patterns. Nature Communications 10:1–9.

Vincent, I., B. Paudel, H. Guo, E. Rosskopf, F. Di Gioia, J. Hong, D. McNear, N. Xu, L. Anrecio, J. Colee, and X. Zhao. 2022. Spatial and Temporal Changes of Soil Microbial Communities in Field Tomato Production as Affected by Anaerobic Soil Disinfestation. Frontiers in Sustainable Food Systems 6.

Wang, C., D. Liu, and E. Bai. 2018. Decreasing soil microbial diversity is associated with decreasing microbial biomass under nitrogen addition. Soil Biology and Biochemistry 120:126–133.

Wang, S., T. Lamy, L. M. Hallett, and M. Loreau. 2019. Stability and synchrony across ecological hierarchies in heterogeneous metacommunities: linking theory to data. Ecography 42:1200–1211.

Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. Ecological Applications 21:335–342.

Weller, D. M., J. M. Raaijmakers, B. B. M. Gardener, and L. S. Thomashow. 2002. Microbial populations responsible for specific soil suppressiveness to plant pathogens. Annual Review of Phytopathology 40:309–348.

Werbin Z. R., B. Hackos, M.C. Dietze, and J.M. Bhatnagar. 2021. The National Ecological Observatory Network's soil metagenomes: assembly and basic analysis [version 1; peer review: 2 approved with reservations]. F1000 Research 10:299.

Wieder, W. R., C. C. Cleveland, W. K. Smith, and K. Todd-Brown. 2015. Future productivity and carbon storage limited by terrestrial nutrient availability. Nature Geoscience 8:441–444.

Wilhelm, S. 1961. Verticillium wilt of strawberry controlled by fumigation of soil with chloropicrin and chloropicrin-methyl bromide mixtures. Phytopathology 51:744–748.

Wilhelm, S., and A. O. Paulus. 1980. How soil fumigation benefits the California strawberry industry. Plant disease 64:264–270.

Williams, J. W., and S. T. Jackson. 2007. Novel climates, no-analog communities, and ecological surprises. Frontiers in Ecology and the Environment 5:475–482.

Williams, S. E., L. P. Shoo, J. L. Isaac, A. A. Hoffmann, and G. Langham. 2008. Towards an Integrated Framework for Assessing the Vulnerability of Species to Climate Change. PLoS Biology 6:e325.

Woolgar, S., and J. Lezaun. 2013. The wrong bin bag: A turn to ontology in science and technology studies? Social Studies of Science 43:321–340.

Wright, E., S. 2016. Using DECIPHER v2.0 to analyze big biological sequence data in R. The R Journal 8:352.

Xu, X., N. Wang, D. Lipson, R. Sinsabaugh, J. Schimel, L. He, N. A. Soudzilovskaia, and L. Tedersoo. 2020. Microbial macroecology: in search of mechanisms governing microbial biogeographic patterns. Global Ecology and Biogeography 29:1870–1886.

Yilmaz, P., J. A. Gilbert, R. Knight, L. Amaral-Zettler, I. Karsch-Mizrachi, G. Cochrane, Y. Nakamura, S.-A. Sansone, F. O. Glöckner, and D. Field. 2011. The genomic standards consortium: bringing standards to life for microbial ecology. The ISME Journal 5:1565–1567.

Zhou, J., and D. Ning. 2017. Stochastic community assembly: Does it matter in microbial ecology? Microbiology and Molecular Biology Reviews 81:1–32.

Zinger, L., P. Taberlet, H. Schimann, A. Bonin, F. Boyer, M. D. Barba, P. Gaucher, L. Gielly, C. Giguet-Covex, A. Iribar, M. Réjou-Méchain, G. Rayé, D. Rioux, V. Schilling, B. Tymen, J. Viers, C. Zouiten, W. Thuiller, E. Coissac, and J. Chave. 2019. Body size determines soil community assembly in a tropical forest. Molecular Ecology 28:528–543.