UNIVERSITY OF CALIFORNIA,
IRVINE

Microbial Functional and Genetic Variation with Climate Change

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Bahareh Sorouri

Dissertation Committee:
Professor Steven D. Allison, Chair
Professor Brandon S. Gaut
Professor Jennifer B. H. Martiny
Professor Adam C. Martiny

2023

DEDICATION

To

my family

in recognition of their unconditional love and support.

To

my parents

Simin Daryaee and Mohammad Javad Sorouri

Thank you for your sacrifices, dedication, inspiration, and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

To my advisor, Steven:

Thank you for your time, mentorship, kindness, and support throughout this challenging process. You are one of the few people who I have seen embody and advocate for diversity, equity, and inclusion in all aspects of your life. I have learned so much from you on how to be a better researcher and advocate. Thank you for giving me the opportunity to join your lab, it has been a privilege to work with you these past few years.

To my committee, Brandon, Jen, and Adam:

Thank you all for pushing me to be a better scientist and critical thinker. I appreciated your thought-provoking questions and advice during our meetings. Thank you for sharing your resources and for your guidance throughout my dissertation research. In particular, Brandon, thank you for time, patience, support, and mentorship. I learned so much from you during our weekly meetings and am very grateful to have worked with you during this journey.

Many thanks to my collaborators, Cynthia I. Rodriguez and Nicholas C. Scales. Also, thank you to Tiffany N. Batarseh. It has been an honor to work with and learn from each of you.

Thank you to my family, friends, community, mentors, and colleagues. I truly appreciate your support and encouragement throughout the years. Special thanks to Maman, Baba, Tahereh, Hedieh, Moein, Aga Houman, Zaynab, Fatemeh, and Zahra. Your patience, support, and guidance has shaped me into the person that I am today. Each of you inspire me beyond words, and I am unbelievably blessed to be a part of your family. This dissertation would not have been possible without you.

This text of Chapter 1 is a reprint of the material as it appears in Elementa: Science of the Anthropocene. This is an open-access article “distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any media, provided that the original author and source are credited. See http://creativecommons.org/licenses/by/1.0/”. Steven D. Allison, a c-author listed in this publication, directed and supervised the research, which forms the basis for this dissertation.

VITA

Bahareh Sorouri

EDUCATION

2023	Ph.D., Biological Sciences, University of California, Irvine (UCI)
2021	M.S., Biological Sciences, University of California, Irvine (UCI)
2017	B.S., Genetics, University of California, Irvine (UCI)
2015	A.A., Biology, Chemistry, and Physical Sciences, Irvine Valley College (IVC)

PUBLICATIONS

Sorouri, B., Rodriguez, CI., Gaut, BS., Allison, SD. (2023) Variation in *Sphingomonas* Traits across Habitats and Phylogenetic Clades. *Frontiers in Microbiology* (In Revision).

Sorouri, B., Allison, SD. (2022) Microbial Extracellular Enzyme Activity with Simulated Climate Change. Elementa: Science of the Anthropocene 10:1.

PUBLICATION IN PREPARATION

Sorouri, B., Scales, NC., Gaut, BS., Allison, SD. *Sphingomonas* Clade and Functional Distribution with Simulated Climate Change.

NON-ACADEMIC SCIENCE WRITING

2020-2023	Edited approximately 300 science podcast scripts covering recent scientific papers across all disciplines. Loh Down on Science.
2020	Wrote and recorded three podcast episodes on the science and history behind pandemics. **Bahareh Sorouri**. Loh Down on Science: Special Pandemic Edition.
2019-2020	Wrote ten radio stories, each covering a recent scientific paper across all disciplines. **Bahareh Sorouri**. Loh Down on Science.

SELECT FELLOWSHIPS AND AWARDS

2022-2023	American Geophysical Union Voices for Science Advocate
2019-2023	Loh Down on Science Fellowship

2022	Graduate Dean's Dissertation Year Fellowship
2022	UCI Associated Graduate Student Symposium Biological/Life Sciences 2nd Place Oral Presentation Winner
2019	NSF Graduate Research Fellowship Honorable Mention
2019	UCI Environmental Research Symposium Best Poster

SELECT TEACHING EXPERIENCES

2023	UCI BIO 94 Organisms to Ecosystems Teaching Assistant
2021	UCI/CSU Long Beach Pre-Professor Program Fellow
2021	UCI Preparing for Future Faculty Careers Certificate Program
2021	UCI E 118 Ecosystem Ecology Teaching Assistant
2021	UCI BIO 94 Organisms to Ecosystems Teaching Assistant
2020	UCI BIO 93 DNA to Organisms Teaching Assistant

SELECT MENTORING AND OUTREACH EXPERIENCES

2020-2022	ReachOut TeachOut at UCI Co-President
2018-2021	Allison Lab Mentor for undergraduate students; 3 mentees won UCI UROP Grants and Excellence in Research Awards
2017-2020	Letters to a Pre-Scientist Mentor
2018-2019	ReachOut TeachOut at UCI Secretary
2018-2019	Irvine Unified School District Science Fair Judge and Volunteer
2018-2019	University High School Invited Speaker
2018	UCI Freshman Seminar Invited Panelist
2017-2018	ReachOut TeachOut at UCI Volunteer

PRESENTATIONS

Invited Talks

Sorouri B. (2022) *Voices for Science.* American Geophysical Union. Chicago, IL.

Sorouri B. (2022) *Uncovering Traits within the Microbial Tree of Life*. Soka University of America. Aliso Viejo, CA.

Sorouri B., Allison SD. (2020) *Microbial Extracellular Enzyme Activity Along a Climate Gradient*. Society for Advancement of Chicanos/Hispanics and Native Americans in Science Annual Conference. Virtual. (***Session Speaker***)

Berhe AA., **Sorouri B.,** Ulrich R., Handley H., Emani S., Marcus T. (2019). *The Power of Science Lies in Its Diverse Voices*. American Geophysical Union Town Hall. San Francisco, CA. (***Invited Panelist***). Featured article

Conference Presentations

Sorouri B., Scales, NC, Gaut BS, Allison SD. (2022) *Sphingomonas Clade Distribution Across a Southern California Climate Gradient.* American Geophysical Union. Chicago, IL. **(*Poster*)**

Sorouri B., Rodriguez CI., Gaut BS., Allison SD. (2021) *Variation in Bacterial Traits Across Habitats and Phylogenetic Clades*. Ecological Society of America & Canadian Society for Ecology and Evolution. Montréal, Québec, Canada. (***Oral)***

Sorouri B. *Blast to the Past: Variation in Bacterial Traits across Habitats and Phylogenetic Clades.* (2022) UCI Associated Graduate Student Symposium. Irvine, CA. ***(Oral)***

Finley BK., **Sorouri B.,** Treseder KK., Martiny JBH., Martiny AC., Rodriguez-Verdugo A., Goulden ML., Wang S., Brodie EL., Allison SD. (2022) *Litter Microbial Trait-Based Strategies in Response to Drought: A Multi-omics Approach*. Department of Energy Genomics Science Program Annual Principal Investigator Meeting. Virtual.

Sorouri B. (2022) *Sphingomonas: Solution to Pollution?*. UCI Grad Slam Semifinals. Virtual. ***(Oral)***

Sorouri B., Rodriguez CI., Allison SD. (2021) *Variation Bacterial Traits Across Phylogenetic Clades and Environments*. American Geophysical Union. New Orleans, LA. (**Poster**)

Sorouri B. (2020) *Who Run the World: Environmental Microbes*. UCI Grad Slam Semifinals. Irvine, CA. ***(Oral)***

Sorouri B., Allison SD. (2020) *Microbial Extracellular Enzyme Activity Along a Climate Gradient*. Society for Advancement of Chicanos/Hispanics and Native Americans in Science Annual Conference. Virtual. ***(Oral)***

Sorouri B., Allison SD. (2019) *Microbial Communities Exhibit Resilient Extracellular Enzyme Activity Along a Climate Gradient*. UCI Environmental Research Poster Symposium. Irvine, CA. **(*Awarded Best Poster***)

Sorouri B., Allison SD. (2019) *Microbial Communities Exhibit Resilient Extracellular Enzyme Activity Along a Climate Gradient*. American Geophysical Union Meeting. San Francisco, CA. **(*Poster*)**

Sorouri B., Allison SD. (2019) *Environmental Microbes: (Friendly) Snakes in the Grass*. Associated Graduate Student Symposium. Irvine, CA. **(Oral)**

ABSTRACT OF THE DISSERTATION

Microbial Functional and Genetic Variation with Climate Change

by

Bahareh Sorouri

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2023

Professor Steven D. Allison, Chair

Microbial communities are integral for the survival of organisms and ecosystems. Anthropogenic influences like pollution and climate change drastically impact their environment, and microbial responses to these phenomena are uncertain. My dissertation investigates microbial functional and genetic variation with climate change. I aimed to address the following objectives, 1) determine how microbial community extracellular enzyme activity varied during a reciprocal transplant that simulated climate change; 2) assess the evolutionary history, phylogeny, and habitat preference in a comparative genomic analysis of publicly available *Sphingomonas* genomes found worldwide; 3) uncover how *Sphingomonas* clade and functional distribution vary under simulated climate change. To address these objectives, I used computational, statistical, and bioinformatic techniques to analyze proteomic, genomic, and metagenomic data.

In my first chapter, I investigated the variation in extracellular enzyme activity and litter decomposition of microbial communities from a Southern California climate gradient after an 18-month reciprocal transplant. Communities were from five sites that varied inversely with temperature and precipitation (desert, grassland, mountains, etc.), and the reciprocal transplant simulated future climate change conditions. During the reciprocal transplantation, microbial

communities from each site were inoculated onto sterile grassland leaf litter, placed in bags that allowed for the transfer of nutrients, and distributed back into each site. Enzyme activity suggested microbial communities were not specialized to their native environment. Additionally, there was rarely a reduction in enzyme function after microbial communities were transplanted into new climate conditions. I found significant differences in decomposition rates; however, they were not related to enzyme activities. These results suggest that direct, physiological impacts of climate are potentially important for enzyme-mediated decomposition, but climate specialization will not constrain the microbial response to climate change in our system.

In my second chapter, I used the bacterium *Sphingomonas* to explore why certain bacteria are present in specific habitats, by analyzing how microbial traits vary with evolutionary history. The *Sphingomonas* genus inhabits a wide variety of environments and hosts, making it ideal for examining the distribution of habitat preference traits. Furthermore, with appropriate management and manipulation, *Sphingomonas* can rehabilitate polluted locations. In this project, I downloaded publicly available *Sphingomonas* genomes, quality filtered them, curated them into eight habitat categories based on their isolation source (plants, animals, contaminated sites, etc.), analyzed their gene content, and assessed their evolutionary history. I found that closely related *Sphingomonas* genomes shared similar accessory genes, and genomes from similar habitats clustered together in phylogenetic clades. Moreover, the frequencies of functional genes significantly varied by habitat, suggesting habitat preference. Understanding environmental and host influence on *Sphingomonas* evolutionary history at a genomic level will aid future functional predictions and restoration of polluted habitats.

In my third chapter, I expand on my previous findings to inspect the clade and functional distribution of *Sphingomonas* along the Southern California climate gradient, before and after the

reciprocal transplantation. Using metagenomic data, I trimmed and quality filtered sequences, extracted *Sphingomonas* core genes, determined *Sphingomonas* clade composition, and inferred the distribution of gene-based functional traits. I confirmed that prior to transplantation, sites have distinct *Sphingomonas* clade compositions. The clade and functional composition shift after the 18-month transplant, and site conditions had the most significant effect on both clade and functional composition. In combination with previous research from the Southern California climate gradient, these findings support consistent bacterial response to climate change at multiple phylogenetic levels. In summary, this work will help assess and predict microbial response to climate change.

INTRODUCTION

Anthropogenic activities and pollution are changing the climate and causing major environmental perturbations for the world and drastic consequences for its inhabitants (Barnosky, 2011; Johnson, 2017; Ripple, 2017; IPCC, 2022). More research is needed to understand the consequences for microbial inhabitants that are important for ecosystems and for the survival of organisms (Cavicchioli et al., 2019; Tiedje et al., 2022). Specifically, microbial communities are crucial for driving global nutrient cycling – such as carbon, nitrogen, and phosphorus – that maintain ecosystem function and support organisms (Falkowski et al., 2008). It is instrumental to understand how microbes will respond in the face of their changing habitats, because they have the potential to mitigate the effects of climate change (Cavicchioli et al., 2019).

To predict microbial responses to climate change, it is important to understand why some microbes are found in one location rather another, and how their biodiversity shifts over time (Brown et al., 1998; Martiny et al., 2006). The Baas-Becking hypothesis proposes that microbes are found "everywhere" and that the "environment selects" for those that can thrive, and ultimately contributes to variation in microbial diversity (Baas-Becking, 1934). There are different processes – such as environmental filtering and selective pressures from resource availability, climate, and competition – that drive microbial community assembly and select for advantageous traits within populations (Ackerly, 2003; Martiny et al., 2006; Kraft et al., 2015; Friedman et al., 2017). Many microbes possess climate response traits that are deeply conserved within their phylogeny and evolutionary history (Isobe et al., 2020). Some of these traits are conserved at more shallow levels, (Martiny et al., 2013; Chase et al., 2018; Isobe et al., 2019), even within strains (Chase et al., 2021). All these traits may play a role in microbial responses to

climate change at the community level (Glassman et al., 2018). Investigating microbial genetics can inform the functional traits driving physiological responses to environmental perturbations and improve predictions of microbial response to climate change.

My dissertation seeks to assess microbial genetic and functional responses to climate change. Throughout the dissertation, I analyzed data from a field study that simulated climate change by reciprocally transplanting microbial communities from a Southern California climate gradient. The climate gradient contained ecosystems that varied inversely with temperature and precipitation. In my first chapter, I investigated the variations in extracellular enzyme activity and litter decomposition of microbial communities from the Southern California climate gradient reciprocal transplant. I found that the physiological impacts of climate are potentially important for enzyme-mediated decomposition; however, climate specialization will not constrain microbial community response to shifting environments. In my second chapter, I explored the global distribution of the *Sphingomonas* bacterial clade to determine whether there is a relationship between habitat and phylogeny, and whether habitat preference is reflected in key genome-based traits. Using publicly available data, I found that *Sphingomonas* gene content and phylogeny reflects habitat preference. Finally in my third chapter, I returned to the Southern California climate gradient to inquire how *Sphingomonas* clade and functional composition shift after experiencing environmental perturbations from the reciprocal transplant. I found that *Sphingomonas* clade and functional composition change to reflect site conditions, and this change is consistent with previous bacterial research at other phylogenetic levels from the climate gradient. Altogether, my dissertation improves the understanding of how microbes respond to shifting environments.

CHAPTER 1

Microbial extracellular enzyme activity with simulated climate change

ABSTRACT

It is critical to understand the consequences of environmental change for the microbial regulation of carbon and nutrient cycling. Specifically, understanding microbial community traits, such as extracellular enzyme activity, can help inform nutrient cycling models and address knowledge gaps. We analyzed data on extracellular enzyme activities and litter decomposition from an 18-month experiment in which microbial communities were reciprocally transplanted along a climate gradient in Southern California. Communities were from desert, scrubland, grassland, pine–oak, and subalpine ecosystems. We aimed to test how enzyme activities responded to climate change following transplantation and how those responses related to decomposition rates. We hypothesized that microbial communities would specialize on their native climate conditions, resulting in higher enzyme activities when transplanted back into their native climate. We investigated the relationship between extracellular enzyme Vmax values, substrate mass loss, and microbial biomass as well as variation in these variables across the climate gradient. We found little evidence for climate specialization, and there was rarely a reduction in enzyme functioning after microbial communities were transplanted into new climate conditions. Moreover, observed differences in decomposition were not related to changes in extracellular enzyme potential, although there were significant differences in enzyme activities and decomposition rates across sites. These results suggest that direct, physiological impacts of climate are likely to be important for enzyme-mediated decomposition, but climate specialization will not constrain the microbial response to climate change in our system.

INTRODUCTION

Understanding the implications of global warming for decomposition and subsequently, carbon and nutrient cycling is an urgent need (Cavicchioli et al., 2019). Microbial decomposition depends on climate variables, such as temperature and precipitation, that directly affect the abiotic environment (Allison 2006; Chapin et al. 2012; Swift et al. 1979). Climate may also have indirect effects on decomposition by altering substrate inputs (Hobbie, 1992). Therefore, investigating the impacts of climate change on microbial decomposition is critical for predicting carbon and nutrient cycling.

There is increasing evidence that decomposition responses to climate change may depend not only on abiotic conditions, but also on microbial community composition. For example, some studies have found that plant litter from a given site is decomposed more rapidly by microbial communities from that same site (Gholz et al., 2000; Ayres et al., 2009). Still, the mechanisms underlying such community effects remain unclear, making it difficult to predict how soil carbon and nutrient cycling will respond to climate change across ecosystems.

One of the key community-level traits relevant to decomposition is the production of extracellular enzymes. Microbes produce extracellular enzymes to break down complex polymers into soluble carbon and nutrients that can be taken up and used to fuel microbial metabolism (Burns et al. 2013; Sinsabaugh and Moorhead 1994). The costs of enzyme production may result in specialization to produce only enzymes that degrade the specific litter compounds in the local environment (Allison et al. 2010). In a new environment with different litter chemistry, those specialized enzymes might be less effective, leading to a “home field advantage” for the native microbial community (Gholz et al., 2000; Ayres et al., 2009).

Although there is evidence that microbial communities and their extracellular enzymes specialize on different litter substrates, it is less clear if communities and enzymes also specialize on different climate conditions. In previous studies with reciprocal litter transplants, both climate and litter substrates varied across sites, making it difficult to disentangle these potential drivers of community specialization (Ayres et al. 2009; Gholz et al. 2000). Like substrate, climate variables could select for production of extracellular enzymes with site-specific properties that affect decomposition activity. For example, extracellular enzyme temperature or moisture sensitivity might be tuned to the local climate in a way that alters enzyme kinetic properties (Alster et al., 2020).

Testing for climate specialization requires holding litter substrate constant and measuring the performance of microbial communities across different climate conditions. Recently, Glassman et al. (2018) conducted such a test using a reciprocal transplant design along a climate gradient in Southern California with desert, scrubland, grassland, and forest ecosystems (Glassman et al., 2018). Although there was no evidence for climate specialization with overall decomposition (i.e. home field advantage based on climate), there were microbial community effects on decomposition rates (Glassman et al., 2018). The microbial community from the grassland ecosystem decomposed a common litter substrate fastest in the scrubland ecosystem and vice versa.

Building on the Glassman et al. (2018) study of decomposition rates and microbial composition, our goal was to test whether microbial enzyme activities showed climate specialization. Using Glassman et al.'s (2018) reciprocal transplant design, we simulated climate change effects on microbial communities while holding litter substrate constant. We tested the climate specialization hypothesis, which predicts that extracellular enzyme activities of native

microbial communities should be greater than those of non-native microbial communities (Figure 1.1). Although Glassman et al. (2018) found no evidence for climate specialization with overall litter decomposition, we tested whether the decay rates of specific litter chemical compounds were specialized to climate and related to specific extracellular enzyme classes or microbial biomass.

METHODS

Climate Gradient

Our climate gradient includes five sites in desert, scrubland, grassland, pine-oak, and subalpine ecosystems (Figure 1.2; Table S1.1; Baker and Allison 2017). Temperature and precipitation vary inversely along the gradient: the desert site shows the hottest (22.8 ± 0.8°C) and driest climate (100 ± 24 mm mean annual precipitation) and the subalpine is the coldest site (10.3 ± 1.8 °C) with the most precipitation (~265 mm; Table S1.1).

Reciprocal Transplant Design

Microbial communities from each site along the climate gradient were transplanted into all sites to simulate climate change (Figure 1.2). A microbial community inoculum was created by collecting four samples from each of the sites on September 11, 2015, and homogenizing the samples within each site. Fifty mg of the site inoculum was added to 5 g irradiated, homogenized grassland leaf litter in sterilized litter bags. To keep the experiment manageable, we only used senesced grass as the litter type because grasses were present at all climate gradient sites. The grassland leaf litter was homogenized with coffee grinders. The litter bags were nylon membrane

bags with 0.22 µm pores (cat. no. SPEC17970; Tisch Scientific Cleaves, Ohio, USA) that allowed nutrients and water to travel freely though the bag. However, the bags did not allow for the movement of microbes and leaf litter.

On October 19, 2015, 300 bags were deployed along the gradient (4 replicates x 5 sites x 5 microbial community inocula x 3 timepoints = 300 bags). This design and level of replication had the statistical power to detect significant site by inoculum interactions that explained as little as 16% of the variation in mass loss within a timepoint (Glassman et al., 2018). The bags were placed in four 1 m x 1 m plots in each ecosystem and destructively sampled every 6 months over an 18-month period. At each of the three timepoints, 100 bags were collected (Timepoint 1: April 5, 2016; Timepoint 2: October 24, 2016; Timepoint 3: April 18, 2017). For information on microbial community taxonomic composition and shifts in litter chemistry, see Glassman et al. (2018).

Substrate Mass Loss

We re-analyzed litter chemistry data collected previously by calculating the ash-free mass and concentrations of cellulose, hemicellulose, lignin, crude protein, and starch (Baker and Allison, 2017; Glassman et al., 2018). For individual substrate mass loss, we subtracted the Timepoint 2 substrate mass from the initial mass of each substrate. Litter chemistry data were only collected for the initial litter and Timepoint 2. Statistical tests were conducted on the mass change, not the percent loss. Final substrate mass values are available on Github: https://github.com/stevenallison/UCIClimateExperiment/

Microbial Biomass

Fungal hyphal lengths were measured with microscopy, and bacterial cell counts were measured using flow cytometry (Glassman et al., 2018). Subsequently, both hyphal lengths and bacterial counts were converted to carbon content per gram litter (g C/g) using previously established methods in the literature described by Baker et al. (2017) and Alster et al. (2013). Bacterial cells were considered to have C density of 2.2×10^{-13} g μm^{-3} and treated as spherical with a radius of 0.6 µm (Bratbak, 1985). We assumed fungal hyphae had a diameter of 5.2 µm, 40% C in dry mass, 33% dry mass, and a density of 1.1 g cm^{-3} of hyphae (Paul and Clark, 1996). Microbial biomass is reported as the sum of both bacterial and fungal biomass. Fungal hyphal lengths were not measured at Timepoint 3, so fungal and total microbial biomass are only reported for Timepoints 1 and 2. Fungal and bacterial abundance data are also available on Github: https://github.com/stevenallison/UCIClimateExperiment/

Extracellular Enzyme Assays

Fluorometric and oxidative assays measuring the extracellular enzyme activities of α-glucosidase (AG), β-glucosidase (BG), β-xylosidase (BX), cellobiohydrolase (CBH), leucine aminopeptidase (LAP), and polyphenol oxidase (PPO) were conducted using previously published methods (German et al., 2011; Baker and Allison, 2017). Litter samples (0.4 g) stored at -80°C were thawed and combined with 150 mL maleate buffer (2 5mM) at pH 6.0 to create a homogenate. Next 125 µL of litter homogenate was combined with 125 µL of fluorometric substrate solution in microplate wells. Standards and substrate solution were made in maleate buffer. The assays were incubated at 4, 16, 22, 28, or 34°C for 4 hours. Enzymes were also assayed at a range of 8 substrate concentrations at each of the temperatures (Table 1.1). A plate

reader was then used to measure substrate fluorescence or absorbance values. We calculated activity as μmol hr^{-1} g^{-1} dry litter based on standard calculations (German et al., 2011).

Activities from the assays were fit to the Michaelis-Menten model using the "nls" R package to obtain maximum reaction velocities, V_{max} (German et al., 2011; Baty et al., 2015). The Michaelis-Menten model describes enzyme reaction velocity as a function of substrate concentration. We fit natural log-transformed V_{max} values at all the temperatures to a linear model and extracted the natural log V_{max} at 22°C from the model for further analysis.

Statistical Analysis

All statistical analyses were completed with R version 4.1.0 (R Core Team, 2021). First, we performed multivariate analysis of variance (MANOVA) on natural log-transformed potential extracellular enzyme V_{max} to test the fixed effects of site, inoculum, and time on overall enzyme activities. Analysis of variance (ANOVA) was then applied to the individual extracellular enzyme V_{max} values using the "car" package in R (Fox and Weisberg, 2019) to test for site, inoculum, and time effects. Also using the R "car" package, Type III tests were applied to enzymes that showed a significant three-way interaction, and Type II tests were applied to those that did not (Fox and Weisberg, 2019). This sequential approach was used to avoid type-1 errors in hypothesis testing. Post-hoc ANOVAs on individual enzymes were only run to determine the enzymes driving significant results from the MANOVA.

A significant interaction effect ($p < 0.05$) between site and inoculum may indicate support for climate specialization. However, climate specialization also requires that the native community V_{max} is significantly higher than the non-native communities' V_{max} values. Most

often, the extracellular enzymes did not have a three-way interaction effect, so we averaged the extracellular enzyme potential activity across timepoints for statistical analyses (Figures S1.1-S1.3). LAP was the only enzyme with a significant three-way interaction effect, so we also analyzed LAP activity without averaging over time (Figure S1.4). We applied pairwise comparisons, specifically Tukey post hoc analyses, to test for significant differences among communities within a site. We performed the Tukey post hoc analyses with "multcomp" package, using the "glht" (general linear hypotheses) and "cld" (compact letter display) functions (Hothorn et al. 2008). Again, to avoid type-1 errors, post-hoc Tukey comparisons were only run if the enzyme ANOVA showed a significant site by inoculum interaction.

To further test for mechanistic relationships between extracellular enzymes and substrate mass changes, we performed Pearson's product-moment correlation tests between natural log-transformed extracellular enzyme V_{max} values and their associated substrate mass losses. Similarly, we applied correlation tests to examine the relationship between natural log-transformed microbial biomass and substrate mass loss. We used the base R "stats" package with the "cor.test" function for the correlation tests (R Core Team, 2021).

RESULTS

Potential Extracellular Enzyme Activities

MANOVA results indicated that site, inoculum, and time had significant ($p < 0.001$) effects on the collective, natural log-transformed extracellular enzyme V_{max} values (Table 1.2). Furthermore, there were significant ($p < 0.05$) interactions between the fixed factors. ANOVA results indicated that inoculum, site, and timepoint each had significant effects on activity for all

enzymes except BX, which did not show a significant timepoint effect (Table 1.2). There was a significant site by timepoint interaction for all extracellular enzymes, except those that act on cellulose, BG and CBH (Table 1.2). All enzymes, except PPO, showed a significant inoculum by site interaction (Table 1.2). Further Tukey post-hoc tests only supported climate specialization in the scrubland site for BX and in the subalpine site for BG, LAP, and PPO (Figures S1.1-S1.3). LAP was the only EE with a significant timepoint x inoculum x site interaction effect ($p < 0.05$; Table 1.2). For LAP, potential activities at all sites and timepoints revealed only two potential instances of climate specialization: the subalpine site at Timepoint 1 and the desert site at Timepoint 3 (Figure S1.4).

Extracellular Enzyme Activities and Mass Loss

Compared to other substrates, cellulose had the highest mass loss, similar to the total mass loss patterns (Figure 1.3a; Glassman et al., 2018). At Timepoint 2, there were no significant differences in BG V_{max} values across the transplanted microbial communities within each site that would explain differences in cellulose decomposition (Figure 1.3b). Furthermore, at Timepoints 1 and 2, the enzymes and their respective substrate mass changes did not exhibit any significant ($p < 0.05$) correlations (Table S1.2). Only AG V_{max} at Timepoint 1 showed a positive correlation with starch decomposition.

Microbial Biomass and Substrate Mass Loss

Overall, the strongest positive correlation with individual substrate mass loss was observed with bacterial biomass (Table 1.3). Most often, the strongest positive correlations

occurred between the bacterial biomass at Timepoint 1 and substrate mass loss at Timepoint 2 (Table 1.3). Protein mass loss had the strongest positive correlation with bacterial biomass at Timepoint 2. In contrast, lignin mass loss at Timepoint 2 had a negative correlation with bacterial biomass at Timepoint 1 (Table 1.3). A negative correlation means there is a higher bacterial biomass with lower lignin mass loss, suggesting that the microbes present in the early stages of decay, mainly bacteria, are not lignin-degraders. Later, at Timepoint 2, there was a positive correlation with the bacterial biomass and lignin mass loss. The total microbial biomass showed mass loss relationships similar to bacterial biomass, but not fungal biomass (Table 1.3). Even when the fungal biomass exhibited a significant correlation — as seen with hemicellulose, lignin, and starch at Timepoint 2 — a similar correlation was not observed in the total microbial biomass correlations (Table 1.3).

DISCUSSION

Understanding the microbial community functional response to climate change is especially important for predicting future carbon and nutrient cycling. We hypothesized that microbial communities specialize on climate conditions, meaning that their extracellular enzyme activities should be greatest in their native climate. Instead, we found minimal evidence for climate specialization and no relationship between enzymes and substrate mass loss (Figure 1.3). Surprisingly, non-native communities often met or exceeded the V_{max} values of native communities within a site. To the best our knowledge, this result is the first field test of climate specialization with microbial extracellular enzymes. A previous study along our climate gradient found that access to native microbes did not increase litter decomposition or enzyme V_{max}, but that study's design could not be used to test for climate specialization (Baker et al., 2018).

There are several potential explanations for a lack of climate specialization in our study. Rather than being climate specialists, microbes and their enzymes may have broad climate tolerances given the high climatic variability within sites along our gradient. It is also possible that there is climate specialization of some microbes and traits, but not enzyme activities (Chase et al., 2021). Notably, other studies have also found a lack of support for microbial community specialization on litter chemistry. John et al. (2011) rejected the home field advantage hypothesis in a grassland-forest reciprocal transplant. Furthermore, home field advantage was not observed for leaf litter decomposition in the Atlantic rainforests of Brazil (Gießelmann et al., 2011).

For all enzymes, there was statistical evidence for activity differences across inoculated communities; in most cases, there were also significant inoculum x site interactions (Table 1.2). A few of these interactions were consistent with climate specialization, such as slightly higher BG activity in the subalpine community in the subalpine site. More frequently, though, non-native communities matched or exceeded the V_{max} values of native communities—in the scrubland site for example, the desert community showed the highest BG activity (Figure S1.2). It is possible that these differences were driven by stochastic variation in community assembly or priority effects as the inoculated communities established on the grassland litter (Zhou and Ning, 2017; Albright et al., 2019). Seeing as we only measured decomposition of grassland litter, different patterns might emerge on different litter types. A forest litter type, for example, would have selected for a different set of communities that might have exhibited a different degree of climate specialization. Additional community transplants across climate gradients with other litter types would be needed to know for sure.

As with extracellular enzymes, we found no evidence for climate specialization in decomposition rates of specific chemical substrates. Furthermore, there was almost no

relationship between extracellular enzyme activity and decomposition, suggesting that other factors mediate variation in decomposition rates (Graham et al., 2016). For cellulose degradation, which involves a complex multi-enzyme system, it is possible that endoglucanases control the rate-limiting degradation step, whereas the CBH and BG enzymes we measured are active further downstream (Xie et al. 2007; Singh et al. 2016). Another possibility is that the abundance of specific microbial taxa controls decomposition rates.

Overall, our microbial biomass results suggest that bacteria promote the decomposition of non-lignin litter compounds (Wohl et al. 2004). We found positive correlations between bacterial biomass at Timepoint 1 and cellulose, hemicellulose, and starch mass loss, whereas lignin showed a negative correlation under the same conditions (Table 1.3). Protein mass loss at Timepoint 2 correlated positively with bacterial biomass at the second timepoint. In general, there were not strong correlations between fungal biomass and substrate decomposition, although we did find significant positive relationships with hemicellulose and starch loss, and a negative relationship with lignin loss (Table 1.3). These results support Glassman et al.'s (2018) finding that for grassland litter types, bacterial communities seem to have a stronger effect than fungal communities. Furthermore, a more detailed analysis of microbial functional composition could help explain why some enzyme potentials and decomposition rates varied across communities.

Taken together, our enzyme and decomposition results indicate that microbial communities vary in their enzymatic efficiency, defined as substrate mass loss per unit of extracellular enzyme activity (Alster et al. 2013). For instance, BG efficiency of cellulose degradation was lower for microbial communities in the desert site relative to other sites at Timepoint 2 (Figure 1.3). Interactions between microbial community members could explain variation in enzymatic efficiency and support enzymatic functioning outside the native climate.

Some microbes “cheat,” or benefit from extracellular enzymes without incurring the costs of enzyme production, which could limit enzyme efficiency, even in native communities (Allison 2005). Differences in production strategies or spatial structure across communities may result in different *in-situ* decomposition rates despite similar lab-measured extracellular enzyme potentials (Burns 2013). Additionally, resource acquisition traits and life history strategies vary between microbial communities, which could affect their overall efficiency of substrate turnover (Malik et al., 2020).

CONCLUSIONS

In summary, we found little support for climate specialization of extracellular enzymes. Correspondingly, we did not find much evidence for microbial community specialization to climate, and even if it did occur, there was no associated reduction in enzyme functioning with climate change. Likewise, there was no evidence for climate specialization in litter substrate decomposition, and the differences in decomposition that Glassman et al. (2018) observed across microbial communities were not driven by changes in extracellular enzyme potential. There were significant differences in enzyme V_{max} values and decomposition rates across sites, suggesting that direct, physiological impacts of climate change are likely to be more important than indirect effects from community compositional change. The microbial communities in our system appear to have a high degree of metabolic flexibility, meaning that climate specialization is not likely to constrain decomposition as climate changes.

ACKNOWLEDGEMENTS

We thank J.B.H Martiny, A.C. Martiny, D.R. Campbell, A.A. Malik, N.R. Baker, E. Abs, S.I. Glassman, and C. Weihe for their helpful discussions and comments. We thank the two anonymous reviewers whose comments improved the clarity of the manuscript. Finally, we thank D.M. Nisson, E. Stogner, G.A. Barajas, J. Allison, and M.Y. Farah for their hours of service performing enzyme assays for this project. This work was supported by the US National Science Foundation Ecosystem Studies Program (DEB-1457160) and the US Department of Energy, Office of Science, BER (DE-SC0020382).

FIGURES AND TABLES

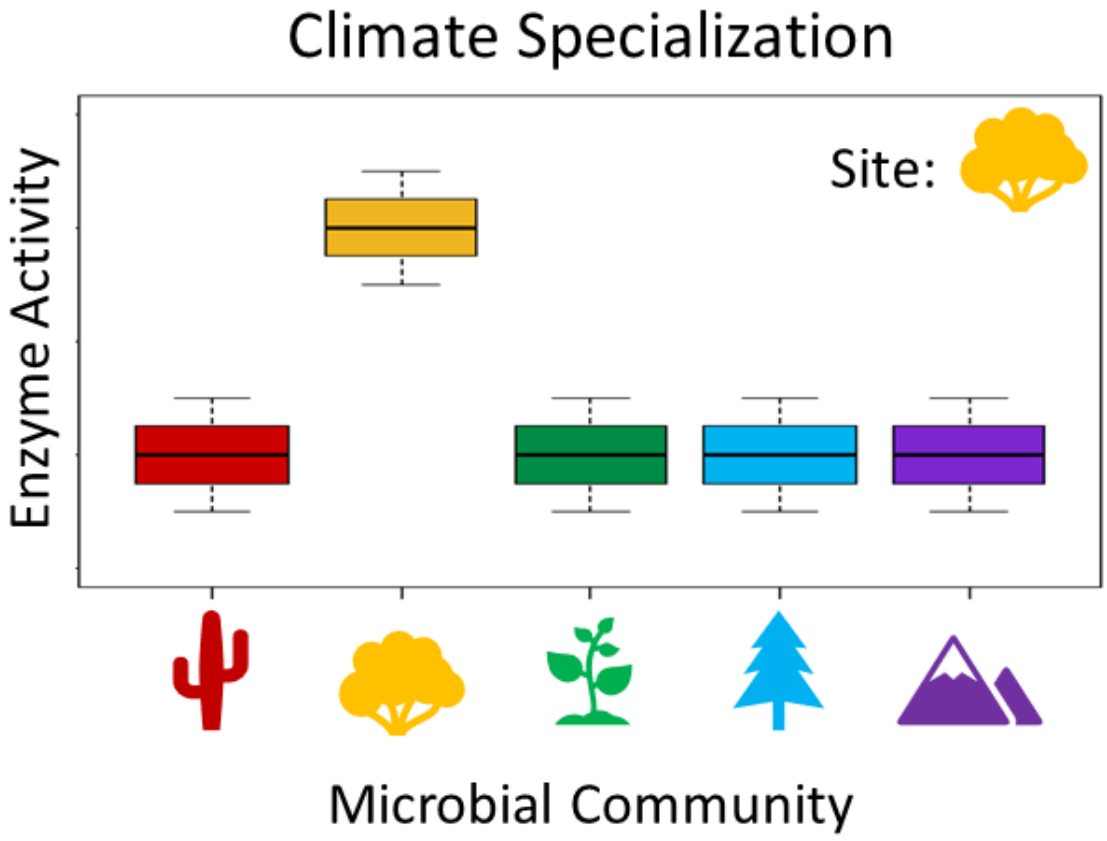


Figure 1.1. Hypothetical example of climate specialization in which the scrubland microbial community shows the highest extracellular enzyme activity in the scrubland site.

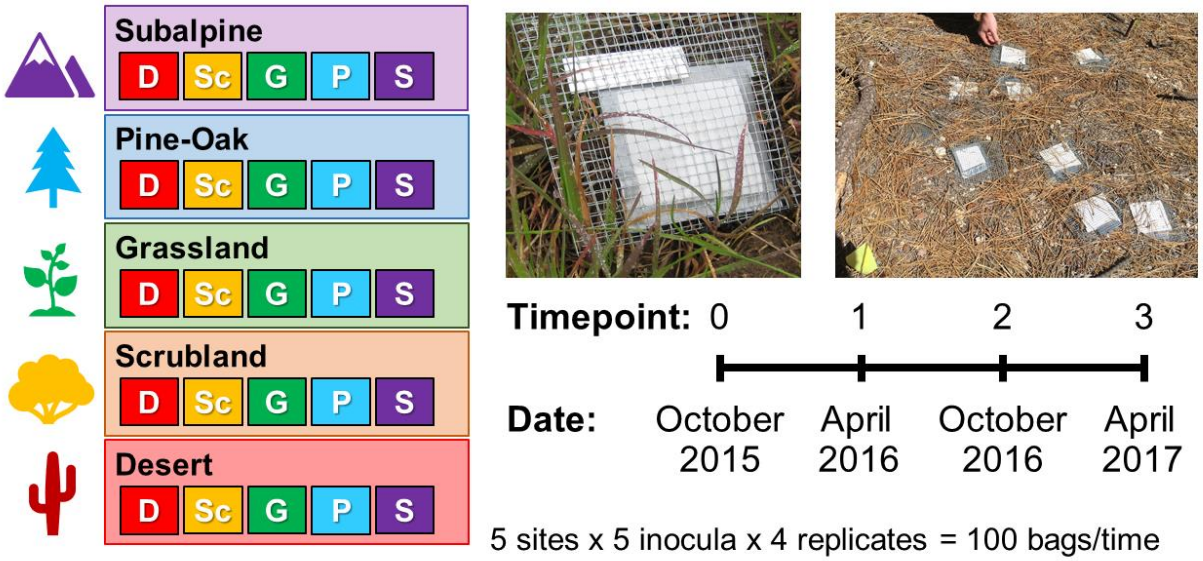
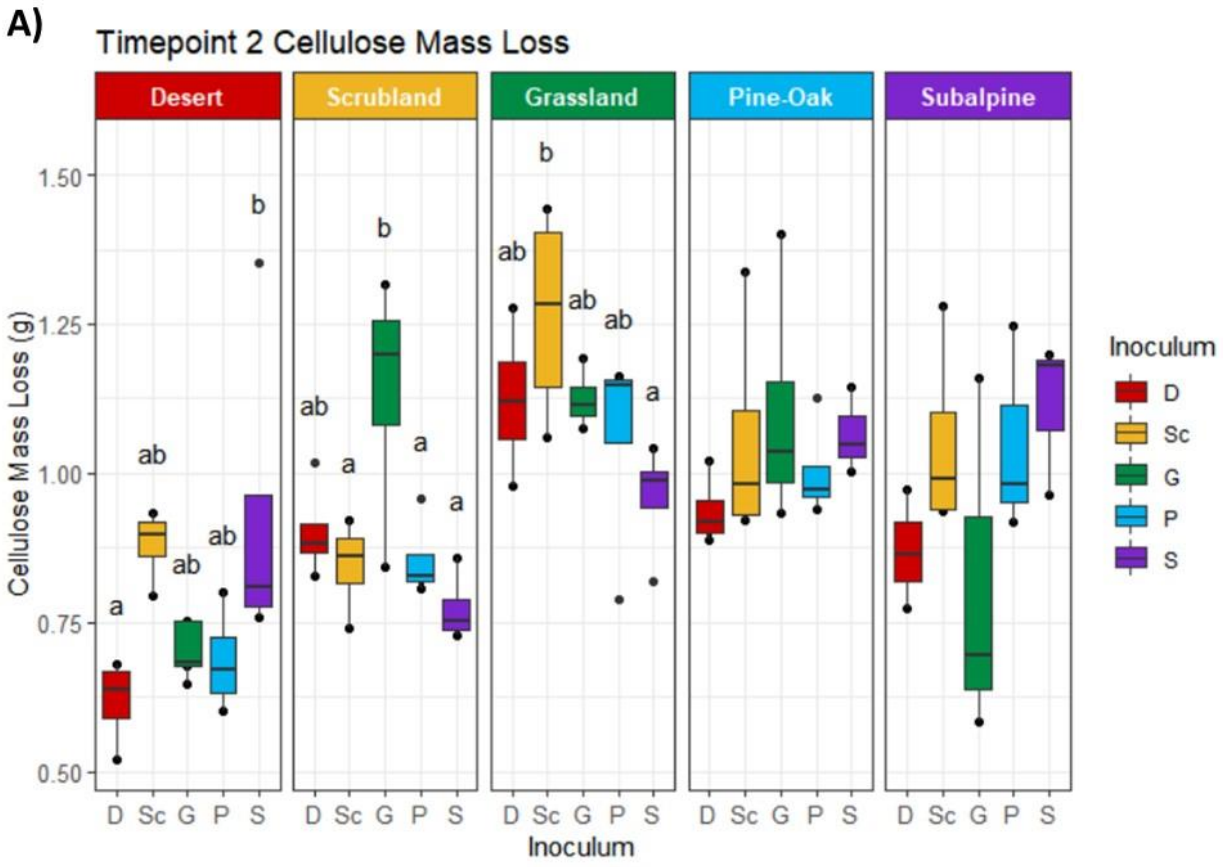


Figure 1.2. Climate gradient and experimental setup.

Microbial communities were reciprocally transplanted across 5 sites: desert (D), scrubland (Sc), grassland (G), pine-oak (P), and subalpine (S). These sites vary inversely with temperature and precipitation Table S1.1). Litter was collected from each site and individually inoculated onto gamma-irradiated grassland leaf litter substrate. Nylon membrane bags containing the inoculum and the substrate were placed in cages and randomly distributed within sites. The bags remained in the field for 18 months and were destructively sampled every 6 months.


A)
Timepoint 2 Cellulose Mass Loss
Desert
Scrubland
Grassland
Pine-Oak
Subalpine
Cellulose Mass Loss (g)
1.50
1.25
1.00
0.75
0.50
a
ab
ab
ab
b
ab
a
b
a
a
ab
ab
ab
a
b
D Sc G P S
Inoculum
Inoculum
D
Sc
G
P
S

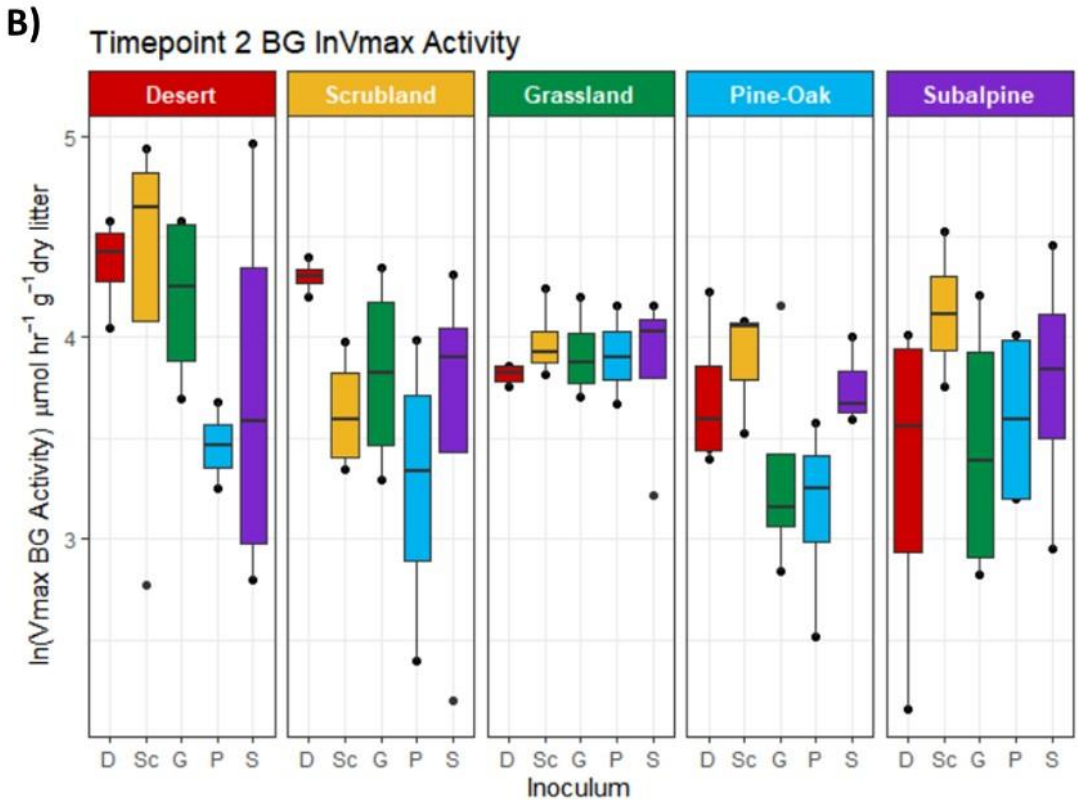
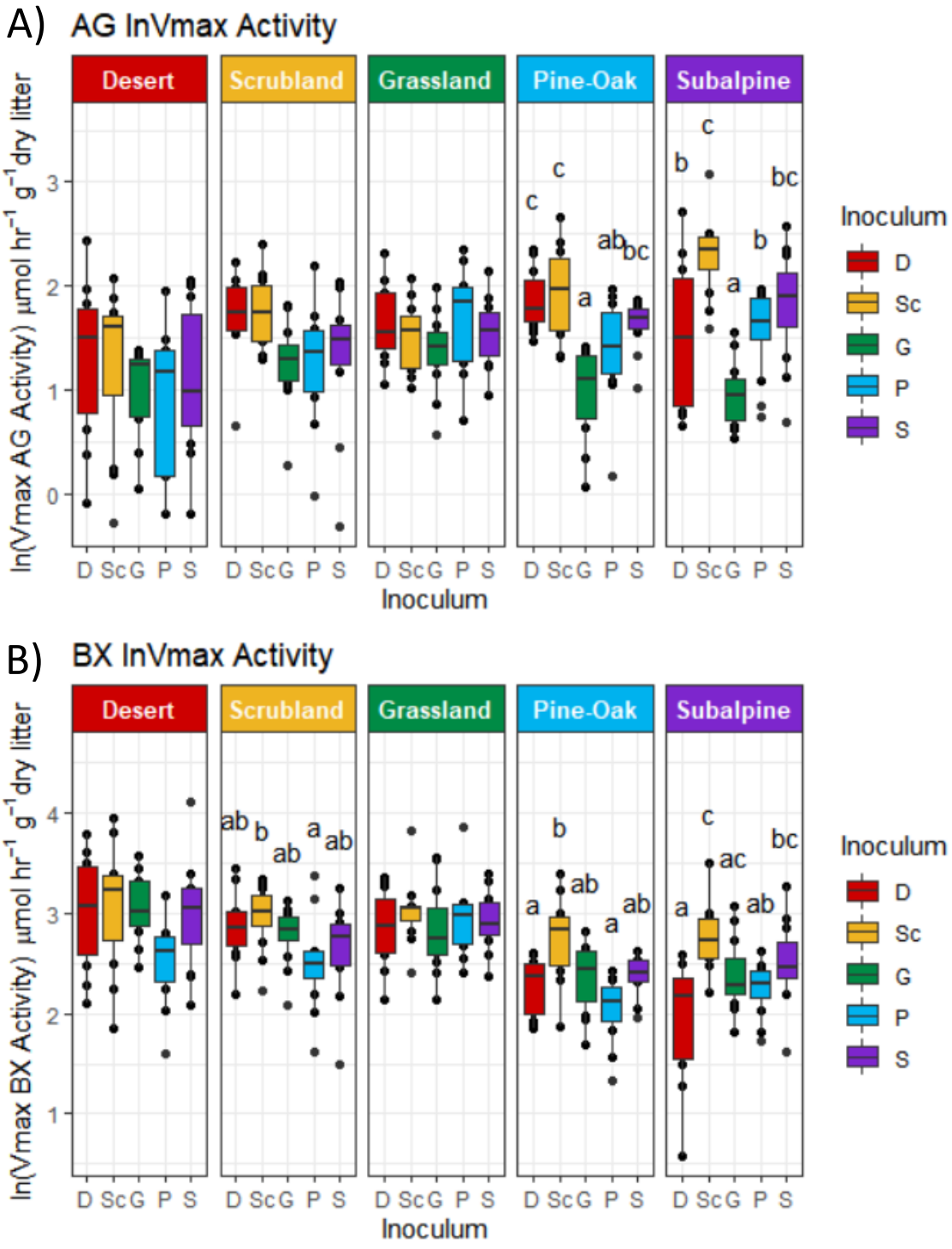
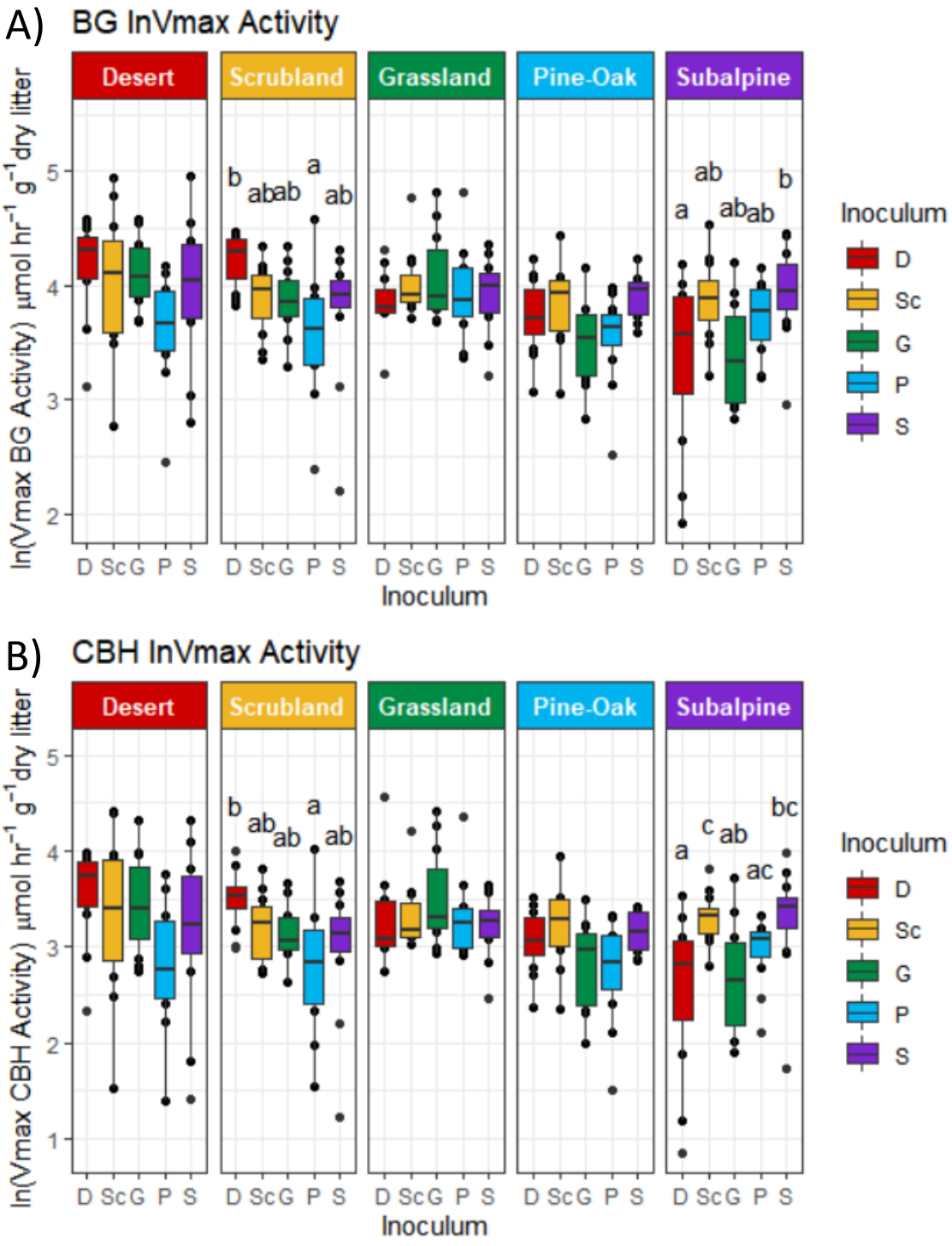

B)
Timepoint 2 BG lnVmax Activity
Desert
Scrubland
Grassland
Pine-Oak
Subalpine
ln(Vmax BG Activity) µmol hr^{-1} g^{-1} dry litter
5
4
3
D Sc G P S
Inoculum

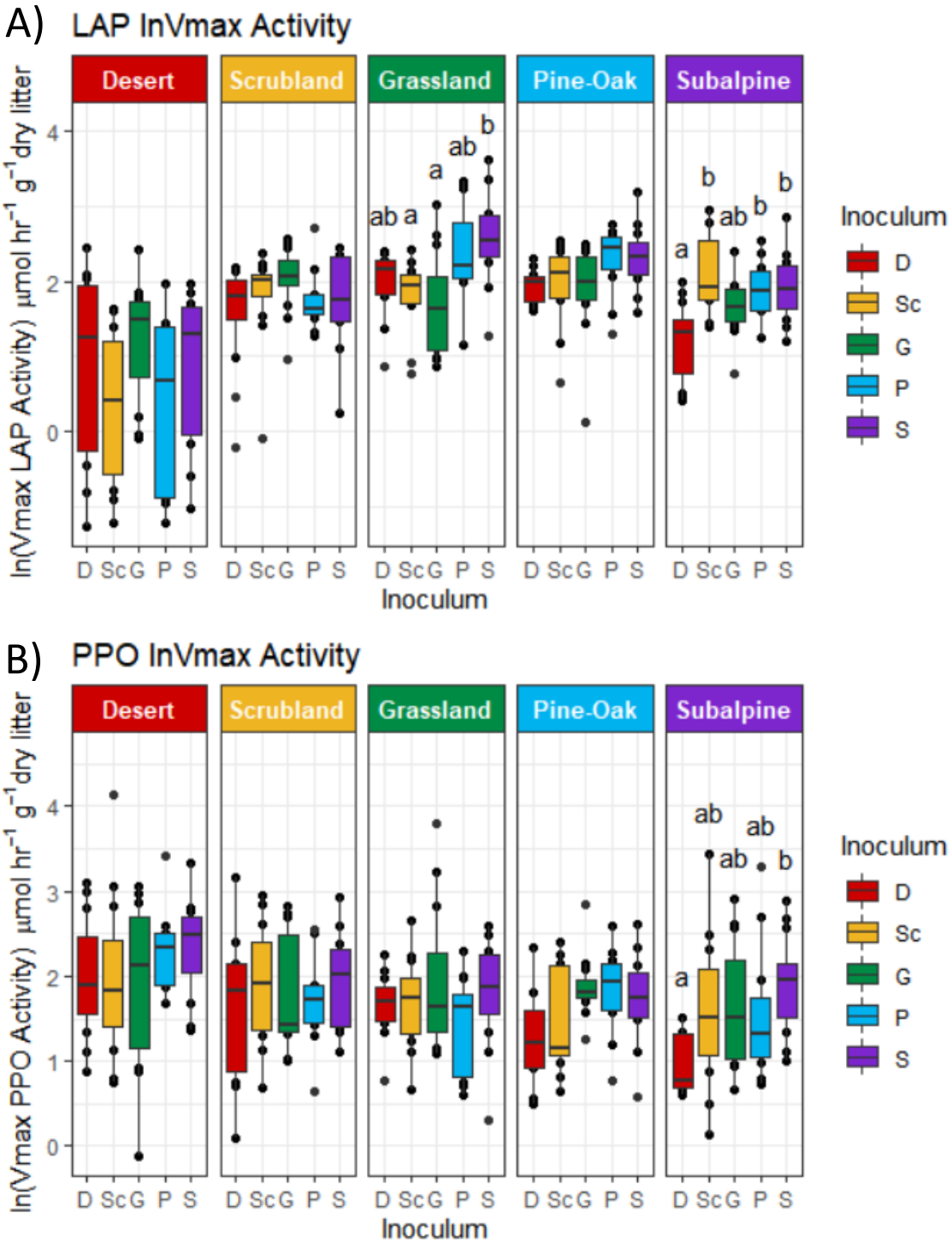
Figure 1.3. Extracellular enzyme activity and mass loss at Timepoint 2.
A) Cellulose-specific mass loss at Timepoint 2. **B)** Log transformed V_{max} of BG (β-glucosidase), a carbohydrate degrading extracellular enzyme, at Timepoint 2. Boxplots represent median values with the upper and lower quartiles. Whiskers cover the data range, excluding outlying data points. Within a site, inocula sharing the same Tukey letters above the boxplots are not significantly different from one another. Panels without Tukey letters did not show any significant differences. Extracellular enzyme activity did not reflect climate specialization since there were no significant differences in BG activity within sites. Additionally, BG activity did not have a significant correlation with cellulose mass loss (Table S1.2) and did not reflect substrate mass loss patterns.



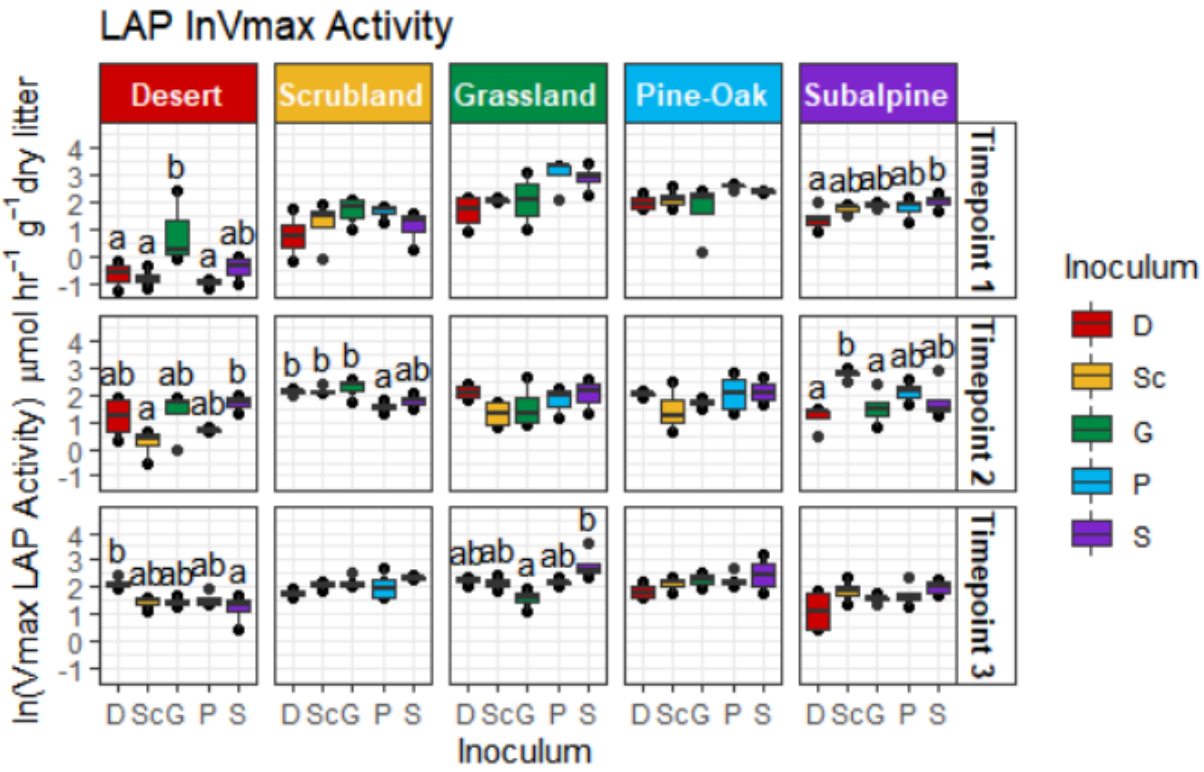
Supplemental Figure 1.1. Boxplots of natural log transformed **A)** AG and **B)** BX Vmax values averaged across time. Letters above inoculum indicate Tukey Test pairwise comparisons. Values that share the same letter are not significantly different within a site.



Supplemental Figure 1.2. Boxplots of natural log transformed **A)** BG and **B)** CBH Vmax values averaged across time. Letters above inoculum indicate Tukey Test pairwise comparisons. Values that share the same letter are not significantly different.



Supplemental Figure 1.3. Boxplots of natural log transformed **A)** LAP and **B)** PPO Vmax values averaged across time. Letters above inoculum indicate Tukey Test pairwise comparisons. Values that share the same letter are not significantly different.



Supplemental Figure 1.4. Boxplots of natural log transformed LAP faceted at each timepoint and site. Columns are sites and rows are times. Letters above inoculum indicate Tukey Test pairwise comparisons. Values that share the same letter are not significantly different.

Table 1.1. Enzyme classes, the substrates they act upon, and the synthetic substrates used in this study.

Enzyme Class	Substrate	Synthetic Substrate	[Substrate](μM)[a]
α-glucosidase	Starch	4-MUB-α-D-glucopyranoside	1.5625–200
β-glucosidase	Cellulose	4-MUB-β-D-glucopyranoside	3.125–400
β-xylosidase	Hemicellulose	4-MUB-β-D-xylopyranoside	3.125–400
Cellobiohydrolase	Cellulose	4-MUB-β-D-cellobioside	1.5625–200
Leucine aminopeptidase	Protein	L-leucine-7-amido-4-methylcoumarin hydrochloride	1.5625–200
Polyphenol oxidase	Lignin	Pyrogallol	7.8125–1,000

[a] The substrate concentration refers to the concentrations of synthetic substrate. There were 8 concentrations: the maximum concentration and 7 serial dilutions.

[a] The substrate concentration refers to the concentrations of synthetic substrate. There were eight concentrations: the maximum concentration and seven serial dilutions.

Table 1.2. MANOVA and ANOVAs p-values of natural log transformed microbial extracellular enzyme V_{max} at 22°C.

Variables	MANOVA	AG	BG	BX	CBH	LAP	PPO
Inoculum	**<0.001**	**<0.001**	**0.010**	**<0.001**	**0.003**	**0.001**	**0.001**
Site	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Timepoint	**<0.001**	**<0.001**	**0.001**	0.308	**<0.001**	**<0.001**	**<0.001**
Inoculum × Site	**<0.001**	**<0.001**	**<0.001**	**0.020**	**<0.001**	**<0.001**	0.407
Inoculum × Timepoint	**0.025**	0.257	0.226	0.251	0.081	**0.038**	0.553
Site × Timepoint	**<0.001**	**<0.001**	0.056	**0.002**	0.052	**<0.001**	**<0.001**
Inoculum × Site × Timepoint	**0.002**	0.090	0.832	0.802	0.413	**<0.001**	0.265

Only statistically significant *P* values are shown in bold font. AG = α-glucosidase; BG = β-glucosidase; BX = β-xylosidase; CBH = cellobiohydrolase; LAP = leucine aminopeptidase; PPO = polyphenol oxidase.

[a] Abbreviations: AG (α-glucosidase), BG (β-glucosidase), BX (β-xylosidase), CBH (Cellobiohydrolase), LAP (Leucine aminopeptidase), PPO (Polyphenol oxidase)

Table 1.3. Pearson correlations of substrate mass loss at Timepoint 2 and natural log transformed biomass.

Substrate[a]	Cellulose		Hemicellulose		Lignin		Protein		Starch	
Timepoint[b]	1	2	1	2	1	2	1	2	1	2
Bacterial[c]	**0.494***** [d]	N.S.	**0.581*****	N.S.	**−0.269****	**0.285****	N.S.	**0.436*****	**0.329****	N.S.
Fungal[c]	N.S.	N.S.	N.S.	**0.296***	N.S.	**−0.305****	N.S.	N.S.	N.S.	**0.352****
Microbial[c]	**0.501*****	N.S.	**0.589*****	N.S.	**−0.282****	**0.251***	N.S.	**0.442*****	**0.313****	N.S.

N.S. = not significant.

[a] Substrate mass loss at only Timepoint 2.

[b] Timepoint that the natural log transformed biomass was collected and tested against the substrate mass loss.

[c] Bacterial indicates only bacterial biomass, fungal indicates only fungal biomass, and microbial indicates the sum of fungal and bacterial biomass.

[d] Only statistically significant correlation coefficients (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$) are shown in bolded font.

Supplemental Table 1.1. Site description adapted from Baker and Allison 2017.

Sites	Latitude, Longitude	Elevation (m)	Mean Annual Temperature (°C)	Mean Annual Precipitation (mm)	Dominant Litter Type
Desert	33.65, -116.37	275	22.8 ± 0.8	100 ± 24	grasses, shrub leaves
Scrubland	33.61, -116.77	1280	15.6 ± 0.8	193 ± 33	grasses, needles, shrub leaves
Grassland	33.74, -117.70	470	16.4 ± .03	242 ± 76	grasses, forb leaves
Pine-Oak	33.81, -116.77	1710	12.3 ± 0.6	402 ± 118	needles, tree leaves
Subalpine	33.80, -116.69	3000	10.3 ± 1.8	~265	needles, tree leaves

Supplemental Table 1.2. Table of Pearson's product-moment correlation values of natural log transformed microbial EE activity and its associated substrate mass loss at Timepoints 1 and 2. Significant values ($p < 0.05$) are bolded.

EE	Timepoint 1	Timepoint 2
AG	**0.410**	-0.00257
BG	0.107	-0.0354
BX	0.242	-0.245
CBH	0.126	-0.0140
LAP	-0.153	0.0616
PPO	-0.250	0.305

CHAPTER 2

Variation in *Sphingomonas* traits across habitats and phylogenetic clades

ABSTRACT

Whether microbes show habitat preferences is a fundamental question in microbial ecology. If different microbial lineages have distinct traits, those lineages may occur more frequently in habitats where their traits are advantageous. *Sphingomonas* is an ideal bacterial clade in which to investigate how habitat preference relates to traits because these bacteria inhabit diverse environments and hosts. Here we downloaded 440 publicly available *Sphingomonas* genomes, assigned them to habitats based on isolation source, and examined their phylogenetic relationships. We sought to address whether: 1) there is a relationship between *Sphingomonas* habitat and phylogeny, and 2) whether there is a phylogenetic correlation between key, genome-based traits and habitat preference. We hypothesized that *Sphingomonas* strains from similar habitats would cluster together in phylogenetic clades, and key traits that improve fitness in specific environments should correlate with habitat. Genome-based traits were categorized into the Y-A-S trait-based framework for high growth yield, resource acquisition, and stress tolerance. We selected 252 high quality genomes and constructed a phylogenetic tree with 12 clades based on an alignment of 404 core genes. *Sphingomonas* strains from the same habitat clustered together within the same clades, and strains within clades shared similar clusters of accessory genes. Additionally, key genome-based trait frequencies varied across habitats. We conclude that *Sphingomonas* gene content reflects habitat preference. This knowledge of how environment and host relate to phylogeny may also help with future functional predictions about *Sphingomonas* and facilitate applications in bioremediation.

INTRODUCTION

Bacteria occur in a wide diversity of habitats, but the factors that control habitat preference are unclear (Fierer and Jackson, 2006; Martiny et al., 2006; Merino et al., 2019). Given that habitats vary in their abiotic and biotic conditions, different habitats may select for different organismal traits (Noble and Slatyer, 1977). These traits can be phylogenetically conserved (Martiny et al., 2013; Dolan et al., 2017; Isobe et al., 2019, 2020) or horizontally transferred (Ochman et al., 2000), but likely also represent trade-offs for life-history strategies. For environmental microbes, one way to organize these trade-offs is the Y-A-S framework, which posits that bacterial life-history strategies are driven by tradeoffs in resource allocation to growth Yield, resource Acquisition and Stress tolerance responses (Malik et al., 2020). Investigating functional traits related to the Y-A-S strategies has the potential to yield insights into factors that affect the distributions of microbial taxa.

Sphingomonas is an excellent bacterial genus to investigate the distribution of habitat preference traits because it is found in a wide range of habitats. Within the Proteobacteria phylum, the *Sphingomonas* genus are Gram-negative, strictly aerobic, chemoheterotrophic, yellow-pigmented bacteria that possess glycosphingolipids in their cell envelope (Balkwill et al., 2006; Yabuuchi et al., 1990). *Sphingomonas* species have been isolated from soils, plant roots, water distribution systems, human samples, and hospital machines (White et al., 1996; Leys et al., 2004). Some species cause animal disease, while others are antagonistic toward phytopathogenic fungi that infect commercially important plants (White et al., 1996). Additionally, *Sphingomonas* species have also been used on the International Space Station to aid the extraction of rare earth elements (Cockell et al., 2020). On planet Earth, *Sphingomonas* serves as biocatalyst for bioremediation and can be found in soils that are contaminated with

pollutants (Leys et al., 2004). Understanding the distribution of *Sphingomonas* is especially important because with appropriate management strategies, this lineage can be a tool to clean up polluted environments (Onder Erguven & Demirci, 2019). Furthermore, *Sphingomonas* is able to degrade cellulose and hemicellulose and is therefore involved in organic carbon decomposition (Koskinen et al., 2000). Hence, the distribution and functional abilities of *Sphingomonas* make it an ideal genus for investigating phylogenetic histories of habitat preference traits.

Despite the potential importance and widespread distribution of *Sphingomonas* species, there has not yet been a comprehensive, in-depth study of the comparative genomics and phylogenetics of the genus from a trait-based perspective. Most studies thus far look at the distribution and phylogeny of select genomes from 16S rRNA perspective, and often do not consider genome-based traits (Leung et al., 1999; Leys et al., 2004; Asaf et al., 2020). Moreover, the *Sphingomonas* genus classification is still evolving; *Sphingomonas* has five sub-genus classifications, and although additional strains continue to be identified, it is difficult to place them in specific clades (Takeuchi et al., 2001; Jogler et al., 2013; Asaf et al., 2020). Additionally, some *Sphingomonas* species have been shown to improve plant growth during stressful drought and salinity conditions (Halo et al., 2015; Asaf et al., 2017). Currently, there are knowledge gaps in the literature with respect to *Sphingomonas* phylogenetics, taxonomy, and genome mapping in the context of stress tolerance and bioremediation (Asaf et al., 2020). Therefore, it is useful to explore the phylogenomics of *Sphingomonas* from a whole-genome and trait-based perspective. Since *Sphingomonas* has important bioremediation qualities, understanding the genetics and distributions of these traits can provide preliminary knowledge towards harnessing *Sphingomonas* to rehabilitate natural habitats (Schmidt et al., 1992).

The knowledge of how environment and host influences genetics may also help with future functional predictions. In this study, we downloaded over 400 available *Sphingomonas* sequences from public databases, assigned them to a habitat based on where they were isolated from, and assessed their phylogenetic relationships. With this information, we sought to address two questions. First, are there significant relationships between habitat and phylogeny? Second, do key, genome-based traits demonstrate phylogenetic clustering and correlate with habitat preference?

We used the genome-based traits as proxies for the Y-A-S life history categories: growth yield, resource acquisition, and stress tolerance (Figure 2.1; Malik et al., 2020). For growth yield, we investigated the distribution of amino acid related enzymes, lipid biosynthesis proteins, and lipopolysaccharide biosynthesis proteins. Carbohydrate-active enzymes (CAZymes) reflected resource acquisition strategies. Finally, for stress tolerance traits we explored chaperones, folding catalysts, prokaryotic defense system genes, as well as peptidoglycan biosynthesis and degradation proteins. Collectively, these traits underlie habitat preference. We hypothesize that *Sphingomonas* strains from similar habitats will cluster together in phylogenetic clades. Furthermore, key traits that improve fitness in specific environments should correlate with the isolation habitat. For example, CAZyme genes should be most prevalent in genomes of *Sphingomonas* associated with plants, and prokaryotic defense system genes would be the highest in *Sphingomonas* genomes found at locations with a contaminant. Ultimately, these findings will improve the understanding of *Sphingomonas* distribution across habitats, as well as illuminate the link between habitat preference and life history strategies.

METHODS

Library collection and curation

We downloaded 440 publicly available *Sphingomonas* genomes from the PATRIC database on July 31, 2020 (Wattam et al., 2014) and used the metadata for each strait to identify the isolation source (Table 2.1). The sequences were categorized by their isolation source and assigned to one of these eight groups based on the strain description: animal (n = 10), clinical (n = 43), contaminated site (n = 13), industrial (n = 13), environmental (n = 54), plant (n = 68), water (n = 34), and other (n = 17; Table 2.1). More specifically, strains in the animal category were isolated from living, non-human sources. Strains in the clinical category came from hospital settings and included bodily samples from human beings, like blood. Any strain with the word "contaminated" in the description was placed in the contaminated site category. The environment category consisted of strains from abiotic, outdoor sources that were not water-based, like soils. The industrial category included samples from bioreactors, mines, and wastewater facilities (which contained the key phrase "activated sludge" in the description). Strains isolated from the plant kingdom were placed in the plant category; these strains were isolated from different parts of plants such as the seed, root, stem, and leaf. The water category consisted of strains isolated from a water source and sediments that did not include "contaminated" in its description. Finally, strains that could not be assigned to one of the previous 7 distinct groups were placed in the other category, such as samples from lichens and dust (Table 2.1).

Genomes with unspecified isolation sources were removed from our analyses. Next, we checked the completeness of the genomes against the *Sphingomonadales* order using the BUSCO (Benchmarking Universal Single Copy Orthologs) v4.1.4 program (Seppey et al., 2019).

Genomes with a BUSCO completeness score of less than 95% were filtered out. We used the online QUAST (Quality Assessment Tool for Genome Assemblies) server v5.0.2 to investigate the quality of the remaining genomes (Gurevich et al., 2013). From the initial genome library, 254 high quality *Sphingomonas* genome sequences remained for further analysis. These genomes consisted of 23 completed genomes and 231 fragmented genomes. All genomes were annotated with Prokka v1.14.6, and the core genes and accessory genes were identified with Roary v3.13.0 (Page et al., 2015; Seemann, 2014). For comparison to the larger subset that included fragmented genomes, we also used Prokka and Roary to quantify the pangenome for just the 23 completed genomes (Seemann, 2014; Page et al., 2015).

Outgroup Optimization

Zymomonas, *Rhizobium*, and *Rhodospirillum* are three closely related genera to *Sphingomonas* (Leys et al., 2004). To select the best outgroup or combinations of outgroups, we used Roary core gene counts. Specifically, we compared the core genes of the *Sphingomonas*-only ingroup to the core genes of the ingroup with various combinations of outgroups. We also included *E. coli* as a distantly related outgroup for further confirmation (Zhao et al., 2017). We selected *Rhodospirillum centenum SW* (GenBank Accession: CP000613) as an outgroup because it yielded a core gene count that was closest to the *Sphingomonas*-only ingroup. Furthermore, previous phylogenetic analysis (Leys et al., 2004) confirmed that *Rhodospirillum* is not part of the ingroup.

Reference Tree Visualization

We made a phylogenetic tree with core genes present in *Sphingomonas* genomes and the *Rhodospirillum* outgroup using methods from Rodriguez and Martiny (2020). In short, we ran Roary again, this time with a 50% blastp sequence identity for the *Sphingomonas* ingroup and *Rhodospirillum* outgroup. We identified 401 core genes and generated a bootstrapped Maximum Likelihood tree of the alignment with RAxML v8.2.12 with the PROTGAMMABLOSUM62 substitution model and 100 rapid Bootstrap searches (Stamatakis, 2014). Two of the 254 *Sphingomonas* sequences were removed from the analyses since RAxML deemed them identical. Therefore, we removed the duplicate sequences and re-ran Roary with the 252 *Sphingomonas* genomes to generate an alignment of 404 core genes. We used the core gene alignment to construct a phylogenetic tree with RAxML and subsequently visualized the tree with the iTOL v6.5 interactive tool (Figure 2.3; Letunic and Bork, 2019).

Clade Designation

We manually designated phylogenetic clades based on their divergence from the common ancestor. We marked the first clade by starting from the most distant, large monophyletic group. Subsequently, we moved along the tree until we came across another large, monophyletic group that was interpreted as another clade. Clades were defined in this manner until we identified a total of 12. Two strains that resembled an outgroup within two separate monophyletic clades were not included as part of the clade. We confirmed the clades and genome clusters by identifying pairwise average amino acid and nucleotide identities with the Enveomics tool (Rodriguez-R and Konstantinidis, 2016). Additionally, clades possessed a bootstrap identity of at least 86.

Genome-Based Traits

We quantified the abundances of genome-based traits involved in high growth yield, resource acquisition, and stress tolerance strategies. To identify the traits, we used the CAZy (Cantarel et al., 2009) and KEGG databases (Kanehisa and Goto, 2000). For CAZymes we looked at glycoside hydrolase and carbohydrate binding module abundances. Specifically we identified cellulase and glycoside hydrolase genes from Prodigal protein annotations using dbCAN2, a metaserver based on the CAZy database (Hyatt et al., 2010; Zhang et al., 2018). In our analysis, we only selected genes that were found with all three tools available on dbCAN2: HMMER, DIAMOND, and Hotpep. Additionally, we used the GhostKOALA v2.2 automatic annotation server to annotate the remaining genes based on KEGG Orthology (Kanehisa et al., 2016). We selected these genome-based traits for further analyses: lipopolysaccharide biosynthesis proteins, lipid biosynthesis proteins, amino acid related enzymes, prokaryotic defense system, peptidoglycan biosynthesis and degradation proteins, and finally chaperones and folding catalysts. These genes were grouped into the Y-A-S microbial life history trait-based framework developed by Malik et al., (2020) based on their role in high growth yield, resource acquisition, and stress tolerance strategies.

Statistical Analyses

After quantifying gene abundances, we natural log transformed the gene counts of the genome-based traits. Subsequently, we confirmed the normality of residuals using histograms and the Shapiro-Wilk tests, then ran Kruskal-Wallis rank sum tests to identify differences across habitats. We performed Kruskal-Wallis tests since not all the functional gene data were normally

distributed. Additionally, we conducted phylogenetic least squares (PGLS) statistical analyses to test whether there was an association between the habitat and the genome-based traits, independent of phylogenetic history (Mundry, 2014). We used PGLS statistics to also confirm if significant Kruskal-Wallis results were influenced by phylogenetic relatedness.

We used R v4.1.0 to run all the statistical analyses, and specially incorporated the "nlme", "geiger", "phytools", and "ape" packages (Revell, 2012; Pennell et al., 2014; Paradis and Schliep, 2019; Pinheiro et al., 2021)

Additionally, we ran ANOSIM tests to determine whether phylogeny was related to habitat preference. Using the "ape" package in R, we called the tree in R and subsequently used the "cophenetic" function in the "stats" package to calculate a distance matrix (Paradis and Schliep, 2019; R Core Team, 2021). Then, we used the "anosim" function in the R "vegan" to run ANOSIM tests (Oksanen et al., 2020).

RESULTS

Pangenome

We downloaded 440 publicly available *Sphingomonas* genomes, selected 252 high-quality genomes, and carefully curated them into 8 habitat categories based on the isolation source. Roary and Prokka pangenome analysis for the 252 *Sphingomonas* genomes revealed a total of 113,816 genes. Specifically, there were 444 core genes found in at least 99% of the genomes, 304 soft core genes found in 95 to 99% of genomes, 4,070 shell genes found in 15 to 95% of genomes, and 108,998 cloud genes present in less than 15% of genomes (Table 2.2).

When the *Rhodospirillum centum* outgroup was included in the pangenome analysis, there was a total of 115,874 genes with 404 core genes, 321 soft core genes, 4,091 shell genes, and 111,058 cloud genes (Supplemental Table 2.1; Figure 2.2). Some of the core gene functions include but are not limited to those associated with ribosomes, transcription factors, translation factors, and ATP synthases.

The pangenome analysis of just the 23 complete *Sphingomonas* genomes revealed a total of 33,131 genes comprised of 758 core genes, 184 soft core genes, 4,452 shell genes, and 27,737 cloud genes (Supplemental Table 2.2).

Phylogenetic Tree

Phylogenetic analysis of 252 *Sphingomonas* sequences with a *Rhodospirillum centum* outgroup yielded a phylogenetic tree assembled from an alignment of 404 core genes (Figure 2.3). The tree leaves clustered into 12 clades with a minimum bootstrap value of 86. After running Enveomics, pairwise comparisons within clades revealed a minimum average amino acid identity of 33.24% and a minimum average nucleotide identity of 76.37%.

Significant ANOSIM tests ($p < 0.05$) showed that *Sphingomonas* strains from the same habitat clustered together based on phylogeny, meaning that the distribution of taxa within a clade were not random and exhibited similar habitat preferences. For example, clade 7 was mostly composed of clinical samples, whereas clade 12 was dominated by strains from contaminated regions (Figure 2.3; Supplemental Figure 2.1). Some known lineages clustered in specific clades. Clade 2 contained *Sphingomonas melonis*, which is a pathogen of yellow Spanish melon fruits and causes brown spots (Buonaurio et al., 2002). Clade 3 included

Sphingomonas sanguinis, which causes dry rot of mango (Liu et al., 2018). *Sphingomonas naasensis* was found in clade 6 and was first isolated from forest soil in South Korea (Kim et al., 2014). Clade 7 contained *Sphingomonas koreensis*, which was first isolated from natural mineral water and can be a human pathogen in patients with meningitis (Lee et al., 2001; Marbjerg et al., 2015). Clade 8 included strains of *Sphingomonas japonica* (Supplemental Figure 2.2) that were isolated from the red king crab from the Sea of Japan (Romanenko et al., 2009). Moreover, *Sphingomonas* strains within the same clade shared similar clusters of accessory genes (Figure 2.2b).

Functional Genes

There was a significant correlation between genome size and gene counts for most of the traits we analyzed (Supplemental Figure 2.3). The largest genome with 6,899,075 bases belonged to a strain isolated from a contaminated site, and the shortest genome came from the animal classification with 2,861,323 bases. On average, genomes from contaminated sites were the largest and those from animals were the smallest (Supplemental Table 2.3). Similarly, genomes from contaminated sites typically had more genome-based traits, whereas strains from animals often had the fewest when compared to the other habitats (Figure 2.4). The prokaryotic defense system gene group was highest within contaminated habitats. There was also a high enrichment of chaperones and folding catalysts within genomes isolated from the clinical habitat; on average, the genome size of clinical strains was the second largest (Figure 2.4; Supplemental Table 2.3).

A subset of KEGG Orthology genes were chosen for investigating genome-based habitat preference traits based on their classification within the Y-A-S framework. Kruskal-Wallis rank sum tests between habitat and genome-based trait counts yielded significant ($p < 0.05$) differences for all traits except CAZymes and peptidoglycan biosynthesis and degradation proteins (Figure 2.4). Similarly, analysis of variance tests (ANOVA) on the Phylogenetic Generalized Least Squares (PGLS) models indicated that trait frequencies differed significantly ($p < 0.05$) by habitat for all traits except the prokaryotic defense system and CAZymes. Kruskal-Wallis rank sum tests between clades and genome-based trait counts also yielded significant ($p < 0.05$) differences for all traits.

DISCUSSION

Using comparative genomics, we investigated the association between *Sphingomonas* habitat and phylogeny. Our hypothesis that *Sphingomonas* strains from similar habitats would cluster together in phylogenetic clades was supported as depicted in the phylogenetic tree with a significant (ANOSIM test $p < 0.05$) association between habitat and phylogeny (Figure 2.3). Furthermore, within clades, strains shared similar accessory genes (Figure 2.2). Moreover, we found partial support for the hypothesis that key, genome-based traits related to fitness in specific environments would correlate with the isolation habitat (Figure 2.3). A closer investigation of functional genes associated with life history strategies revealed significant differences in gene counts across habitats (Figure 2.3). Some of the patterns reflected what we anticipated, while others did not. Ultimately, these findings bring us one step closer towards understanding the relationship between habitat preference and phylogeny.

The phylogenetic tree indicates that there is an association between *Sphingomonas* habitats and phylogeny, supporting our hypothesis that strains from similar habitats are more closely related (Figure 2.3; ANOSIM test $p < 0.05$). These findings are also supported in other bacterial systems such as *Bifidobacteria, Curtobacterium*, and *Xyllela fastidiosa* (Chase et al., 2018; Rodriguez and Martiny, 2020; Batarseh et al., 2022). It appears that abiotic factors as well as biological conditions, such as hosts, contribute to the environmental filtering and evolution of *Sphingomonas* within each habitat (Martiny et al., 2006; Kraft et al., 2015). Although there is not a perfect alignment between habitat and phylogeny, it is possible that our 8 habitat categories (Table 2.1) may be too broad or too narrow, or perhaps dispersal between sources influences the evolutionary history (Finlay, 2002). Most of the environmental samples consisted of soils, while the plant samples could be separated into the root, stem and leaves (Supplemental Figure 2.4). The rhizosphere consists of soils in the vicinity of plant roots, and could include lineages that are selected by both soil and plant properties (Berendsen et al., 2012). Additionally, dispersal between habitats could bring together *Sphingomonas* strains from different sources in the same location (Finlay, 2002; Albright et al., 2019; Walters et al., 2022). Dispersal is particularly likely across environment, plant, water, and contaminated site habitats.

Since we found that habitat preference is phylogenetically conserved, we sought to disentangle potential genome-based traits that underlie habitat preference. *Sphingomonas* clades share similar accessory genes (Figure 2.2), and genome-based trait counts varied by habitat, together suggesting that adaptation to the local environment has shaped habitat preference (Figure 2.4). Strains from contaminated sites had more genes associated with the prokaryotic defense system, while clinical strains had higher averages for chaperones and folding catalysts (Figure 2.4). It is possible that in the Y-A-S life history framework, strains from both of these

habitats may depend on stress tolerance strategies for survival (Figure 2.1; Malik et al., 2020). Chaperones and folding catalysts serve as signaling molecules to blood cells to promote immunity and inflammation (Henderson and Pockley, 2010), two common processes in clinical settings. Immune responses are stressful to bacterial infectious agents, and bacterial stress proteins such as chaperones may even trigger the immune response of hosts (Henderson et al., 2006). Moreover, compared to the other habitats, contaminated sites also had more genome-based traits associated with high growth yield (Figure 2.4). Since *Sphingomonas* can break down pollutants and has the potential for bioremediation (Schmidt et al., 1992), it is possible that strains in contaminated habitats invested in resource uptake rather than stress tolerance.

For traits that did not possess significant differences in their frequencies, such as CAZymes and peptidoglycan biosynthesis and degradation proteins, there are two potential possibilities (Figure 2.4). These traits may be part of the core genome and are required by all strains for basic functioning. Alternatively, there may be finer-scale differences in specific genes that are not detected because our traits are defined as broad sums of multiple genes. Moreover, proteins may have overlapping functions in metabolic pathways, making it difficult to assign them to a single life history strategy.

Although the genomics field and sequencing technologies have made tremendous advancement (Heather and Chain, 2016), there are still challenges with assembling complete genomes. In publicly available data, there will be differences in the quality of the genomes since sampling and sequencing methods vary across studies. Therefore, to mitigate variability, we were very selective with the *Sphingomonas* genomes that we decided to investigate further. Although we included fragmented genomes, all sequences had a minimum BUSCO score of 95% from the *Sphingomonadales* order (Seppey et al., 2019). Still, fragmented genomes may not

accurately reflect the total core gene count of *Sphingomonas* pangenome analysis due to missing genes. Therefore, for comparison, we performed pangenome analysis on the 23 complete *Sphingomonas* sequences in our dataset (Supplemental Table 2.2), revealing 758 core genes. This analysis indicates that our core gene count of 404 for the genus is reasonable. As the diversity and frequency of genomes increases, the number of core genes will decrease.

We investigated the genomic variation and phylogeny of *Sphingomonas* across different habitats. Additionally, we used a trait-based framework to explore differences in genome-based traits and life history strategies. We found that strains from similar habitats group together in clades and share accessory genes. Although our results did not reveal distinct life history strategies for all habitats, genome-based trait counts varied by habitat. These findings indicate that *Sphingomonas* gene content reflects habitat preference. Considering the relationships between habitat, genomics, and phylogeny may help us predict *Sphingomonas* habitat preference and better exploit its potential for bioremediation.

ACKNOWLEDGEMENTS

Thanks to the following people for discussion and feedback on the manuscript: Adam C Martiny, Alex B Chase, Brittni L Bertolet, Edwin Solares, Elsa Abs, Galen T Martin, Jennifer BH Martiny, José M Murúa Royo, Lucas Ustick, Luciana Chavez Rodriguez, Nicholas C Scales, Renaud Berlemont, and Tiffany N Batarseh. Thanks to Nadya Williams and the HPC Team for technical support and maintenance of the high-performance computing clusters. This project was funded by the US Department of Energy, Office of Science, Biological and Environmental Research, under award DE-SC0020382.

FIGURES AND TABLES

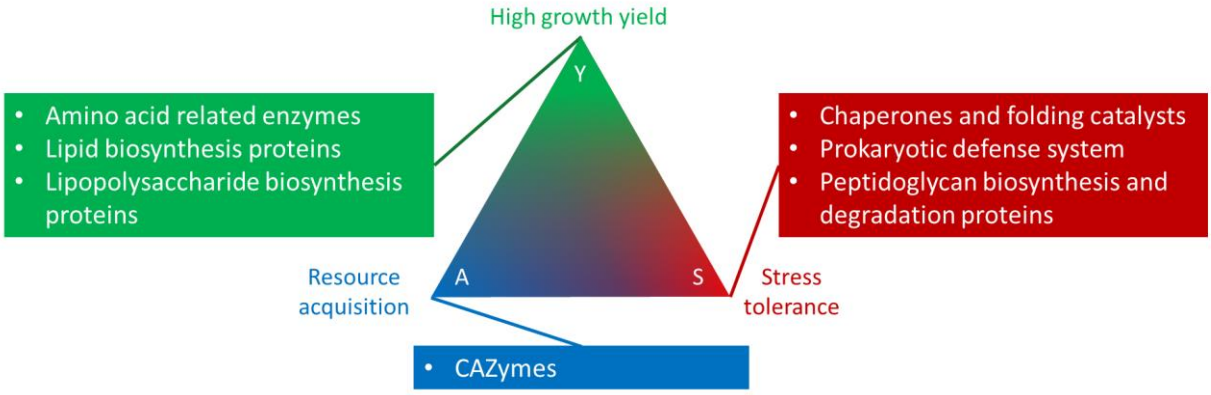


Figure 2.1. Genome-based trait groupings into the YAS life history strategy framework developed by Malik et al. (2020).

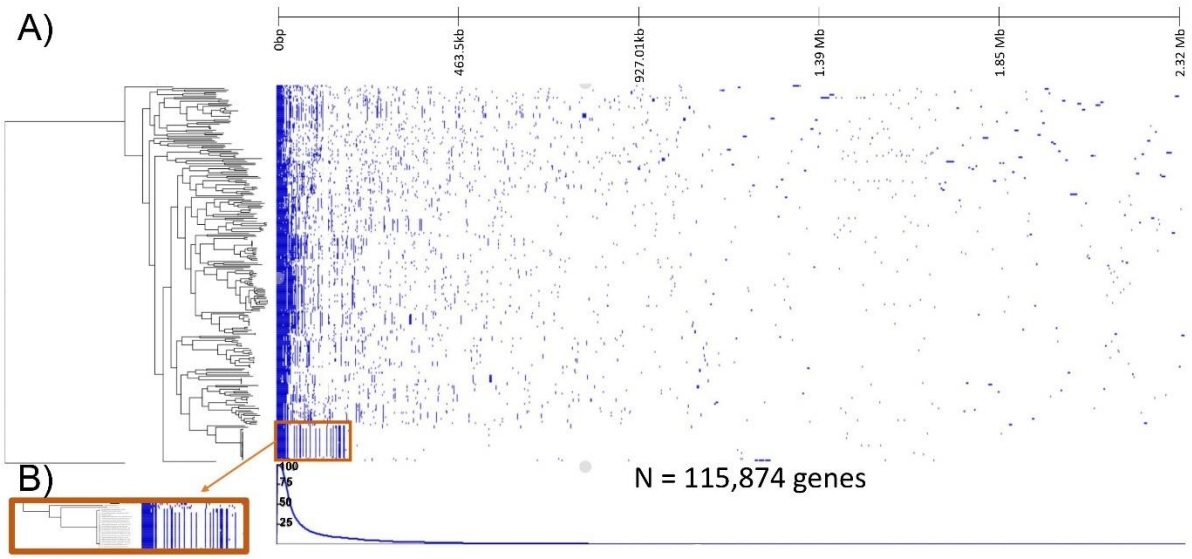


Figure 2.2. Pan-genome analysis of 252 *Sphingomonas* genomes and the *Rhodospirillum centum* outgroup. A) Gene presence-absence heatmap where vertical blue lines represent presence of a gene within rows corresponding to the *Sphingomonas* genome, and white reflects gene absence. The line graph underneath indicates the percentage of strains possessing the corresponding gene. B) Close-up of the gene patterns within a clade shows how clades contain similar gene clusters.

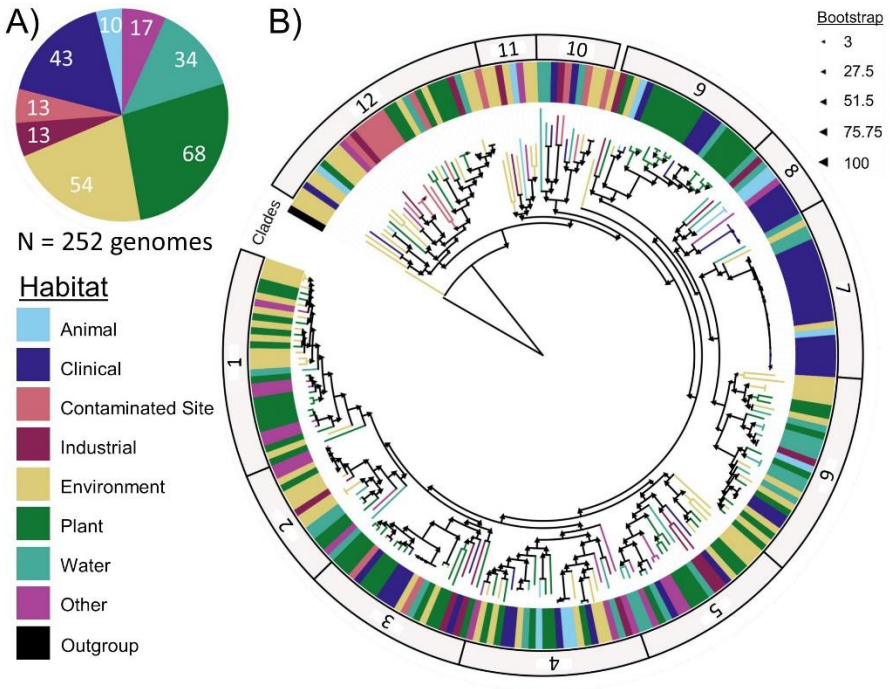


Figure 2.3. *Sphingomonas* A) habitat and B) phylogenetic tree constructed with 252 *Sphingomonas* genomes and 404 core genes, separated into 12 clades. The closely related *Rhodospirillum centum* was used as the outgroup to identify the core gene alignment and construct the tree. Significant ($p < 0.05$) ANOSIM results reflect that *Sphingomonas* sequences cluster together in clades by their habitat preference.

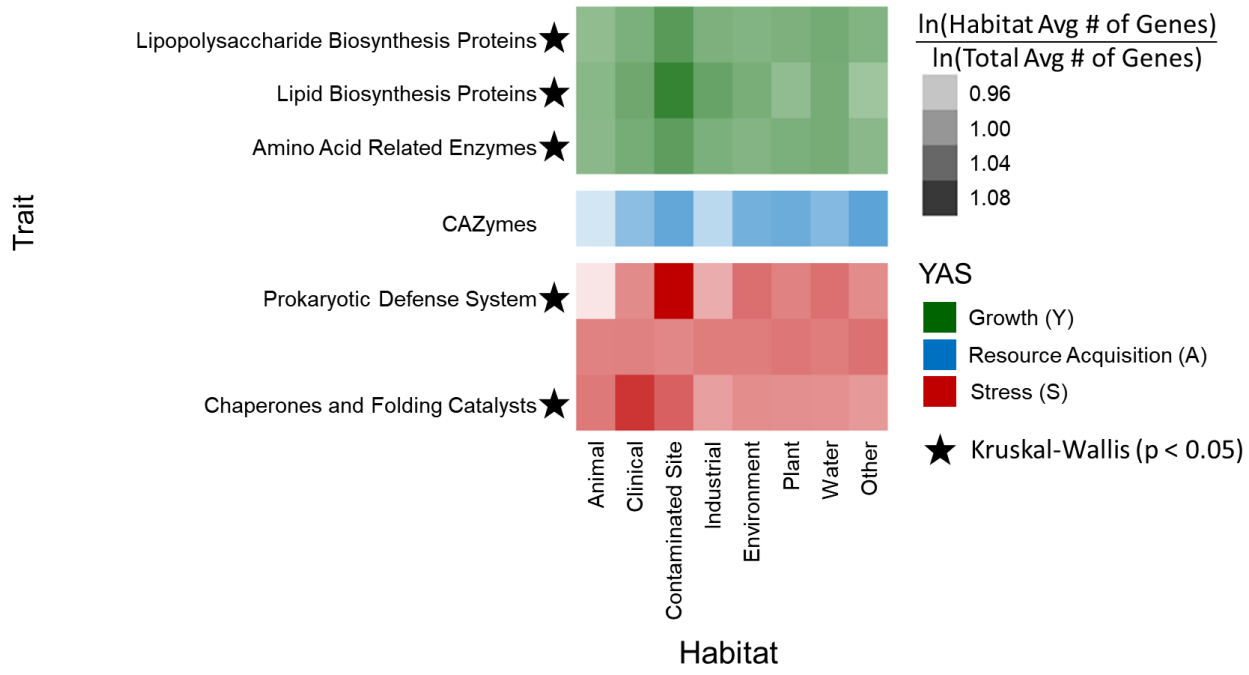
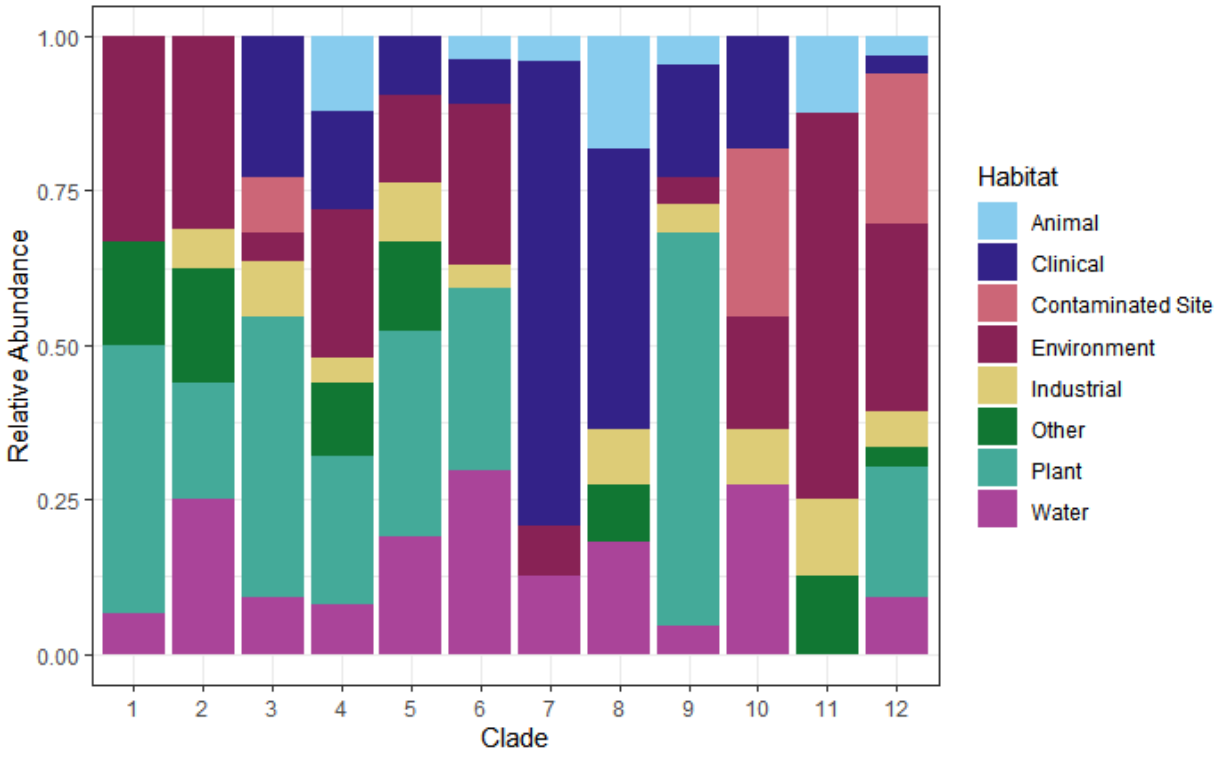
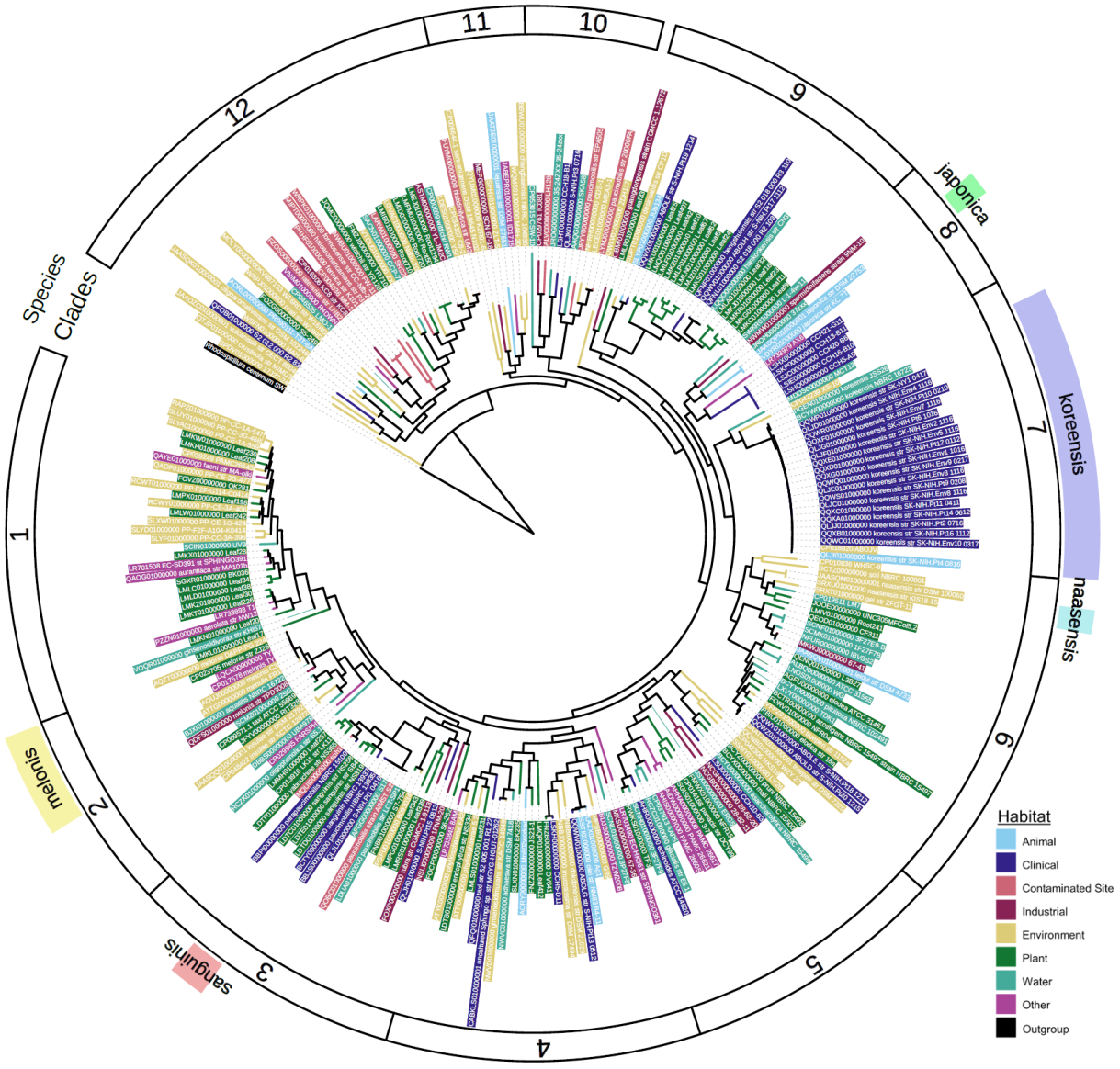


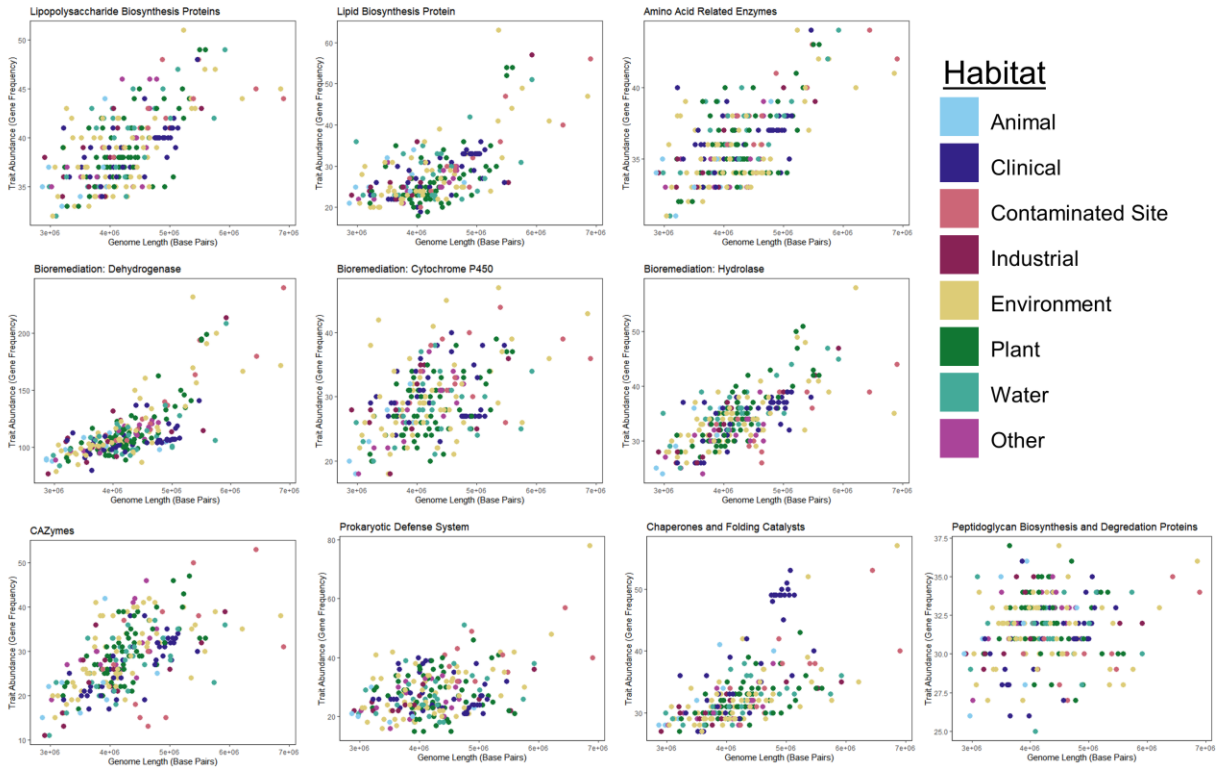
Figure 2.4. Heatmap depicting the enrichment of genome-based traits by habitat. Traits are grouped together based on their YAS classification: top green rows are growth traits, the middle blue CAZymes row is a resource acquisition trait, and the bottom, red rows are stress tolerance traits. Traits with stars indicate significant (Kruskal-Wallis, $p < 0.05$) differences of natural log transformed gene counts between habitats.



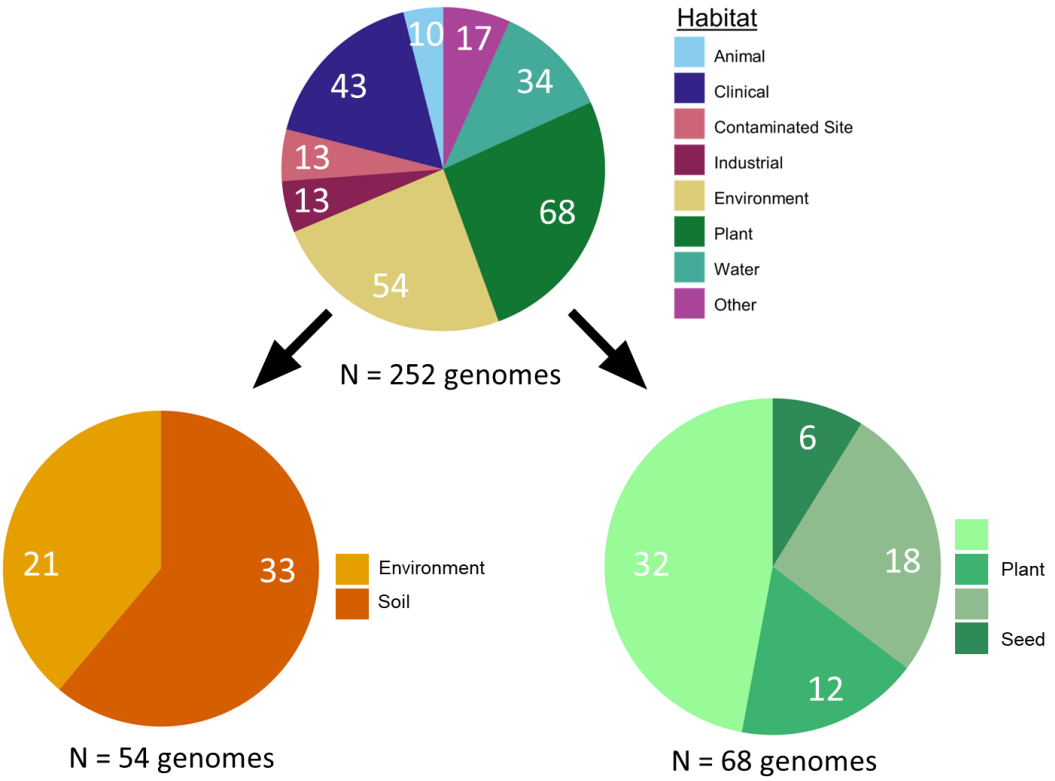
Supplemental Figure 2.1. Relative abundances of habitats within each clade.



Supplemental Figure 2.2. *Sphingomonas* phylogenetic tree from Figure 2.3 in main text with visible names. Genomes are labeled by the GenBank Accession Number and the associated *Sphingomonas* strain.



Supplemental Figure 2.3. Genome frequencies and traits. All traits had a significant ($p < 0.05$) correlation between the frequency of the associated genes and genome size.



Supplemental Figure 2.4. Demographic information of *Sphingomonas* genomes (N = 252) where the environment (N = 67) and plant (N = 60) classifications are further broken down into additional subcategories. Genomes labeled "Environment" were not further specified by the authors uploading the genomic sequences.

Table 2.1. Classification descriptions of the isolation sources for *Sphingomonas* samples.

Classification	Description
Animal	Isolation source is from a living, non-human, non-plant source
Clinical	Samples from a hospital that caters towards human beings, includes blood samples
Contaminated Site	Any sample that contains "contaminated" in the description
Environment	Isolated from abiotic, outdoor sources, such as soil
Industrial	Sources from bioreactors, mines, and wastewater facilities (contains "activated sludge")
Plant	Samples isolated from the plant kingdom, can come from seeds, leaves, and roots
Water	Any sample from a water source and sediment that does not contain "contaminated" in the description
Other	Samples that do not fit in to the other categories, can be lichens, dust

Table 2.2. Pangenome analysis for the 252 *Sphingomonas* genomes.

Gene	Description	Frequency
Core genes	99% <= strains <= 100%	444
Soft core genes	95% <= strains <= 99%	304
Shell genes	15% <= strains <= 95%	4,070
Cloud genes	0% <= strains <= 15%	108,998
Total	0% <= strains <= 100%	113,186

Supplemental Table 2.1. Pangenome analysis for the 252 *Sphingomonas* strains and the *Rhodospirillum centum* outgroup.

Gene	Description	Frequency
Core genes	99% <= strains <= 100%	404
Soft core genes	95% <= strains <= 99%	321
Shell genes	15% <= strains <= 95%	4,091
Cloud genes	0% <= strains <= 15%	111,058
Total	0% <= strains <= 100%	115,874

Supplemental Table 2.2. Pangenome analysis for the 23 complete genomes that we included in our data analysis.

Gene	Description	Frequency
Core genes	99% <= strains <= 100%	758
Soft core genes	95% <= strains <= 99%	184
Shell genes	15% <= strains <= 95%	4,452
Cloud genes	0% <= strains <= 15%	27,737
Total	0% <= strains <= 100%	33,131

Supplemental Table 2.3. Average genome length for strains in the habitat categories.

Habitat	Average Genome Size (bases)
Animal	3,678,369
Clinical	4,414,984
Contaminated Site	4,947,991
Environment	4,059,478
Industrial	4,188,510
Plant	4,251,944
Water	4,296,399
Other	4,092,009

CHAPTER 3

Sphingomonas clade and functional distribution with simulated climate change

ABSTRACT

Microbes are essential for the functioning of all ecosystems, and as global warming and anthropogenic pollution threaten ecosystems, it is critical to understand microbial responses to shifting environments. Here we investigated the climate response of *Sphingomonas*, a widespread Gram$^-$ bacterial genus, after an 18-month microbial community reciprocal transplant experiment across a Southern California climate gradient. Communities were taken from each ecosystem (desert, scrubland, grassland, pine-oak, and subalpine), inoculated onto sterilized grassland litter, and transplanted back into each of the sites. We aimed to determine whether *Sphingomonas* responds to climate change in line with previously published patterns from both Gram$^+$ bacteria and whole bacterial communities across the same climate gradient. We hypothesized that after 18 months, the transplanted *Sphingomonas* clade and functional composition would correspond with site conditions and reflect the *Sphingomonas* composition of native communities. Alternatively, since all the transplanted microbial communities were inoculated onto grassland litter, we hypothesized that the *Sphingomonas* clade and functional composition might converge on the litter type. Using publicly available *Sphingomonas* genomes, we built a phylogenetic tree and grouped taxa into 12 clades. Next, we identified *Sphingomonas* sequences in metagenomic data across the gradient and inferred their clade and functional composition. Representatives of all 12 clades were found at varying relative abundances along the climate gradient, and transplanted *Sphingomonas* clade composition shifted after 18 months. Furthermore, our results supported both hypotheses, indicating that both site and substrate determine *Sphingomonas* clade and functional composition. Climate likely drives shifts in

composition because site had the most significant effect (PERMANOVA; $p < 0.001$) on the distribution of both *Sphingomonas* functional ($R^2 = 0.446$) and clade composition ($R^2 = 0.359$). Additionally, for both *Sphingomonas* clade and functional composition, principal coordinate centroids of the 18-month transplanted samples were closer on average to the native *Sphingomonas* composition of the grassland site compared to the site they were transplanted into. Thus, the grassland substrate was also responsible for influencing *Sphingomonas* clade and trait convergence. These findings are in line with previous microbial research from the same climate gradient and together support consistent bacterial response to climate at different scales of genetic variation. This understanding of how microbes respond to perturbation at varying genetic scales may aid future predictions of microbial responses to climate change.

INTRODUCTION

Microorganisms are critical for the maintenance of healthy ecosystems and are threatened by the anthropogenic effects of climate change (Cavicchioli et al., 2019; Seidel et al., 2022). Furthermore, microbial communities drive planetary biogeochemical cycles - such as carbon and nitrogen flux- that organisms require for survival (Falkowski et al., 2008). Therefore, it is important to understand the implications of climate change for microbial composition and functioning. Still, due to high microbial abundance and diversity, it is difficult to predict how microbial communities will collectively respond to environmental shifts (Torsvik et al., 2002). Moreover, the ecological niche of a single microbial strain – or genetic variant of a species – can vary depending on the geographic origin of the strain or prior exposure to stress, which further complicates predictions (Kvitek et al., 2008; Wang et al., 2011).

When investigating microbial response to climate change, it is important to consider the scale of genetic variation. For example, the response to climate change varies across domains, as well as broad clades within bacteria and fungi (Castro et al., 2010; Danovaro et al., 2016; Zhang et al., 2017; Glassman et al., 2018). There are also finer-scale differences in compositional responses to the environment within a genus and between bacterial strains (Martiny et al., 2013; Chase et al., 2021). These finer-scale differences in compositional responses are reflected in conserved functional traits; therefore, genomics and ecologically relevant functional traits can help predict the ecosystem implications of climate change (Martiny et al., 2013; Evans and Wallenstein, 2014).

Along these lines, Glassman et al., (2018) reported that differences in microbial community composition and assembly impact litter decomposition under simulated climate change conditions. They reciprocally transplanted microbial communities across a Southern California climate gradient consisting of 5 sites that vary inversely with temperature and precipitation: desert, scrubland, grassland, pine-oak, and subalpine (Baker et al., 2018; Glassman et al., 2018). They found that the decomposition responses of microbial communities with respect to climate shifts depended on the community composition (Glassman et al., 2018). Furthermore, bacterial communities shifted more rapidly in response to changes in climate, and after 18 months, the bacterial community composition most closely reflected the abiotic conditions of the site (Glassman et al., 2018).

Differential responses to climate not only occur at broad genetic scales, (Cadotte et al., 2008; Gravel et al., 2011; Martiny et al., 2013), but also at finer scales. For instance, within a bacterial genus Chase et al., (2021) found that $Gram^{+}$ bacterial species of *Curtobacterium* were

also differentially adapted to local climates and experienced a shift in their composition after transplantation.

Building on the Chase et al., (2021) study of *Curtobacterium*, we aimed to test whether bacterial responses to climate change are similar across genera. In particular, we wanted to compare these responses across $Gram^+$ and $Gram^-$ bacteria. There are physiological differences between the cell membranes of $Gram^+$ and $Gram^-$ bacteria; namely $Gram^-$ bacteria have an outer membrane while $Gram^+$ bacteria do not (Slonczewski and Foster, 2017). These physiological differences may yield different functional and compositional responses to the selective pressures of climate change and in particular, to desiccation stress (Schimel et al., 2007a), which varies along our climate gradient. Since $Gram^+$ bacteria have a thicker layer of peptidoglycan within their cell walls, they may be more resistant to osmotic pressures and better tolerate stress than $Gram^-$ bacteria (Kieft et al., 1987; Schimel et al., 2007b; Evans and Wallenstein, 2014). The *Sphingomonas* bacterial genus is the most abundant $Gram^-$ bacteria found among our climate gradient sites (Chase et al., 2018). Therefore, we asked whether the *Sphingomonas* genus shows a similar response to changes in climate compared to $Gram^+$ bacteria (Chase et al., 2018, 2021) and bacterial communities as a whole (Glassman et al., 2018).

To address this question, we investigated the phylogenetic and functional diversity of *Sphingomonas* across the climate gradient. Using metagenomic data, we tested how the composition and functional potential of the *Sphingomonas* genus shifted in response to climate change as simulated by an 18-month reciprocal transplant experiment (Figure 3.1A). Based on previous results with *Curtobacterium* (Chase et al., 2021), we hypothesized that the clade and functional composition in the transplants would correspond with the site environment and climate (Figure 3.1B). Alternatively, because all the transplanted microbial communities were

inoculated onto a common grassland litter, the *Sphingomonas* clade and functional composition might converge on the litter type (Figure 3.1C). Since substrate is a potentially important control of community assembly (Zhalnina et al., 2018; Finks et al., 2021), the functional and clade composition of transplants might converge on that of grassland litter. Ultimately, our findings aimed to improve understanding of how consistently microbes respond to climate change at multiple phylogenetic levels.

METHODS

Reciprocal Transplant Design and Metagenomic Sampling

We analyzed metagenomic data from an 18-month reciprocal transplant across a Southern California climate gradient, as previously described in Glassman et al., (2018; Figure 3.1A). Briefly, the climate gradient consists of 5 sites across which temperature and precipitation vary inversely: desert, scrubland, grassland, pine-oak, and subalpine. Leaf litter was collected from each site on September 11, 2015. Subsequently, the leaf litter was homogenized with coffee grinders and used to inoculate irradiated grassland leaf litter in sterilized nylon bags. The nylon litter bags had 0.22 μm pores (cat. No. SPEC17970; Tisch Scientific, Cleves, OH, USA) such that nutrients and water can move freely into and out of the bags, but bacteria and fungi cannot.

On October 19, 2015, the transplant bags were placed in the 5 sites and destructively sampled after 6, 12, and 18 months. In total, 300 bags (5 sites x 5 inoculum x 4 replicates x 3 timepoints) were deployed, and 100 bags were collected at each timepoint. Additionally, at each timepoint, survey samples comprising native microbial communities on their native litter were

collected adjacent to each plot. Timepoint 0 (T0) refers to the time that all litter bags were deployed into the field.

Metagenomic data was generated for all survey samples (N = 20), the initial transplant communities (N = 20), and the 18 month transplant bags (N = 99) as previously described (Chase et al., 2021). Briefly, DNA was extracted from 0.05 g of ground leaf litter using the FastDNA SPIN Kit for Soil (Mo Bio). The DNA was subsequently cleaned with the Genomic DNA Clean and Concentrator kit (Zymo Research). Clean samples were diluted and used for the Nextera XT library Prep kit and sequenced with the Illumina HiSeq4000 instrument with paired-end reads of 150 bp.

We trimmed and quality filtered the metagenomic data using trimmomatic v0.36 (Bolger et al., 2014). We used bwa v0.7.17 and samtools v1.10 (Li and Durbin, 2009; Danecek et al., 2021) to filter out abundant grass (*Lolium perenne*; Accession: MEHO01000000) and fungus (*Pyrenophora teres*; Accession: NZ_AEEY00000000). For our analyses, we used the forward reads.

Sphingomonas identification from metagenomic samples

Using previously described methods, we selected 252 high-quality, publicly available *Sphingomonas* genomes to build a phylogenetic tree with 12 clades, rooted with a *Rhodospirillum* outgroup (Sorouri et al., *in review*). We identified 444 shared core genes within the *Sphingomonas* genomes, and of those, we selected 23 marker core genes (Supplemental Table 3.1) that appear in a reference genomic amino acid database developed by Chase et al., (2017). We appended the 23 core genes from each *Sphingomonas* genome to the Chase et al.,

(2017) reference genomic database, which together served as our reference database for DIAMOND BLASTX (Buchfink et al., 2014). We included the genes from the reference genomic database because we wanted to extract only *Sphingomonas* core gene hits and avoid matching metagenomic sequences that were not *Sphingomonas*. Forward reads from metagenomic samples were translated to amino acids and queried against the reference database with default BLASTX parameters (Buchfink et al., 2014). Furthermore, to ensure the sequences belonged to *Sphingomonas*, we selected hits with a percent identity value of at least 98% and an E value of less than $1e^{-20}$ as previously done (Scales et al., 2022).

When querying the metagenomic sequences, some hit only one clade, whereas others hit multiple clades. Those that hit only one clade were assigned to that clade. For the sequences that hit multiple clades, we first checked whether all hits to one of the clades showed consistently higher percent identity (e.g., 100%); if so, the sequence was assigned to that clade. If not, we assigned the sequence to a "pseudo clade" consisting of all clades that matched the query sequence with the same, highest identity. For example, if a query sequence matched 5 reference sequences from clade 1 with 100% identity, 4 reference sequences from clade 2 with 100% identity, and 3 reference sequences from clade 3 with 98% identity, then the query was assigned to pseudo clade 1-2. Because there are hundreds of potential pseudo clades, we grouped the sequences assigned to rarer pseudo clades into a single broad pseudo clade for subsequent analyses. Pseudo clades that had less then 10 hits were combined into the broad pseudo clade.

For each of the climate gradient samples, we summed the *Sphingomonas* sequences by clade. We rarefied the data with the EcolUtils v0.1 R package (Salazar, 2022), calculated the clade relative abundances for each sample, and visualized clade relative abundances using the ggpubr v0.4.0 R package (Kassambara, 2020). One Timepoint 3 (T3) subalpine survey sample

and several T3 bag samples were removed prior to rarefaction due to low sequence coverage. With the clade relative abundance data, we performed a principal coordinate analysis (PCoA) using a Bray-Curtis dissimilarity matrix to visualize compositional differences between samples. A permutational multivariate analysis of variance (PERMANOVA) is a distance-based method that tests whether microbial composition is associated with the covariates (Tang et al., 2016); in our case was used it to determine site and inoculum effect sizes. All statistical analyses and data visualizations across the climate gradient were done in R v4.1.0. (R Core Team, 2021). The PCoA and PERMANOVA analyses were performed with the vegan v2.5-7 R package, and the PCoA was again visualized with ggpubr v0.4.0 (Kassambara, 2020; Oksanen et al., 2020).

For the ordination, we averaged the points within sites to calculate the T3 transplant and survey centroids. Additionally, we used the average distances between centroids to determine which hypothesis was supported by results. To test the hypothesis that climate is responsible for convergence, we calculated the average distance between the T3 transplant centroids and their corresponding T3 survey centroid within the same site. To test the hypothesis that the grassland litter is responsible for convergence, we calculated the average distance between the T3 transplant centroids and the T3 grassland survey centroid.

Sphingomonas functional genes

We used KEGG and CAZy databases to identify *Sphingomonas* functional genes in publicly available genomes, selected a subset of traits, and assigned them to a Y-A-S life history category depending on their role in growth, resource acquisition, or stress tolerance (Sorouri et al., *in review*). For growth, these genome-based traits included amino acid related enzymes, lipid

biosynthesis proteins, and lipopolysaccharide biosynthesis proteins. CAZymes were used to indicate resource acquisition. Chaperones and folding catalysts, peptidoglycan biosynthesis and degradation proteins, and prokaryotic defense system proteins were the genome-based traits attributed to stress tolerance. We calculated the relative abundances of each of these trait categories for each of the 12 *Sphingomonas* clades designated by the publicly available genomes.

To infer the distribution of the 7 genome-based traits across the climate gradient, we multiplied the trait relative abundances from the publicly available genomes by the clade relative abundances for the transplant and survey samples at 18 months. For this product, we recalculated the clade relative abundances across the climate gradient for just the 12 clades and removed the pseudo clade categories. We averaged the values by sample and used them to construct a matrix with rows corresponding to samples and columns corresponding to traits. PCoA and PERMANOVA analyses were done on this matrix using the same methods described earlier for *Sphingomonas* clade relative abundances.

RESULTS

Sphingomonas clade composition

We aimed to evaluate the representation of *Sphingomonas* clades (Sorouri et al., *in review*; Figure 3.2) in metagenomic sequences across the climate gradient. Of 49,044 sequences identified as *Sphingomonas*, 34,003 matched to one of the 12 phylogenetic clades, and the remaining 15,041 were assigned to pseudo clades. Each of the 12 main clades was found across the climate gradient in both transplant and survey samples at varying abundance (Sorouri et al., *in review*; Figure 3.3). *Sphingomonas* composition was similar in the initial inoculum at T0 and

the survey samples at T3 (Figures 3.3A & 3.3B). However, the grassland site had a higher relative abundance of clade 1 in the T3 survey samples compared to T0. When comparing the survey and transplant samples at 18 months, there was a distinct difference in the distribution of *Sphingomonas* clades (Figures 3.3B & 3.3C).

Environmental conditions at the sites influenced the composition of *Sphingomonas* clades. Although both site and inoculum had significant ($p < 0.001$) effects on the distribution of transplanted *Sphingomonas*, site was the strongest predictor of composition with an R^2 value of 0.359, whereas inoculum had an R^2 value of 0.090 (Table 3.1). The strong site effect suggests that climate or other site conditions influence *Sphingomonas* composition following transplantion. There was also a significant ($p < 0.01$) site by inoculum interaction, which had a stronger effect on *Sphingomonas* composition than inoculum ($R^2 = 0.170$). This indicates that the inoculum effect varies by site (Figure 3.3C). The T3 transplanted samples at the grassland site had the tightest clusters in the PCoA, meaning that the *Sphingomonas* clade composition within the grassland was similar across inocula. For the other sites, there was more variation in *Sphingomonas* clade composition across inocula (Figure 3.4). In the desert site, the relative abundances of clades 1, 2, 5, and 9 varied most across inocula. Within the scrubland site, relative abundances of clades 1, 2, 5, 9, and 12 varied most, while clades 1, 5, and 12 varied most in the subalpine site (Figure 3.3C). Clade 7 was absent in most T3 transplant samples, and in the higher elevation pine-oak and subalpine sites, clades 1, 5, and 12 were the most abundant. Overall, clades 1 and 5 played an important role in driving the site by inoculum interaction. Clades 1 and 5 varied widely across the transplanted inocula within the desert, scrubland, and subalpine sites. However, clade 1 was rare within the grassland site and consistent within the pine-oak site.

If *Sphingomonas* composition is determined primarily by litter substrate, we would expect transplanted communities to converge on the grassland survey community (Figure 3.1C). The T3 survey samples not only clustered together by site, but also partially converged on the grassland survey community (Figure 3.4). The centroids of the *Sphingomonas* communities transplanted into the desert and scrubland were the closest to the grassland survey. The pine-oak, subalpine, and grassland transplants were further away from the grassland survey. Overall, the average distance (0.211) between the centroids of the transplant samples and the grassland survey was smaller than the average distance between the transplant centroids and survey samples within the same site (0.381). Therefore, there was also support for the hypothesis that *Sphingomonas* clades converged on the grassland litter.

Sphingomonas functional composition

Since *Sphingomonas* functional gene content reflects habitat preferences (Sorouri et al., *in review*), we investigated the distribution of several *Sphingomonas* genome-based functional traits to determine whether they supported our hypotheses (Figures 1B & 1C). We identified 3,615 unique genes from the KEGG database and 274 CAZymes. We analyzed a subset of the unique KEGG genes to investigate genes involved in growth (N = 61) and stress tolerance (N = 154). Based on our analysis of publicly-available genomes, *Sphingomonas* clades had significant ($p < 0.05$, Kruskal Wallis) differences in their functional gene content (Figure 3.5A; Sorouri et al., *in review*). For the genome-based traits associated with high growth yield, the amino acid related enzymes and lipopolysaccharide biosynthesis proteins were the most abundant in clade 8, whereas lipid biosynthesis proteins were most abundant in clade 12. CAZymes linked to the resource acquisition strategy were more abundant in clade 1 relative to other clades. With respect

to the genome-based stress tolerance traits, chaperones and folding catalysts were most abundant in clade 7, clade 12 had the highest abundance of prokaryotic defense system genes, and peptidoglycan and biosynthesis proteins were most abundant in clade 2 (Figure 3.5A).

Using the functional trait distributions of each clade, we next predicted the distribution of *Sphingomonas* functional genes across the climate gradient. Similar to the clade composition patterns, site and inoculum had significant ($p < 0.05$) effects on *Sphingomonas* functional composition in transplants by T3; however, the site * inoculum interaction was not significant (Table 3.2; Figure 3.5B). As with the clade composition, site had the strongest effects on the *Sphingomonas* functional gene distribution ($R^2 = 0.446$, $p < 0.001$), while inoculum had a weaker effect ($R^2 = 0.058$, $p < 0.05$). The patterns of convergence on the grassland survey for functional traits were nearly identical to the patterns for clade composition in that the average distance between T3 bag centroids and the T3 grassland survey centroid (0.00782) was smaller than the average distance between the T3 bag centroids and their respective T3 survey samples (0.0125) from the same site.

DISCUSSION

Our results suggest that microbial responses to environmental change are consistent across bacterial clades at the genus level. The *Sphingomonas* clade composition varies across the climate gradient, indicating that clades may be differentially adapted to site conditions, including climate. Furthermore, *Sphingomonas* clade composition shifted within the genus during an 18-month reciprocal transplant experiment (Figure 3.3). These findings thus support the hypothesis that climate and abiotic conditions of the sites drive shifts in *Sphingomonas* clade and trait

composition (Figures 3.1B and 3.4; Table 3.1). Moreover, there was also support for our alternative hypothesis that the grassland substrate was responsible for clade and trait convergence (Figures 3.1C; Fig 3.4, 3.5B), suggesting indirect effects of climate change via plant community shifts may also influence *Sphinogomonas* response to climate. Therefore, *Sphinogomonas* clade and functional composition is determined by both site and litter.

The outcome that site and abiotic climate conditions drive clade and functional compositional shifts is line with previous findings across this Southern California climate gradient. Both Glassman et al., (2018) and Chase et al., (2021) found that climate drives distinct bacterial composition at both community and strain levels, respectively. Our study shows that these results extend to *Sphingomonas*, an abundant Gram- bacterial clade (Figure 3.4 and Table 3.1). This phenomenon that environmental variables are strong predictors for microbial processes and composition is not only observed across the Southern California climate gradient, but also globally (Graham et al., 2016). Additionally, the patterns in the distribution of *Sphingomonas* clades and functional genes across the climate gradient to some extent reflect the climate conditions of the sites (Figures 3.4 & 3.5). Within both survey and transplant samples at T3, oftentimes the desert and scrubland sites grouped together, as did the pine-oak and subalpine sites. These groupings may reflect similarities with respect to temperature, precipitation, and litter chemistry (Baker et al., 2018).

Although site conditions are a strong predictor of shifts in *Sphingomonas* clade composition (Figure 3.4; Table 3.1), there was still variation among the transplanted inocula within each site after 18 months even though the grassland litter they resided on was uniform (Figure 3.3). It is likely that the different inocula had varying levels of resilience to environmental changes, and microbial legacy effects from previous historical events may have

prevented complete convergence (Allison and Martiny, 2008; Martiny et al., 2017; Glassman et al., 2018; Hawkes et al., 2020; Chase et al., 2021). Additionally, *Sphingomonas* clades 1, 5, and 6 contain taxa that are found at high relative abundances in both plant and environmental habitats (Sorouri et al., *in review*); these clades shifted the most in the T3 transplant after 18 months (Figure 3.3). Perhaps these clades are more abundant in the T3 transplant because they are better suited to survive on grassland litter across the climate gradient. Legacy effects also impact traits and potential functioning of microbial communities (Glassman et al., 2018). Even though some climate response traits are deeply conserved (e.g. Martiny et al., 2015; Amend et al., 2016; Isobe et al., 2020), climate response also occurs at fine scales, such as our results suggest for *Sphingomonas* and prior research reflects for *Prochlorococcus* (Martiny et al., 2013). Therefore, responses to climate change are likely to occur across different scales of genetic variation.

We were surprised to find that in both clade and functional composition, the survey and transplant samples within the grassland site did not converge and in fact were very dissimilar (Figures 3.3-5). In both *Sphingomonas* functional and clade composition, after 18 months the grassland survey was more similar to the scrubland and desert transplant samples than to the grassland transplant samples. Given that the initial *Sphingomonas* clade composition of the grassland inoculum was similar to the grassland survey and they were both on grassland litter, for both hypotheses we anticipated that the grassland survey and transplant samples would converge after 18 months (Figure 3.1). However, it is possible that there were microclimate effects within bags that prevented the convergence. Additionally, the grassland substrate may vary from year to year, and microbial succession may have further contributed to the variation between the two groups (van der Valk, 1981). The combination of both litter substrate and bag microclimate may explain distinct *Sphingomonas* composition in grassland survey and transplant

samples, given that the response of bacterial microdiversity to environmental perturbations is substrate and ecosystem dependent (Scales et al., 2022).

Trait-based approaches are useful for predicting and interpreting microbial responses to climate change. Trait responses to climate change simulated by transplantation closely followed clade responses, consistent with the trait variation across clades (Figure 3.5A). For example, clade 1 had the highest abundance of CAZymes, and was responsible for driving the differences in *Sphingomonas* clade and functional composition. We also observed that clade 7 was rare in our litter metagenomes, consistent with this clade's dominance by clinical strains that may not possess the traits to live in a surface soil environment (Sorouri et al., *in review*). Still, across the climate gradient, differences in the relative abundance of other environmentally prevalent clades such as 5 and 6 were not clearly related to the genome-based traits we measured. We recognize that there are likely finer-scale differences between traits, such as individual glycoside hydrolases (Berlemont and Martiny, 2013), that we did not assess here. Additionally, there may be functional differences in other ecologically relevant traits that we did not analyze.

CONCLUSION

We investigated the distribution of *Sphingomonas* clades and functional potential across a Southern California climate gradient. We found that the clade and functional composition shifted during an 18-month reciprocal transplant. Our findings indicate that the Gram$^-$ *Sphingomonas* genus had compositional and functional responses similar to Gram$^+$ *Curtobacterium* (Chase et al., 2021) and microbial communities (Glassman et al., 2018). Collectively, these studies suggest that compositional and functional responses to climate change occur at various genetic scales,

ranging from within *Curtobacterium* strains, to within *Sphingomonas* clades, and across clades within microbial communities. Understanding how microbes respond to perturbation at all these genetic scales may aid future predictions of microbial responses to climate change.

ACKNOWLEDGEMENTS

We would like to thank the following people for discussion and feedback on the manuscript: Adam C Martiny, Alberto Barron, Alex B Chase, Brittni L Bertolet, Claudia Weihe, Edwin Solares, Elsa Abs, Jennifer BH Martiny, José M Murúa Royo, Lucas Ustick, Luciana Chavez Rodriguez, Moein Hosseini, and Tiffany N Batarseh. Thanks to Nadya Williams and the HPC Team for technical support and maintenance of the high-performance computing clusters. This project was funded by the US Department of Energy, Office of Science, Biological and Environmental Research, under award DE-SC0020382.

FIGURES AND TABLES

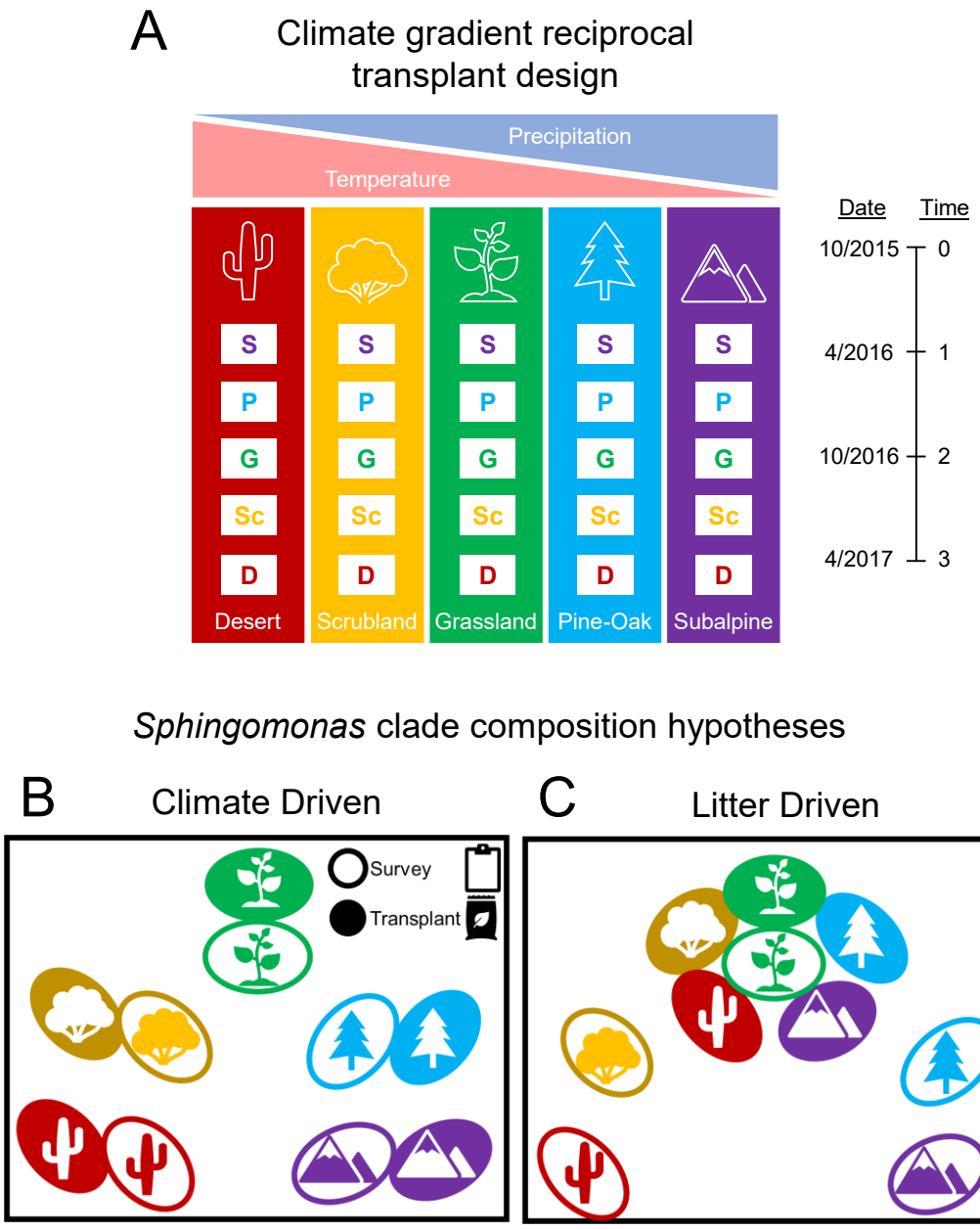


Figure 3.1. Microbial community reciprocal transplant design and hypotheses driving *Sphingomonas* clade composition after 18 months in the field. A) Schematic of reciprocal transplant experiment across a climate elevation gradient across which temperature and precipitation very inversely. All colors and icons remain consistent across the figures. We hypothesized that B) site environment (e.g., climate) determines *Sphingomonas* composition after 18 months. Thus, the composition within bag transplant and survey samples will be similar at each site. C) Alternatively, since all microbial communities within transplants were inoculated onto grassland litter, the grassland substrate might drive *Sphingomonas* composition, causing transplants to converge on the grassland survey samples.

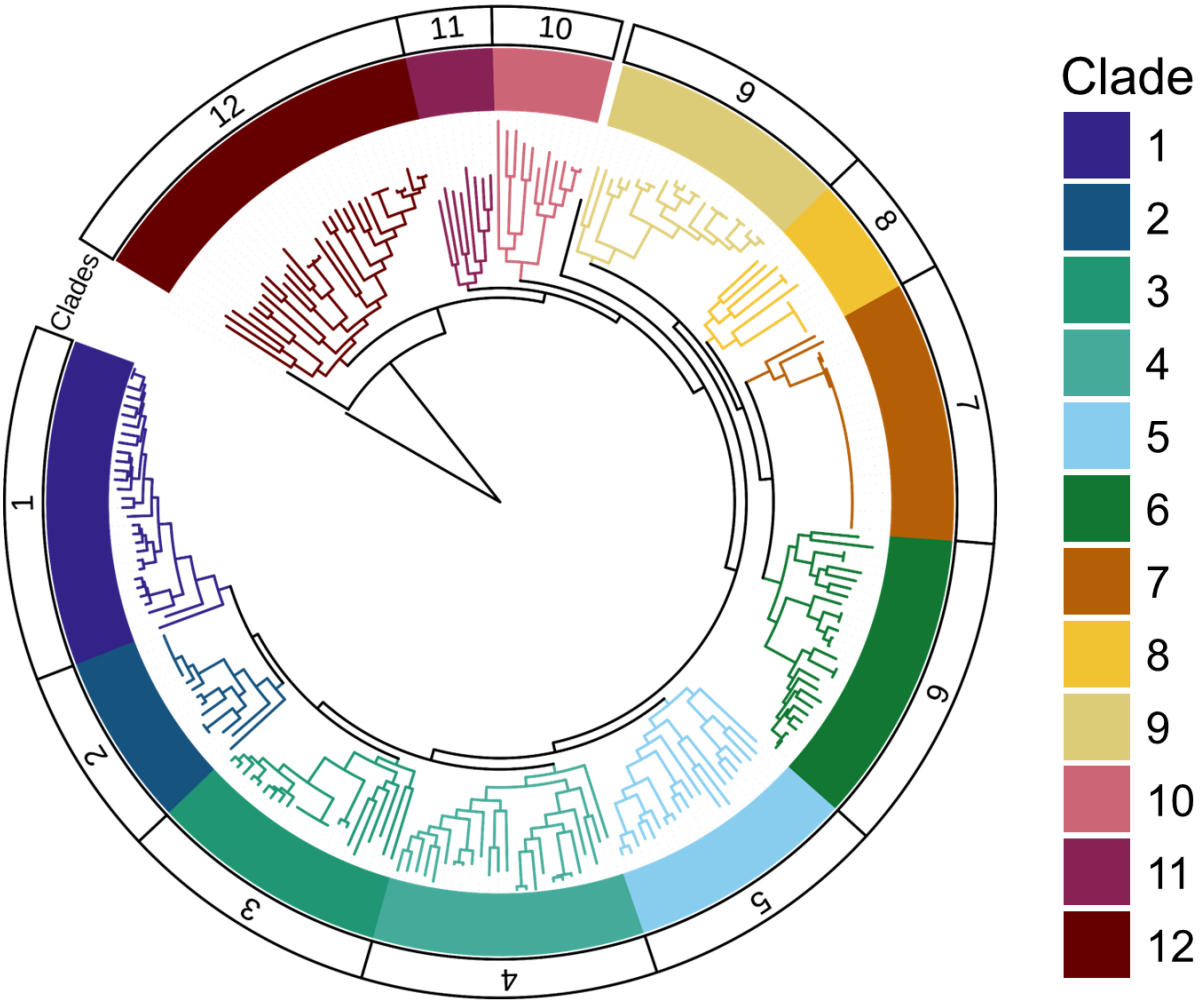


Figure 3.2. Phylogenetic tree of 252 publicly available, high-quality *Sphingomonas* genomes. The tree was built with 404 core genes and separated into 12 clades based on their divergence from a common ancestor. Clades are color coordinated, and the tree is rooted by a *Rhodospirillum centum SW* (Accession: CP000613) outgroup.

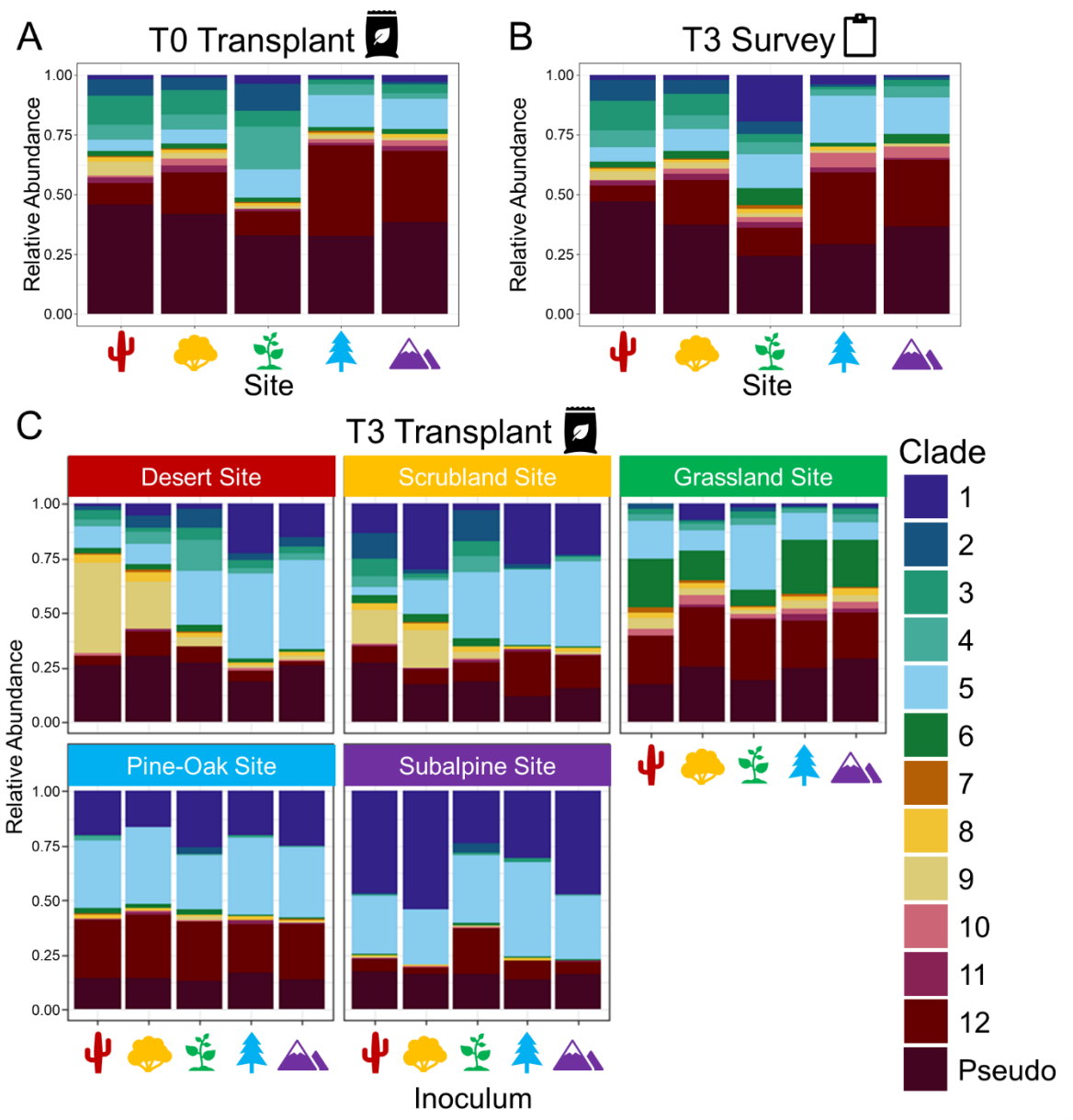


Figure 3.3. *Sphingomonas* clade distribution across the climate gradient in A) T0 transplant samples, B) T3 survey samples, and C) T3 transplant samples. For A) and B), the clade relative abundances are represented for each site. For the T3 transplant, the facet boxes are labeled and colored by the site, and the x-axis indicates inoculum within each site. Facet and icon colors reflect the site, while relative abundances are colored by clade.

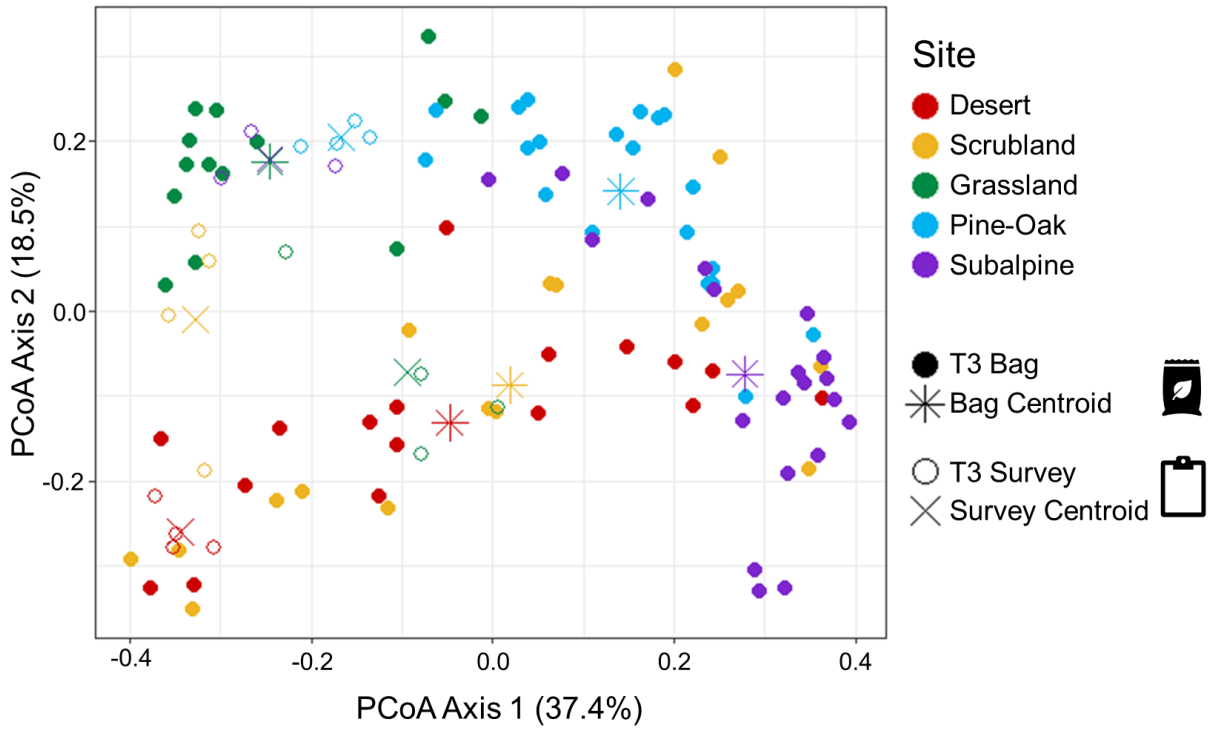


Figure 3.4. Ordination of *Sphingomonas* within survey and transplant samples after 18 months based on clade relative abundances. Points are colored by site, where solid points reflect the transplant samples and open points represent survey samples. Asterisks represent the centroids of transplant samples and x's represent the centroids of survey samples. PERMANOVA statistics revealed significant ($p < 0.001$) site, inoculum, and site* inoculum interaction effects.

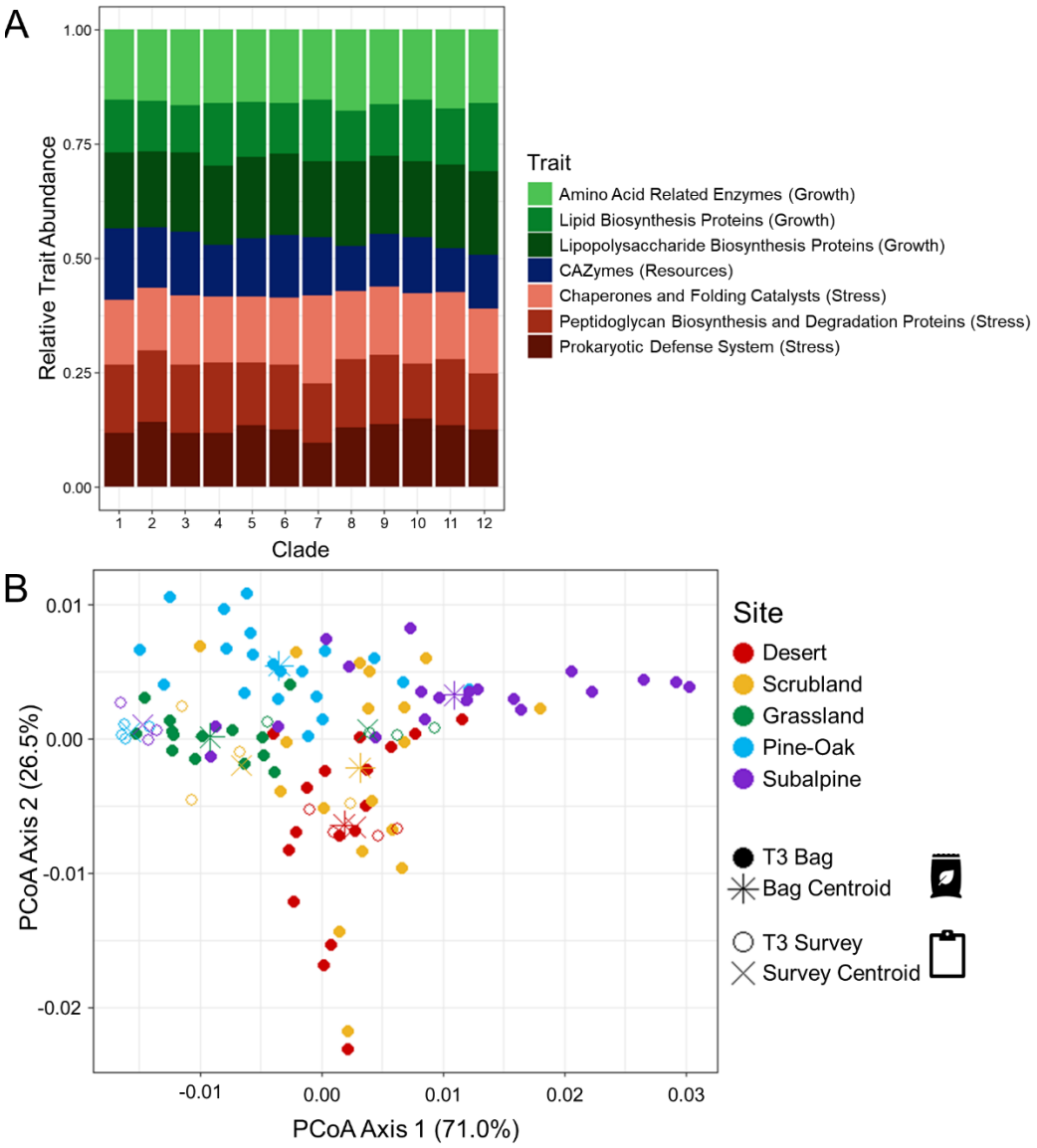


Figure 3.5. Distribution of *Sphingomonas* functional genes within clades and across the climate gradient after 18 months. A) Trait relative abundance of YAS functional genes within clades from publicly available *Sphingomonas* genomes. B) Principal coordinate analysis of predicted trait composition in T3 survey and transplant samples across the climate gradient colored by site and calculated with Bray-Curtis dissimilarity distances. Ordination does not include trait information from pseudo clades.

Table 3.1. PERMANOVA statistics comparing the clade composition of *Sphingomonas* within Timepoint 3 transplanted samples.

	Degrees of Freedom	Sum of Squares	R^2	F	p-value
Site	4	4.859	0.359	14.156	< 0.001
Inoculum	4	1.215	0.090	3.549	< 0.001
Site : Inoculum	16	2.305	0.170	1.679	< 0.005

Table 3.2. PERMANOVA statistics comparing the predicted functional composition of *Sphingomonas* within Timepoint 3 transplants.

	Degrees of freedom	Sum of Squares	R^2	F	p-value
Site	4	$5.47 * 10^{-3}$	0.446	19.126	<0.001
Inoculum	4	$7.13 * 10^{-3}$	0.058	2.492	0.017
Site : Inoculum	16	$1.50 * 10^{-3}$	0.123	1.315	0.164

Supplemental Table 3.1. List of 23 core genes selected from *Sphingomonas* genomes and appended to Chase et al., (2017) reference database

Protein	Name
L1	rplA
L2	rplB
L3	rplC
L4	rplD
L5	rplE
L6	rplF
L10	rplJ
L11	rplK
L13	rplM
L14	rplN
L15	rplO
L16	rplP
L18	rplR
L24	rplX
S2	rpsB
S3	rpsC
S8	rpsH
S9	rpsI
S11	rpsK
S12	rpsL
S13	rpsM
S17	rpsQ
S19	rpsS

REFERENCES

Ackerly, D. D. (2003). Community assembly, niche conservatism, and adaptive evolution in changing environments. *Int. J. Plant Sci.* 164, S165–S184. doi: 10.1086/368401.

Albright, M. B. N., Chase, A. B., and Martiny, J. B. H. (2019). Experimental Evidence that Stochasticity Contributes to Bacterial Composition and Functioning in a Decomposer Community. *MBio* 10. doi: 10.1128/mBio.00568-19.

Allison, S. D. (2005). Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecol. Lett.* 8, 626–635. doi: 10.1111/j.1461-0248.2005.00756.x.

Allison, S. D. (2006). Brown Ground: A Soil Carbon Analogue for the Green World Hypothesis? *Am. Nat.* 167, 619–627. doi: 10.1086/503443.

Allison, S. D., and Martiny, J. B. H. (2008). Resistance, resilience, and redundancy in microbial communities. *Proc. Natl. Acad. Sci.* 105, 11512–11519. doi: 10.1073/pnas.0801925105.

Allison, S. D., Weintraub, M. N., Gartner, T. B., and Waldrop, M. P. (2010). "Evolutionary-Economic Principles as Regulators of Soil Enzyme Production and Ecosystem Function," in (Springer, Berlin, Heidelberg), 229–243. doi: 10.1007/978-3-642-14225-3_12.

Alster, C. J., von Fischer, J. C., Allison, S. D., and Treseder, K. K. (2020). Embracing a new paradigm for temperature sensitivity of soil microbes. *Glob. Chang. Biol.* 26, 3221–3229. doi: 10.1111/gcb.15053.

Amend, A. S., Martiny, A. C., Allison, S. D., Berlemont, R., Goulden, M. L., Lu, Y., et al. (2016). Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. *ISME J.* 10, 109–118. doi: 10.1038/ismej.2015.96.

Asaf, S., Khan, A. L., Khan, M. A., Imran, Q. M., Yun, B. W., and Lee, I. J. (2017). Osmoprotective functions conferred to soybean plants via inoculation with Sphingomonas sp. LK11 and exogenous trehalose. *Microbiol. Res.* 205, 135–145. doi: 10.1016/j.micres.2017.08.009.

Asaf, S., Numan, M., Khan, A. L., and Al-Harrasi, A. (2020). Sphingomonas : from diversity and genomics to functional role in environmental remediation and plant growth. *Crit. Rev. Biotechnol.* 40, 138–152. doi: 10.1080/07388551.2019.1709793.

Ayres, E., Steltzer, H., Simmons, B. L., Simpson, R. T., Steinweg, J. M., Wallenstein, M. D., et al. (2009). Home-field advantage accelerates leaf litter decomposition in forests. *Soil Biol. Biochem.* 41, 606–610. doi: 10.1016/j.soilbio.2008.12.022.

Baas-Becking, L. (1934). Geobiologie of Inleiding Tot de Milieukunde.

Baker, N. R., and Allison, S. D. (2017). Extracellular enzyme kinetics and thermodynamics along a climate gradient in southern California. *Soil Biol. Biochem.* 114, 82–92. doi: 10.1016/J.SOILBIO.2017.07.005.

Baker, N. R., Khalili, B., Martiny, J. B. H., and Allison, S. D. (2018). Microbial decomposers not constrained by climate history along a Mediterranean climate gradient in southern California. *Ecology* 99, 1441–1452. doi: 10.1002/ecy.2345.

Balkwill, D. L., Fredrickson, J. K., and Romine, M. F. (2006). *Sphingomonas and Related Genera*. New York City, USA: Springer New York.

Barnosky, A. (2011). Has the Earth's sixth mass extinction already arrived? *Nature* 471, 51–57.

Batarseh, T. N., Morales-Cruz, A., Ingel, B., Roper, M. C., and Gaut, B. S. (2022). Using

Genomes and Evolutionary Analyses to Screen for Host-Specificity and Positive Selection in the Plant Pathogen Xylella fastidiosa. *Appl. Environ. Microbiol.* 88, e0122022. doi: 10.1128/aem.01220-22.

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., and Delignette-Muller, M.-L. (2015). A Toolbox for Nonlinear Regression in {R}: The Package {nlstools}. *J. Stat. Softw.* 66, 1–21. Available at: http://www.jstatsoft.org/v66/i05/.

Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17, 478–486. doi: 10.1016/j.tplants.2012.04.001.

Berlemont, R., and Martiny, A. C. (2013). Phylogenetic distribution of potential cellulases in bacteria. *Appl. Environ. Microbiol.* 79, 1545–54. doi: 10.1128/AEM.03305-12.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bratbak, G. (1985). Bacterial biovolume and biomass estimations. *Appl. Environ. Microbiol.* 49, 1488–1493. doi: 10.1128/aem.49.6.1488-1493.1985.

Brown, J. H., and others (1998). *Biogeography*.

Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176.

Buonaurio, R., Stravato, V. M., Kosako, Y., Fujiwara, N., Naka, T., Kobayashi, K., et al. (2002). Sphingomonas melonis sp. nov., a novel pathogen that causes brown spots on yellow Spanish melon fruits. *Int. J. Syst. Evol. Microbiol.* 52, 2081–2087. doi: 10.1099/00207713-

52-6-2081.

Burns, R. G., DeForest, J. L., Marxsen, J., Sinsabaugh, R. L., Stromberger, M. E., Wallenstein, M. D., et al. (2013). Soil enzymes in a changing environment: Current knowledge and future directions. *Soil Biol. Biochem.* 58, 216–234. doi: 10.1016/J.SOILBIO.2012.11.009.

Cadotte, M. W., Cardinale, B. J., and Oakley, T. H. (2008). Evolutionary history and the effect of biodiversity on plant productivity. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17012–17017. doi: 10.1073/pnas.0805962105.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663.

Castro, H. F., Classen, A. T., Austin, E. E., Norby, R. J., and Schadt, C. W. (2010). Soil microbial community responses to multiple experimental climate change drivers. *Appl. Environ. Microbiol.* 76, 999–1007. doi: 10.1128/AEM.02874-09.

Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., et al. (2019). Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* 17, 569–586. doi: 10.1038/s41579-019-0222-5.

Chase, A. B., Gomez-Lunar, Z., Lopez, A. E., Li, J., Allison, S. D., Martiny, A. C., et al. (2018). Emergence of soil bacterial ecotypes along a climate gradient. *Environ. Microbiol.* 20, 4112–4126. doi: 10.1111/1462-2920.14405.

Chase, A. B., Karaoz, U., Brodie, E. L., Gomez-Lunar, Z., Martiny, A. C., and Martiny, J. B. H. (2017). Microdiversity of an Abundant Terrestrial Bacterium Encompasses Extensive

Variation in Ecologically Relevant Traits. *MBio* 8, e01809-17. doi: 10.1128/mBio.01809-17.

Chase, A. B., Weihe, C., and Martiny, J. B. H. (2021). Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. *Proc. Natl. Acad. Sci.* 118, e2101254118. doi: 10.1073/pnas.2101254118.

Cockell, C. S., Santomartino, R., Finster, K., Waajen, A. C., Eades, L. J., Moeller, R., et al. (2020). Space station biomining experiment demonstrates rare earth element extraction in microgravity and Mars gravity. *Nat. Commun.* 11, 5523. doi: 10.1038/s41467-020-19276-w.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi: 10.1093/gigascience/giab008.

Danovaro, R., Molari, M., Corinaldesi, C., and Dell'Anno, A. (2016). Macroecological drivers of archaea and bacteria in benthic deep-sea ecosystems. *Sci. Adv.* 2. doi: 10.1126/sciadv.1500961.

Dolan, K. L., Peña, J., Allison, S. D., and Martiny, J. B. H. (2017). Phylogenetic conservation of substrate use specialization in leaf litter bacteria. *PLoS One* 12, e0174472. doi: 10.1371/journal.pone.0174472.

Evans, S. E., and Wallenstein, M. D. (2014). Climate change alters ecological strategies of soil bacteria. *Ecol. Lett.* 17, 155–164. doi: 10.1111/ele.12206.

Falkowski, P. G., Fenchel, T., and Delong, E. F. (2008). The microbial engines that drive earth's biogeochemical cycles. *Science (80-.).* 320, 1034–1039. doi: 10.1126/science.1153213.

Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* 103, 626–631. doi: 10.1073/pnas.0507535103.

Finks, S. S., Weihe, C., Kimball, S., Allison, S. D., Martiny, A. C., Treseder, K. K., et al. (2021). Microbial community response to a decade of simulated global changes depends on the plant community. *Elem. Sci. Anthr.* 9. doi: 10.1525/elementa.2021.00124.

Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science (80-.).* 296, 1061–1063. doi: 10.1126/science.1070710.

Fox, J., and Weisberg, S. (2019). *An {R} Companion to Applied Regression*. Third. Thousand Oaks {CA}: Sage Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Friedman, J., Higgins, L. M., and Gore, J. (2017). Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.* 1. doi: 10.1038/s41559-017-0109.

German, D. P., Weintraub, M. N., Grandy, A. S., Lauber, C. L., Rinkes, Z. L., and Allison, S. D. (2011). Optimization of hydrolytic and oxidative enzyme methods for ecosystem studies. *Soil Biol. Biochem.* 43, 1387–1397. doi: 10.1016/J.SOILBIO.2011.03.017.

Gholz, H. L., Wedin, D. A., Smitherman, S. M., Harmon, M. E., and Parton, W. J. (2000). Long-term dynamics of pine and hardwood litter in contrasting environments: toward a global model of decomposition. *Glob. Chang. Biol.* 6, 751–765. doi: 10.1046/j.1365-2486.2000.00349.x.

Gießelmann, U. C., Martins, K. G., Brändle, M., Schädler, M., Marques, R., and Brandl, R. (2011). Lack of home-field advantage in the decomposition of leaf litter in the Atlantic Rainforest of Brazil. *Appl. Soil Ecol.* 49, 5–10. doi: 10.1016/j.apsoil.2011.07.010.

Glassman, S. I., Weihe, C., Li, J., Albright, M. B. N., Looby, C. I., Martiny, A. C., et al. (2018). Decomposition responses to climate depend on microbial community composition. *Proc. Natl. Acad. Sci. U. S. A.* 115, 11994–11999. doi: 10.1073/pnas.1811269115.

Graham, E. B., Knelman, J. E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., et al. (2016). Microbes as engines of ecosystem function: When does community structure enhance predictions of ecosystem processes? *Front. Microbiol.* 7, 214. doi: 10.3389/fmicb.2016.00214.

Gravel, D., Bell, T., Barbera, C., Bouvier, T., Pommier, T., Venail, P., et al. (2011). Experimental niche evolution alters the strength of the diversity-productivity relationship. *Nature* 469, 89–94. doi: 10.1038/nature09592.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086.

Halo, B. A., Khan, A. L., Waqas, M., Al-Harrasi, A., Hussain, J., Ali, L., et al. (2015). Endophytic bacteria (Sphingomonas sp. LK11) and gibberellin can improve solanum lycopersicum growth and oxidative stress under salinity. *J. Plant Interact.* 10, 117–125. doi: 10.1080/17429145.2015.1033659.

Hawkes, C. V., Shinada, M., and Kivlin, S. N. (2020). Historical climate legacies on soil respiration persist despite extreme changes in rainfall. *Soil Biol. Biochem.* 143. doi: 10.1016/j.soilbio.2020.107752.

Heather, J. M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107, 1–8. doi: 10.1016/j.ygeno.2015.11.003.

Henderson, B., Allan, E., and Coates, A. R. M. (2006). Stress wars: The direct role of host and bacterial molecular chaperones in bacterial infection. *Infect. Immun.* 74, 3693–3706. doi: 10.1128/IAI.01882-05.

Henderson, B., and Pockley, A. G. (2010). Molecular chaperones and protein-folding catalysts as intercellular signaling regulators in immunity and inflammation. *J. Leukoc. Biol.* 88, 445–462. doi: 10.1189/jlb.1209779.

Hobbie, S. E. (1992). Effects of plant species on nutrient cycling. *Trends Ecol. Evol.* 7, 336–339. doi: 10.1016/0169-5347(92)90126-V.

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical J.* 50, 346–363.

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11, 119. doi: 10.1186/1471-2105-11-119.

IPCC (2022). AR6 Climate Change 2022: Mitigation of Climate Change — IPCC. Available at: https://www.ipcc.ch/report/sixth-assessment-report-working-group-3/ [Accessed February 15, 2023].

Isobe, K., Allison, S. D., Khalili, B., Martiny, A. C., and Martiny, J. B. H. (2019). Phylogenetic conservation of bacterial responses to soil nitrogen addition across continents. *Nat. Commun.* 10, 2499. doi: 10.1038/s41467-019-10390-y.

Isobe, K., Bouskill, N. J., Brodie, E. L., Sudderth, E. A., and Martiny, J. B. H. (2020). Phylogenetic conservation of soil bacterial responses to simulated global changes. *Philos.*

Trans. R. Soc. B Biol. Sci. 375, 20190242. doi: 10.1098/rstb.2019.0242.

Jogler, M., Chen, H., Simon, J., Rohde, M., Busse, H. J., Klenk, H. P., et al. (2013). Description of Sphingorhabdus planktonica gen. nov., sp. nov. and reclassification of three related members of the genus Sphingopyxis in the genus Sphingorhabdus gen. nov. *Int. J. Syst. Evol. Microbiol.* 63, 1342–1349. doi: 10.1099/ijs.0.043133-0.

Johnson, C. (2017). Biodiversity losses and conservation responses in the Anthropocene. *Science (80-.).* 356, 270–275.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27.

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006.

Kassambara, A. (2020). ggpubr: "ggplot2" Based Publication Ready Plots. Available at: https://cran.r-project.org/package=ggpubr.

Kieft, T. L., Soroker, E., and Firestone, M. K. (1987). Microbial biomass response to a rapid increase in water potential when dry soil is wetted. *Soil Biol. Biochem.* 19, 119–126. doi: 10.1016/0038-0717(87)90070-8.

Kim, S.-J., Moon, J.-Y., Lim, J.-M., Ahn, J.-H., Weon, H.-Y., Ahn, T.-Y., et al. (2014). Sphingomonas aerophila sp. nov. and Sphingomonas naasensis sp. nov., isolated from air and soil, respectively. doi: 10.1099/ijs.0.055269-0.

Koskinen, R., Ali-Vehmas, T., Kämpfer, P., Laurikkala, M., Tsitko, I., Kostyal, E., et al. (2000).

Characterization of Sphingomonas isolates from Finnish and Swedish drinking water distribution systems. *J. Appl. Microbiol.* 89, 687–696. doi: https://doi.org/10.1046/j.1365-2672.2000.01167.x.

Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., and Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* 29, 592–599. doi: 10.1111/1365-2435.12345.

Kvitek, D. J., Will, J. L., and Gasch, A. P. (2008). Variations in Stress Sensitivity and Genomic Expression in Diverse S. cerevisiae Isolates. *PLoS Genet.* 4, e1000223. doi: 10.1371/journal.pgen.1000223.

Lee, J. S., Shin, Y. K., Yoon, J. H., Takeuchi, M., Pyun, Y. R., and Park, Y. H. (2001). Sphingomonas aquatilis sp. nov., Sphingomonas koreensis sp. nov. and Sphingomonas taejonensis sp. nov., yellow-pigmented bacteria isolated from natural mineral water. *Int. J. Syst. Evol. Microbiol.* 51, 1491–1498. doi: 10.1099/00207713-51-4-1491.

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239.

Leung, K. T., Chang, Y. J., Gan, Y. D., Peacock, A., Macnaughton, S. J., Stephen, J. R., et al. (1999). Detection of Sphingomonas spp in soil by PCR and sphingolipid biomarker analysis. *J. Ind. Microbiol. Biotechnol.* 23, 252–260. doi: 10.1038/sj.jim.2900677.

Leys, N. M. E. J., Ryngaert, A., Bastiaens, L., Verstraete, W., Top, E. M., and Springael, D. (2004). Occurrence and phylogenetic diversity of Sphingomonas strains in soils contaminated with polycyclic aromatic hydrocarbons. *Appl. Environ. Microbiol.* 70, 1944–55. doi: 10.1128/aem.70.4.1944-1955.2004.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324.

Liu, F., Zhan, R. L., and He, Z. Q. (2018). First report of bacterial dry rot of mango caused by Sphingomonas sanguinis in China. *Plant Dis.* 102, 2632. doi: 10.1094/PDIS-04-18-0589-PDN.

Malik, A. A., Martiny, J. B. H., Brodie, E. L., Martiny, A. C., Treseder, K. K., and Allison, S. D. (2020). Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J.* 14, 1–9. doi: 10.1038/s41396-019-0510-0.

Marbjerg, L. H., Gaini, S., and Justesen, U. S. (2015). First report of Sphingomonas koreensis as a human pathogen in a patient with meningitis. *J. Clin. Microbiol.* 53, 1028–1030. doi: 10.1128/JCM.03069-14.

Martiny, A. C., Treseder, K., and Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* 7, 830–838. doi: 10.1038/ismej.2012.160.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341.

Martiny, J. B. H., Jones, S. E., Lennon, J. T., and Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science (80-.).* 350. doi: 10.1126/science.aac9323.

Martiny, J. B. H., Martiny, A. C., Weihe, C., Lu, Y., Berlemont, R., Brodie, E. L., et al. (2017). Microbial legacies alter decomposition in response to simulated global change. *ISME J.* 11, 490–499. doi: 10.1038/ismej.2016.122.

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10, 780. doi: 10.3389/fmicb.2019.00780.

Mundry, R. (2014). "Statistical Issues and Assumptions of Phylogenetic Generalized Least Squares," in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice*, ed. L. Z. Garamszegi (Berlin, Heidelberg: Springer Berlin Heidelberg), 131–153. doi: 10.1007/978-3-662-43550-2_6.

Noble, I. R., and Slatyer, R. O. (1977). Post-fire succession of plants in Mediterranean ecosystems [Eucalyptus]. *USDA For. Serv. Gen. Tech. Rep. WO*.

Ochman, H., Lawrence, J. G., and Grolsman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi: 10.1038/35012500.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). vegan: Community Ecology Package. Available at: https://cran.r-project.org/package=vegan.

Onder Erguven, G., and Demirci, U. (2019). Statistical evaluation of the bioremediation performance of Ochrobactrum thiophenivorans and Sphingomonas melonis bacteria on Imidacloprid insecticide in artificial agricultural field. *J. Environ. Heal. Sci. Eng.*, 1–8. doi: 10.1007/s40201-019-00391-w.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421.

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. *Bioinformatics* 35, 526–528.

Paul, E. A., and Clark, F. E. (1996). Components of the soil biota. *Soil Microbiol. Biochem. ed. EA Paul FE Clark*, 71–106.

Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., Fitzjohn, R. G., et al. (2014). geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30, 2216–2218.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2021). {nlme}: Linear and Nonlinear Mixed Effects Models. Available at: https://cran.r-project.org/package=nlme.

R Core Team (2021). R: A Language and Environment for Statistical Computing. Available at: https://www.r-project.org/.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223.

Ripple, W. (2017). World scientists' warning to humanity: a second notice. *Bioscience* 67, 1026–1028.

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. doi: 10.7287/peerj.preprints.1900v1.

Rodriguez, C. I., and Martiny, J. B. H. (2020). Evolutionary relationships among bifidobacteria and their hosts and environments. *BMC Genomics* 21, 1–12. doi: 10.1186/s12864-019-6435-1.

Romanenko, L. A., Tanaka, N., Frolova, G. M., Mikhailov, V. V, and Lyudmila Romanenko, C. A. (2009). Sphingomonas japonica sp. nov., isolated from the marine crustacean Paralithodes camtschatica. doi: 10.1099/ijs.0.003285-0.

Salazar, G. (2022). EcolUtils: Utilities for community ecology analysis. Available at: https://github.com/GuillemSalazar/EcolUtils.

Scales, N. C., Chase, A. B., Finks, S. S., Malik, A. A., Weihe, C., Allison, S. D., et al. (2022). Differential Response of Bacterial Microdiversity to Simulated Global Change. *Appl. Environ. Microbiol.* doi: 10.1128/aem.02429-21.

Schimel, J., Balser, T. C., and Wallenstein, M. (2007a). Microbial stress-response physiology and its implications for ecosystem function. *Ecology* 88, 1386–1394. doi: 10.1890/06-0219.

Schimel, J., Balser, T. C., and Wallenstein, M. (2007b). Microbial stress-response physiology and its implications for ecosystem function. *Ecology* 88, 1386–1394. doi: 10.1890/06-0219.

Schmidt, S., Wittich, R. M., Erdmann, D., Wilkes, H., Francke, W., and Fortnagel, P. (1992). Biodegradation of diphenyl ether and its monohalogenated derivatives by Sphingomonas sp. strain SS3. *Appl. Environ. Microbiol.* 58, 2744–2750. doi: 10.1128/aem.58.9.2744-2750.1992.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153.

Seidel, L., Ketzer, M., Broman, E., Shahabi-Ghahfarokhi, S., Rahmati-Abkenar, M., Turner, S., et al. (2022). Weakened resilience of benthic microbial communities in the face of climate change. *ISME Commun.* 2, 1–9. doi: 10.1038/s43705-022-00104-9.

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). "BUSCO: Assessing Genome Assembly and Annotation Completeness BT - Gene Prediction: Methods and Protocols," in, ed. M. Kollmar (New York, NY: Springer New York), 227–245. doi: 10.1007/978-1-4939-9173-0_14.

Singh, G., Verma, A. K., and Kumar, V. (2016). Catalytic properties, functional attributes and industrial applications of β-glucosidases. *3 Biotech* 6, 1–14. doi: 10.1007/s13205-015-0328-z.

Sinsabaugh, R. L., and Moorhead, D. L. (1994). Resource allocation to extracellular enzyme production: A model for nitrogen and phosphorus control of litter decomposition. *Soil Biol. Biochem.* 26, 1305–1311. doi: 10.1016/0038-0717(94)90211-9.

Slonczewski, J., and Foster, J. W. (2017). *Microbiology : an evolving science*.

Sorouri, B., Rodriguez, C. I., Gaut, B. S., and Allison, S. D. (n.d.). Variation in Sphingomonas traits across habitats and phylogenetic clades. Unpublishe.

St. John, M. G., Orwin, K. H., and Dickie, I. A. (2011). No "home" versus "away" effects of decomposition found in a grassland-forest reciprocal litter transplant study. *Soil Biol. Biochem.* 43, 1482–1489. doi: 10.1016/j.soilbio.2011.03.022.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033.

Stuart Chapin, F., Matson, P. A., and Vitousek, P. M. (2012). *Principles of terrestrial ecosystem ecology*. Springer New York doi: 10.1007/978-1-4419-9504-9.

Swift, M. J., W., H. O., M., A. J., Swift, M., Heal, O., and Anderson, J. (1979). *Decomposition in*

terrestrial ecosystems. Oxford : Blackwell Scientific.

Takeuchi, M., Hamana, K., and Hiraishi, A. (2001). Proposal of the genus Sphingomonas sensu stricto and three new genera, Sphingobium, Novosphingobium and Sphingopyxis, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol.* 51, 1405–1417. doi: 10.1099/00207713-51-4-1405.

Tang, Z. Z., Chen, G., and Alekseyenko, A. V. (2016). PERMANOVA-S: Association test for microbial community composition that accommodates confounders and multiple distances. in *Bioinformatics* (Oxford University Press), 2618–2625. doi: 10.1093/bioinformatics/btw311.

Tiedje, J. M., Bruns, M. A., Casadevall, A., Criddle, C. S., Eloe-Fadrosh, E., Karl, D. M., et al. (2022). Microbes and Climate Change: a Research Prospectus for the Future. doi: 10.1128/mbio.00800-22.

Torsvik, V., Øvreås, L., and Thingstad, T. F. (2002). Prokaryotic diversity - Magnitude, dynamics, and controlling factors. *Science (80-.).* 296, 1064–1066. doi: 10.1126/science.1071698.

van der Valk, A. G. (1981). Succession in Wetlands: A Gleasonian Appraoch. *Ecology* 62, 688–696. doi: 10.2307/1937737.

Walters, K. E., Capocchi, J. K., Albright, M. B. N., Hao, Z., Brodie, E. L., and Martiny, J. B. H. (2022). Routes and rates of bacterial dispersal impact surface soil microbiome composition and functioning. *ISME J.* 16, 2295–2304. doi: 10.1038/s41396-022-01269-w.

Wang, S., Spor, A., Nidelet, T., Montalent, P., Dillmann, C., De Vienne, D., et al. (2011). Switch

between life history strategies due to changes in glycolytic enzyme gene dosage in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* 77, 452–459. doi: 10.1128/AEM.00808-10.

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 42, D581–D591. doi: 10.1093/nar/gkt1099.

White, D. C., Sutton, S. D., and Ringelberg, D. B. (1996). The genus Sphingomonas: physiology and ecology. *Curr. Opin. Biotechnol.* 7, 301–306. doi: 10.1016/S0958-1669(96)80034-6.

Wohl, D. L., Arora, S., and Gladstone, J. R. (2004). Functional redundancy supports biodiversity and ecosystem function in a closed and constant environment. *Ecology* 85, 1534–1540. doi: 10.1890/03-3050.

Xie, G., Bruce, D. C., Challacombe, J. F., Chertkov, O., Detter, J. C., Gilna, P., et al. (2007). Genome sequence of the cellulolytic gliding bacterium Cytophaga hutchinsonii. *Appl. Environ. Microbiol.* 73, 3536–3546. doi: 10.1128/AEM.00225-07.

Yabuuchi, E., Yano, I., Oyaizu, H., Hashimoto, Y., Ezaki, T., and Yamamoto, H. (1990). Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb, nov., and Two Genospecies of the Genus *Sphingomonas*. *Microbiol. Immunol.* 34, 99–119. doi: 10.1111/j.1348-0421.1990.tb00996.x.

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., Da Rocha, U. N., Shi, S., et al. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3, 470–480. doi:

10.1038/s41564-018-0129-3.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46, W95. doi: 10.1093/NAR/GKY418.

Zhang, Y., Dong, S., Gao, Q., Liu, S., Ganjurjav, H., Wang, X., et al. (2017). Soil bacterial and fungal diversity differently correlated with soil biochemistry in alpine grassland ecosystems in response to environmental changes. *Sci. Rep.* 7, 1–10. doi: 10.1038/srep43077.

Zhao, Q., Yue, S., Bilal, M., Hu, H., Wang, W., and Zhang, X. (2017). Comparative genomic analysis of 26 Sphingomonas and Sphingobium strains: Dissemination of bioremediation capabilities, biodegradation potential and horizontal gene transfer. *Sci. Total Environ.* 609, 1238–1247. doi: https://doi.org/10.1016/j.scitotenv.2017.07.249.

Zhou, J., and Ning, D. (2017). Stochastic Community Assembly: Does It Matter in Microbial Ecology? *Microbiol. Mol. Biol. Rev.* 81. doi: 10.1128/mmbr.00002-17.