Title
Structure, Stability, and Function of the Microbiome of Soil Aggregates

Author
Lin, Jonathan Yueh Chun

Publication Date
2022

Peer reviewed|Thesis/dissertation

Structure, Stability, and Function of the Microbiome of Soil Aggregates

By

JONATHAN YUEH CHUN LIN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Jorge L.M. Rodrigues, Chair

_____

Kate M. Scow

_____

Vanessa L. Bailey

_____

Douglas C. Nelson

Committee in Charge

2022

ACKNOWLEDGEMENTS

Few things in life are guaranteed, including completing a PhD. Many people have contributed to my success at UC Davis over the past six years to whom I am eternally grateful.

I would first like to thank my advisor, Jorge Rodrigues. Thank you for letting me join your lab and for the freedom to develop my own project. I am grateful to my dissertation committee members Kate Scow, Vanessa Bailey, and Doug Nelson. Your expertise and constructive feedback were tremendously helpful. They have helped me to become a better scientist. I also thank my qualifying exam committee members Doug Nelson, Becky Parales, Rachel Vannette, Johan Leaveau, and Angie Gelli.

I am extremely grateful to Daoyuan Wang, for without him the genesis and development of my dissertation project could not have been possible. I greatly enjoyed our discussions, sampling in the field, and processing soil samples together. I would not have been able to complete this PhD without his insight and collaboration.

I am grateful to Laibin Huang, Christian Erikson, Jane Fudyma, Erika Yao, Guadalupe Barajas, David Lipson, and Majdi Abou Najm. I am indebted to you all for helping me develop new ideas and generating data.

I am grateful for the undergraduates that I had the pleasure to mentor during their time at UC Davis: Sung Won, King Law, and Kate Johnson. Thank you for your curiosity and attentiveness. Your youthful tenacity helped keep me afloat during trying times.

Thank you to my lab colleagues – both past and present – whom I have not already mentioned: Jordan Sayre, Rachel Danielson, Alex Gulachenski, Julio Moreira, and Eloi Parlade. I am grateful for the camaraderie that we built.

I am deeply appreciative of Chorong Park and Rene Suleiman. Thanks also to Ariadna Venegas-Li and Megan Shyr, Andrew Wade and Michael Messina, and Jason Loxterkamp, Jess Dill, Zoe Kanavas, and Ngan Mai. You have all enriched my life throughout the past six years.

I would not have sought to do a PhD without the experiences I had in John Wertz's lab at Calvin University. He took a chance on me and gave me the opportunity to work with him both as an undergraduate researcher and research technician. The impact was profound – it was through him that I fell in love in with microbiology. I still think of those formative years fondly.

Finally, I am beyond thankful for my parents Cheng-Ming and Sherry, and my brother Kevin. Your continual support and encouragement have been nothing short of a herculean effort. Everything that I am and all my success I owe to you.

This dissertation research was supported by a USDA NIFA AFRI Predoctoral Fellowship (2020-67034-31937), a Henry A. Jastro Award from the UC Davis College of Agriculture and Environmental Sciences, an Abigail Salyers Endowed Scholarship from the Marine Biological Lab (to attend the Microbial Diversity Course in Woods Hole, MA in summer 2017), and a DOE Joint Genome Institute Community Science Program allocation (proposal #505656).

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF TABLES AND FIGURES vi

ABSTRACT ix

CHAPTERS

1. INTRODUCTION 1

1.1. References 8

2. CHAPTER 1: Compost amendment and cover cropping maintains soil structure and carbon storage by increasing available carbon and microbial biomass in an agricultural soil – a six-year field study 14

2.1. Abstract 14

2.2. Introduction 15

2.3. Materials and Methods 18

2.4. Results 24

2.5. Discussion 26

2.6. References 33

2.7. Tables and Figures 38

2.8. Supplementary Materials 44

3. CHAPTER 2: Differential responses of prokaryotic and fungal communities in soil microenvironments to drying and wetting as affected by soil aggregate isolation method 46

3.1. Abstract 46

3.2. Introduction 47

3.3. Materials and Methods 50

3.4. Results 55

3.5. Discussion 60

3.6. References 71

3.7. Tables and Figures 78

3.8. Supplementary Materials 87

4. CHAPTER 3: Stratification of microbial community taxonomy and function in soil aggregates of different size 98

4.1. Abstract 98

4.2. Introduction 99

4.3. Materials and Methods 101
4.4. Results 107
4.5. Discussion 113
4.6. References 123
4.7. Tables and Figures 131
4.8. Supplementary Materials 138

APPENDIX

A. APPENDIX A: *Geminisphaera* 167
A.1. Abstract 167
A.2. Description 167
A.3. References 175
A.4. Tables and Figures 177

B. APPENDIX B: *Treponema* endosymbionts are the dominant bacterial members with ureolytic potential in the gut of the wood-feeding termite, *Reticulitermes hesperus* 179
B.1. Abstract 179
B.2. Main Text 180
B.3. References 185
B.4. Tables and Figures 187
B.5. Supplementary Materials 190

LIST OF TABLES AND FIGURES

Table 2-1. Distribution of water stable aggregate size fractions 38

Table 2-2. Soil C distribution in soil aggregate fractions 39

Table 3-1. PERMANOVA results for prokaryotes and fungi in aggregates 78

Table 4-1. PERMANOVA results for microbial taxonomy and function in aggregates 131

Table S2-1. Tomato fruit and corn grain yields 44

Table S3-1. Gravimetric water content of bulk soils 87

Table S3-2. Characteristics of bulk soils 88

Table S3-3. Relative abundances of the top 20 prokaryotic taxa at the genus level 89

Table S3-4. Relative abundances of the top 20 fungal taxa at the genus level 90

Table S3-5. PERMANOVA analysis for sieving treatments nested within aggregate size 91

Table S4-1. Metagenome assembly metrics 138

Table S4-2. Microbial taxonomy, functional gene, and metabolite differences by fertilizer treatment 139

Table S4-3. Differentially abundant KEGG genes by aggregate size 140

Table S4-4. Metrics and taxonomic classification of MAGS 147

Table A-1. Differential characteristics between *G. colitermitum* and other type species 177

Table B-1. Abundance and proportion of *ureC* genes in termite guts 187

Table SB-1. Identity, abundance, and ureolytic potential of isolated cultivated from termite guts 196

Table SB-2. Components of trace elements solution 197

Table SB-3. Barcodes used for library preparation of *ureC* gene amplicons 197

Figure 1-1. Simplified diagram depicting soil aggregates 4

Figure 2-1. Soil aggregate stability under different management 40

Figure 2-2. Soil bulk C under different management 40

Figure 2-3. Soil dissolved organic C and microbial biomass C under different management 41

Figure 2-4. 16S gene copy numbers under different management 41

Figure 2-5. ITS copy numbers under different management 42

Figure 2-6. Weighted average 16S rRNA gene and internal transcribed spacer copy numbers 42

Figure 2-7. NMDS plots of prokaryotic communities in aggregates 43

Figure 2-8. Potential for degradation of aromatic C compounds by management 43

Figure 3-1. Conceptual diagram showing experimental manipulations 79

Figure 3-2. NMDS plots of prokaryotes and fungi in soil aggregates 80

Figure 3-3. NMDS plots of prokaryotes and fungi separated by aggregate size 81

Figure 3-4. Alpha diversity of prokaryotes in each sieving treatment separated by aggregate size 82

Figure 3-5. Alpha diversity of fungi in each sieving treatment separated by aggregate size 83

Figure 3-6. Community-weighted mean estimated traits by aggregate size 84

Figure 3-7. Relative abundance of fungal trophic groups and guilds 85

Figure 3-8. Total C and N in each sieving treatment separated by aggregate size 86

Figure 4-1. NMDS plots of microbial taxonomy and function in aggregates 132

Figure 4-2. Alpha diversity of microbial taxonomy and function in aggregates 133

Figure 4-3. Heatmap of microbial taxa at the genus level in aggregates 134

Figure 4-4. Heatmap of microbial functional genes and representation in MAGS in aggregates 135

Figure 4-5. Analysis of metabolites in aggregates 136

Figure 4-6. Cell metabolism diagrams of ammonia-oxidizing archaea MAGS 137

Figure S2-1. Plot layout of the field trial 44

Figure S2-2. NMDS of soil bacterial communities by management and biochar 45

Figure S3-1. Within-group dispersions for prokaryotes and fungi by sieving treatment 92

Figure S3-2. Alpha diversity of prokaryotes and fungi by sieving treatment 92

Figure S3-3. Bubble plot of the 10 most abundant prokaryotic phyla 93

Figure S3-4. Bubble plot of the 10 most abundant fungal phyla 94

Figure S3-5. Community-weighted mean estimated traits by sieving treatment 95

Figure S3-6. Aggregate total C and N by sieving treatment 96

Figure S3-7. Spearman correlations of community-weighted traits with aggregate C and N..... 97

Figure S4-1. Alpha diversity of microbial taxonomy and function by fertilizer treatment..... 159

Figure S4-2. Pathway analysis based on KEGG genes by fertilizer treatment..... 160

Figure S4-3. Taxonomic coverage of differentially abundant genes in macro- vs. microaggregates.... 161

Figure S4-4. Differentially abundant metabolites in aggregates..... 162

Figure S4-5. Analysis of soil enzyme activities in aggregates..... 163

Figure S4-6. Phylogenetic tree and abundance of bacterial and archaeal MAGs..... 164

Figure S4-7. Phylogenetic tree of archaeal MAGs..... 165

Figure S4-8. Relative abundance of MAGs and ammonia-oxidizing archaea in aggregates..... 166

Figure A-1. Scanning electron micrograph of *G. colitermitum*..... 178

Figure A-2. Phylogenetic tree of *G. colitermitum* and other type strains..... 178

Figure B-1. Sampling location, termite identification, *ureC* gene composition in termites..... 188

Figure B-2. Phylogenetic tree of Urec_98 and other *ureC* protein-coding sequences..... 189

ABSTRACT

Most soil microorganisms live in communities within and on the surface of soil aggregates, the three-dimensional complexes composed of organic materials and particles that make up the soil physical structure. Soil aggregates vary considerably in physical and chemical properties by size, making them unique habitats for distinct microbial communities and metabolisms. However, this fine-scale spatial variability for microbes has received relatively little attention, and a better understanding of microbial community dynamics at these levels is crucial for predicting microbially-mediated soil functional responses to changing environments. This dissertation investigates the microbial communities in soil aggregates using three approaches. First, soil carbon (C) and aggregation dynamics were studied in an agricultural field experiment comparing the long-term impacts of manure compost and mineral fertilizer on C storage. Compost amendments increased soil total C, microbial biomass C, and maintained aggregate stability compared to mineral fertilizer. The prokaryotic community differed in composition between aggregates of different size fractions ranging from large macroaggregates (> 2 mm) to silt & clay (< 250 μm) and showed an increased capacity for potential degradation of aromatic C compounds with compost amendment. This suggests that yearly additions of compost can increase the diversity of substrates for microbes to increase biomass and aggregation through microbial activities. Next, because consistent relationships between microbial communities and aggregate size have remained elusive in the literature, two commonly used aggregate isolation methods, dry and wet sieving, were compared to identify their effects on prokaryotic and fungal communities in soils with different starting moistures. While the prokaryotic community was different by sieving treatment in each aggregate size fraction, the alpha diversity and composition of the fungal community were more resistant to change in the large and small macroaggregates than in the microaggregates and silt & clay. Drying soils prior to sieving favored spore-forming fast-growing generalist prokaryotes and fungal saprotrophs, whereas rewetting soils through wet

sieving resulted in more slow-growing specialist prokaryotes and fungal pathogens. These results show that dry and wet sieving soils with different starting moistures can drastically affect the microbial communities in aggregates through drying and wetting dynamics. Finally, to better understand the different microbial taxa driving different functions in soil aggregates, a shotgun metagenomics approach was used to analyze the taxonomic and functional gene composition with the metabolic output of the microbiome in four size fractions of aggregates. Higher abundances of genes for the degradation of plant-derived compounds and biofilm formation were found in the macroaggregates, while the microaggregates and silt & clay were more enriched in genes for biomass recycling and anaerobic respiration. Both taxonomic profiling and reconstruction of genomes from metagenomes revealed a higher abundance of ammonia-oxidizing archaea (AOA) in macroaggregates, and further analysis of their genomic content revealed complementary metabolisms potentially enabling distinct AOA lineages to colonize different niches within the same habitat. These results characterize macroaggregates and microaggregates as resource-rich and resource-poor environments for microbes, respectively, and in conjunction with the other chapters in this dissertation, advances knowledge of the composition, stability, and function of microbial communities stratified in soil aggregates of different size.

INTRODUCTION

Soil microbial diversity and function

Soils are habitats teeming with microbial life – harboring abundances up to 10^{10} bacterial cells and diversities up to 10^{6} species per gram (Delgado-Baquerizo et al., 2016; Bahram et al., 2018). This immense richness underscores soil microbes as an important biological resource where they play crucial roles in sustaining terrestrial nutrient cycling, promoting plant growth, and protecting human health (Samaddar et al., 2021). Unfortunately, many of the beneficial functions performed by the soil microbiome are currently threatened due to changing climate and a fast-expanding human population (Jansson and Hofmockel, 2020). Land use change (Rodrigues et al., 2013; Malik et al., 2018), agricultural intensification (Levine et al., 2011; Hartmann et al., 2015; Hartman et al., 2018; Banerjee et al., 2019), rising temperatures (Dunbar et al., 2012; Woodcroft et al., 2018; Wilson et al., 2021), and drought (de Vries et al., 2018; Schimel, 2018) have resulted in microbial diversity loss, soil carbon (C) depletion, and decreased soil health. In the face of these daunting challenges, improved management practices and restoration techniques are needed to leverage the soil microbiome to build resilience and promote ecosystem functioning.

A better understanding of the fundamental relationship between soil microbial diversity and ecosystem functioning is a critical first step towards developing solutions. However, while it is assumed that high microbial biodiversity enhances ecosystem stability and productivity, consistent relationships between microbial diversity and specific ecosystem functions have not been observed (Allison and Martiny, 2008). Interestingly, several studies manipulating the diversity of the soil microbiome found no links between diversity and composition with microbially-driven processes such as nitrification (Griffiths et al., 2001; Wertz et al., 2006), denitrification (Wertz et al., 2006; Rousk et al., 2011), and C mineralization (Griffiths et al., 2001; Wertz et al., 2006; Rousk et al., 2011). These results contrast with recent reports showing the

importance of soil microbial diversity for controlling denitrification (Philippot et al., 2013), assimilating C from plant residues (Maron et al., 2018) and maintaining multiple functions simultaneously (Delgado-Baquerizo et al., 2016). This lack of a consistent relationship across the literature raises questions about the importance of diversity and specific microbiome compositional states for soil and ecosystem functioning (Raynaud and Nunan, 2014) and makes the interpretation of results for field practices more difficult. Thus, more work is needed to disentangle the factors confounding the relationship between microbiome dynamics with ecosystem functioning in soils.

Spatial scales relevant for microbial interactions in soil

Microbial diversity affects ecosystem processes through interactions between species such as competition for resources, mutualism such as cross-feeding, or predation (Raynaud and Nunan, 2014). These interactions, which together contribute to ecosystem functioning, depend not only on the species present but also the spatial proximity of microbial cells to each other (Nunan et al., 2007; Raynaud and Nunan, 2014). Naturally, the length scales at which microbes interact are miniscule compared to that for plants and animals due to their small size. Previous studies have shown that most bacteria and fungi in soil preferentially occupy pores between 1 μm and 1 mm in size (Chenu and Cosentino, 2011) and that most interactions by bacterial cells occur within a space of 20 μm (Deschesne et al., 2007; Nunan et al., 2007). Yet, despite the vast number of studies focusing on microbial diversity in soils, relatively little attention has been paid to microbial community dynamics at the scales where most microbial interactions occur (Deschesne et al., 2007). One reason for this limited understanding is that most studies typically use a bulk soil sampling approach where entire soil cores are collected and homogenized before analysis. This method, which mixes soil prior to DNA or lipid-based community analysis, is equivalent to analyzing plant communities in a transect spanning from Los Angeles to New York

and, consequently, hinders analysis of microbial community dynamics organized in the microenvironments therein. This is particularly important because soils are some of the most heterogeneous environments in the biosphere with hotspots that can occur in microsites within a single soil core (Parkin, 1987) and extremely high differentiation of physical and chemical properties within micron to millimeter length distances (Kuzyakov and Blagodatskaya, 2015). Hence, to better understand the role that microbial community dynamics play in soil ecosystem functions, it is necessary to study the microscale spatial organization with enough throughput to measure statistical associations of the taxa and their functions (Cordero and Datta, 2016).

Soil physical structure

The soil is a heterogeneous matrix made up of aggregates – three-dimensional complexes composed of organic materials and mineral particles (Wilpiszeski et al., 2019). In all soils, minerals and fine-sized silt particles are bound together by plant and microbial debris and humified organic matter to form microaggregates (< 250 μm) (Tisdall and Oades, 1982; Jastrow et al., 2007). In many soils where organic matter is a major aggregate-binding agent, microaggregates are further enmeshed by labile organic material and plant roots, fungal hyphae, and bacterial polysaccharides to form macroaggregates (> 250 μm) (Tisdall and Oades, 1982; Jastrow et al., 2007; Totsche et al., 2017). This hierarchy of aggregates assembled by size – along with the pores within and between each aggregate – comprises the physical structure of the soil (Tisdall and Oades, 1982; Six et al., 2000, 2004) (**Figure 1-1**).

Soil aggregation is an important property of agricultural ecosystems and plays many crucial roles including promoting soil C storage, contributing to soil water infiltration, and maintaining agricultural productivity (Totsche et al., 2017; Wilpiszeski et al., 2019). Aggregates can also determine the niche availability for resident soil microbes, where they differ in physical and chemical properties from the bulk soil as well as from each other by size. For example, steep

declines in oxygen concentrations have been measured within small distances from the surfaces of wetted macroaggregates, revealing the presence of anoxic microsites (Sexstone et al., 1985; Jastrow et al., 2007). On the other hand, oxygen diffusion into microaggregates is likely constrained to the surface due to pores that are much smaller in diameter compared to the pores in macroaggregates (Sey et al., 2008). In addition to oxygen, differences in pore sizes can also influence the diffusion of water and other nutrients from the soil matrix into aggregate interiors (Totsche et al., 2017) or protect bacteria from predation by microfauna (Vargas and Hattori, 1986). Macroaggregates contain more particulate organic matter incorporated from plant residues that may be more susceptible to microbial decomposition, whereas microaggregates have more phenolic C adsorbed to mineral particles that may be protected from microbial access (Davinic et al., 2012; Yu et al., 2015). These differences provide evidence that the unique environments in aggregates contain microhabitats for distinct microbial communities and metabolic pathways (Foster, 1988; Vos et al., 2013; Rillig et al., 2017).

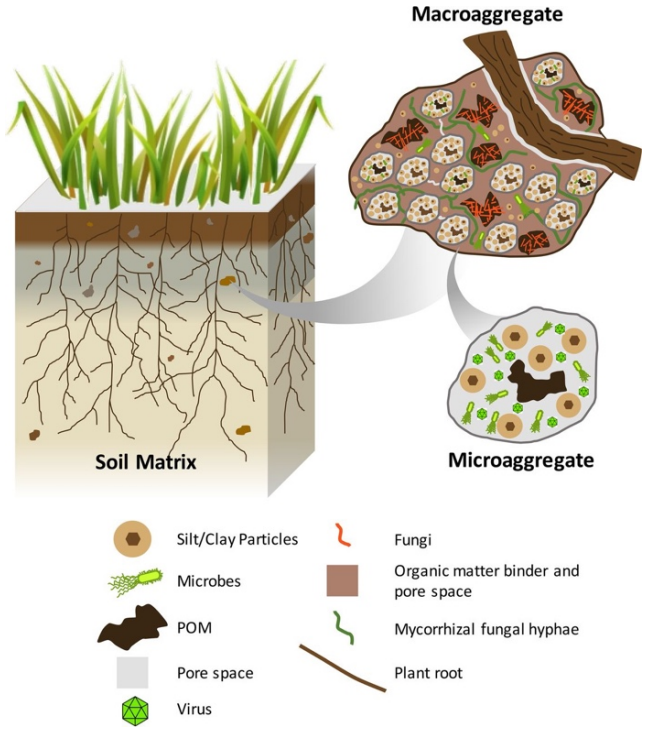


Figure 1-1. Simplified diagram depicting soil macro- and microaggregates and their components within the soil matrix. (From Wilpiszeski et al., 2019).

Soil aggregates as habitats for microorganisms

While aggregates have long been hypothesized as habitats for microorganisms in the soil environment (Allison, 1968; Hattori, 1988), studying microbial communities in aggregates has seen renewed interested due to the development of modern techniques for characterizing microbiomes (Cordero and Datta, 2016; Wilpiszeski et al., 2019). To date, previous studies have shown that different size fractions of aggregates harbor distinct communities of bacteria and archaea (Davinic et al., 2012; Trivedi et al., 2017; Bach et al., 2018; Yang et al., 2019), fungi (Bach et al., 2018; Li et al., 2019; Upton et al., 2019; Yang et al., 2019), and protists (Liao et al., 2021). Different aggregate size fractions also have different total abundances of bacteria (Kanazawa and Filip, 1986; Helgason et al., 2010; Gupta and Germida, 2015; Trivedi et al., 2015, 2017; Wang et al., 2017), nitrogen (N) cycling genes (Wang et al., 2018; Li et al., 2019; Han et al., 2020), and extracellular enzyme activities (Fansler et al., 2005; Bach and Hofmockel, 2014; Trivedi et al., 2017; Wang et al., 2017; Han et al., 2021). Microbial communities have also been studied in single macroaggregates, revealing an overall patchy distribution (Bailey et al., 2013b), variable enzyme rates (Bailey et al., 2012, 2013a), and preferences by a few bacterial taxa for certain pore sizes (Kravchenko et al., 2014) among individual aggregates.

A major theme emerging from these studies, however, is a lack of a consistent relationship between specific taxa, diversity, and functions with aggregates of differing size across the literature. For instance, while some studies found a higher abundance of *Actinobacteria* in microaggregates (Mummey et al., 2006; Bach et al., 2018), others have reported that they preferentially colonize macroaggregates (Davinic et al., 2012; Trivedi et al., 2017). Others have reported contrasting findings of macro- vs microaggregates as hotspots of microbial diversity (Davinic et al., 2012; Bach et al., 2018). To further add to this confusion, a number of studies have found no differences in microbial community composition between different size fractions of aggregates (Ramakrishnan et al., 2000; Schutter and Dick, 2002; Fall

et al., 2004; Kim et al., 2008; Kim and Crowley, 2013; Blaud et al., 2014, 2018). These inconsistent patterns suggest that more studies are needed to arrive at generalizations of microbial community dynamics in soil aggregates.

Dissertation research

The research in this dissertation adds to the body of knowledge on soil aggregate microbial communities by 1) investigating the long-term impacts of different agricultural inputs on soil aggregation, 2) clarifying the effects of different methods used to separate aggregates on prokaryotic and fungal communities, and 3) linking microbial community taxonomy and function in aggregates of different sizes using an omics approach.

Chapter 1 investigates soil aggregates in a long-term field trial comparing the effects of continuous manure compost or mineral fertilizer with or without biochar on soil C and aggregation dynamics in a Mediterranean cropping system. Across all treatments, 6 years of compost addition increased microbial biomass C, dissolved organic C, and maintained soil aggregate stability compared to the mineral fertilizer treatment. 16S rRNA gene sequencing revealed that the prokaryotic community composition was distinct in different aggregate size fractions, with higher abundances of *Micrococcales*, *Streptomycetales*, *Propionibacteriales*, and *Sphingomonadales* in the microaggregate and silt & clay fractions, and higher proportions of *Gentimonadetes*, *Nitrososphaerales*, and *Nitrospirales* in the large and small macroaggregates. Across all aggregate sizes, the abundance of bacteria potentially capable of degrading aromatic C was higher under compost than mineral fertilizer. These findings suggest that continuous compost amendment can potentially generate a positive feedback loop for enhancing C storage by increasing microbially-available substrates, in turn increasing soil microbial biomass, and thereby increasing aggregation through microbial activities.

Chapter 2 compares two common aggregate isolation methods, dry and wet sieving, to investigate their effects on the prokaryotic and fungal communities in four aggregate size fractions originating from field moist and dried soils. Although the composition of the prokaryotic community was different among the sieving treatments in each of the four size fractions, the composition and alpha diversity for fungi were more resistant to change in the large and small macroaggregates than in the microaggregates and silt & clay. While drying soils prior to sieving favored spore-forming fast-growing generalist prokaryotes and fungal saprotrophs, rewetting soils through wet sieving resulted in more slow-growing specialist prokaryotes and fungal pathogens in aggregates in tandem with decreasing aggregate C and N in the smallest aggregate size fractions. These results show that dry and wet sieving soils with different starting moisture conditions can affect the microbial communities in aggregates through drying and wetting dynamics. Hence, researchers are recommended to consider their hypotheses before choosing between methods as each is representative of the microbial community under strikingly different conditions of moisture saturation and disturbance.

Chapter 3 investigates the taxonomic community, functional gene composition, and metabolic output of the microbial community in four size fractions of soil aggregates. The functional gene composition identified using shotgun metagenomics was different by aggregate size, with higher abundances of genes for the degradation of plant-derived compounds and biofilm formation in the macroaggregates and biomass recycling and denitrification in the microaggregates and silt & clay. These differences corroborated with differences in the composition of the metabolome but not specific enzyme activities. Both taxonomic profiling and reconstruction of genomes from metagenomes revealed a higher abundance of ammonia-oxidizing archaea in the macroaggregates, and further analysis of their genomes revealed complementary metabolisms potentially enabling them to colonize different niches within the same habitat. These results characterize macroaggregates and microaggregates as resource-

rich and resource-poor environments for microbes, respectively, acting as environmental filters to drive differences in soil microbial community taxonomy and function in the soil matrix.

REFERENCES

Allison, F.E., 1968. Soil aggregation—Some facts and fallacies as seen by a microbiologist. Soil Science 106, 136.

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

Bach, E.M., Hofmockel, K.S., 2014. Soil aggregate isolation method affects measures of intra-aggregate extracellular enzyme activity. Soil Biology and Biochemistry 69, 54–62.

Bach, E.M., Williams, R.J., Hargreaves, S.K., Yang, F., Hofmockel, K.S., 2018. Greatest soil microbial diversity found in micro-habitats. Soil Biology and Biochemistry 118, 217–226.

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

Bailey, V.L., Bilskis, C.L., Fansler, S.J., McCue, L.A., Smith, J.L., Konopka, A., 2012. Measurements of microbial community activities in individual soil macroaggregates. Soil Biology and Biochemistry 48, 192–195.

Bailey, V.L., Fansler, S.J., Stegen, J.C., McCue, L.A., 2013a. Linking microbial community structure to β-glucosidic function in soil aggregates. The ISME Journal 7, 2044–2053.

Bailey, V.L., McCue, L.A., Fansler, S.J., Boyanov, M.I., DeCarlo, F., Kemner, K.M., Konopka, A., 2013b. Micrometer-scale physical structure and microbial composition of soil macroaggregates. Soil Biology and Biochemistry 65, 60–68.

Banerjee, S., Walder, F., Büchi, L., Meyer, M., Held, A.Y., Gattinger, A., Keller, T., Charles, R., van der Heijden, M.G.A., 2019. Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. The ISME Journal 13, 1722–1736.

Blaud, A., Chevallier, T., Virto, I., Pablo, A.-L., Chenu, C., Brauman, A., 2014. Bacterial community structure in soil microaggregates and on particulate organic matter fractions located outside or inside soil macroaggregates. Pedobiologia 57, 191–194.

Blaud, A., van der Zaan, B., Menon, M., Lair, G.J., Zhang, D., Huber, P., Schiefer, J., Blum, W.E.H., Kitzler, B., Wei, E.H., van Gaans, P., Banwart, S., 2018. The abundance of nitrogen cycle genes and potential greenhouse gas fluxes depends on land use type and little on soil aggregate size. Applied Soil Ecology 125, 1–11.

Chenu, C., Cosentino, D., 2011. Microbial regulation of soil structural dynamics., in: Ritz, K., Young, I. (Eds.), The Architecture and Biology of Soils: Life in Inner Space. CABI, Wallingford, pp. 37–70.

Cordero, O.X., Datta, M.S., 2016. Microbial interactions and community assembly at microscales. Current Opinion in Microbiology 31, 227–234.

Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak, J.C., Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. Soil Biology and Biochemistry 46, 63–72.

De Vries, F.T., Griffiths, R.I., Bailey, M., Craig, H., Girlanda, M., Gweon, H.S., Hallin, S., Kaisermann, A., Keith, A.M., Kretzschmar, M., Lemanceau, P., Lumini, E., Mason, K.E., Oliver, A., Ostle, N., Prosser, J.I., Thion, C., Thomson, B., Bardgett, R.D., 2018. Soil bacterial networks are less stable under drought than fungal networks. Nature Communications 9, 3033.

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

Deschesne, A., Pallud, C., Grundmann, G.L., 2007. Spatial distribution of bacteria at the microscale In soil, in: The Spatial Distribution of Microbes in the Environment. Springer, Dordrecht, pp. 87–107.

Dunbar, J., Eichorst, S.A., Gallegos-Graves, L.V., Silva, S., Xie, G., Hengartner, N.W., Evans, R.D., Hungate, B.A., Jackson, R.B., Megonigal, J.P., Schadt, C.W., Vilgalys, R., Zak, D.R., Kuske, C.R., 2012. Common bacterial responses in six ecosystems exposed to 10 years of elevated atmospheric carbon dioxide. Environmental Microbiology 14, 1145–1158.

Fall, S., Nazaret, S., Chotte, J.L., Brauman, A., 2004. Bacterial density and community structure associated with aggregate size fractions of soil-feeding termite mounds. Microbial Ecology 48, 191–199.

Fansler, S.J., Smith, J.L., Bolton, H., Bailey, V.L., 2005. Distribution of two C cycle enzymes in soil aggregates of a prairie chronosequence. Biology and Fertility of Soils 42, 17–23.

Foster, R.C., 1988. Microenvironments of soil microorganisms. Biology and Fertility of Soils 6, 189–203.

Griffiths, B.S., Ritz, K., Wheatley, R., Kuan, H.L., Boag, B., Christensen, S., Ekelund, F., Sørensen, S.J., Muller, S., Bloem, J., 2001. An examination of the biodiversity–ecosystem function relationship in arable soil microbial communities. Soil Biology and Biochemistry 33, 1713–1722.

Gupta, V.V.S.R., Germida, J.J., 2015. Soil aggregation: Influence on microbial biomass and implications for biological processes. Soil Biology and Biochemistry 80, A3–A9.

Han, S., Delgado-Baquerizo, M., Luo, X., Liu, Y., Van Nostrand, J.D., Chen, W., Zhou, J., Huang, Q., 2021. Soil aggregate size-dependent relationships between microbial functional diversity and multifunctionality. Soil Biology and Biochemistry 154, 108143.

Han, S., Luo, X., Tan, S., Wang, J., Chen, W., Huang, Q., 2020. Soil aggregates impact nitrifying microorganisms in a vertisol under diverse fertilization regimes. European Journal of Soil Science 71, 536–547.

Hartman, K., van der Heijden, M.G.A., Wittwer, R.A., Banerjee, S., Walser, J.-C., Schlaeppi, K., 2018. Cropping practices manipulate abundance patterns of root and soil microbiome members paving the way to smart farming. Microbiome 6, 14.

Hartmann, M., Frey, B., Mayer, J., Mader, P., Widmer, F., 2015. Distinct soil microbial diversity under long-term organic and conventional farming. The ISME Journal 9, 1177–1194.

Hattori, T., 1988. Soil aggregates as microhabitats of microorganisms. Reports of the Institute for Agricultural Research – Tohoku University (Japan).

Helgason, B.L., Walley, F.L., Germida, J.J., 2010. No-till soil management increases microbial biomass and alters community profiles in soil aggregates. Applied Soil Ecology 46, 390–397.

Jansson, J.K., Hofmockel, K.S., 2020. Soil microbiomes and climate change. Nature Reviews Microbiology 18, 35–46.

Jastrow, J.D., Amonette, J.E., Bailey, V.L., 2007. Mechanisms controlling soil carbon turnover and their potential application for enhancing carbon sequestration. Climatic Change 80, 5–23.

Kanazawa, S., Filip, Z., 1986. Distribution of microorganisms, total biomass, and enzyme activities in different particles of brown soil. Microbial Ecology 12, 205–215.

Kim, J.-S., Crowley, D.E., 2013. Size fractionation and microbial community structure of soil aggregates. Journal of Agricultural Chemistry and Environment 02, 75.

Kim, J.-S., Dungan, R.S., Crowley, D., 2008. Microarray analysis of bacterial diversity and distribution in aggregates from a desert agricultural soil. Biology and Fertility of Soils 44, 1003.

Kravchenko, A.N., Negassa, W.C., Guber, A.K., Hildebrandt, B., Marsh, T.L., Rivers, M.L., 2014. Intra-aggregate pore structure influences phylogenetic composition of bacterial community in macroaggregates. Soil Science Society of America Journal 78, 1924–1939.

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry 83, 184–199.

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

Li, F., Qiu, P., Shen, B., Shen, Q., 2019. Soil aggregate size modifies the impacts of fertilization on microbial communities. Geoderma 343, 205–214.

Liao, H., Gao, S., Hao, X., Qin, F., Ma, S., Chen, W., Huang, Q., 2021. Soil aggregate isolation method affects interpretation of protistan community. Soil Biology and Biochemistry 161, 108388.

Malik, A.A., Puissant, J., Buckeridge, K.M., Goodall, T., Jehmlich, N., Chowdhury, S., Gweon, H.S., Peyton, J.M., Mason, K.E., Agtmaal, M. van, Blaud, A., Clark, I.M., Whitaker, J., Pywell, R.F., Ostle, N., Gleixner, G., Griffiths, R.I., 2018. Land use driven change in soil pH affects microbial carbon cycling processes. Nature Communications 9, 3591.

Maron, P.-A., Sarr, A., Kaisermann, A., Lévêque, J., Mathieu, O., Guigue, J., Karimi, B., Bernard, L., Dequiedt, S., Terrat, S., Chabbi, A., Ranjard, L., 2018. High microbial diversity promotes soil ecosystem functioning. Applied and Environmental Microbiology AEM.02738-17.

Mummey, D., Holben, W., Six, J., Stahl, P., 2006. Spatial stratification of soil bacterial populations in aggregates of diverse soils. Microbial Ecology 51, 404–411. Doi:10.1007/s00248-006-9020-5

Nunan, N., Young, I.M., Crawford, J.W., Ritz, K., 2007. Bacterial interactions at the microscale – linking habitat to function in soil, in: The Spatial Distribution of Microbes in the Environment. Springer, Dordrecht, pp. 61–85. Doi:10.1007/978-1-4020-6216-2_3

Parkin, T.B., 1987. Soil microsites as a source of denitrification variability. Soil Science Society of America Journal 51, 1194–1199.

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M., Sarr, A., Maron, P.-A., 2013. Loss in microbial diversity affects nitrogen cycling in soil. The ISME Journal 7, 1609–1619. Doi:10.1038/ismej.2013.34

Ramakrishnan, B., Lueders, T., Conrad, R., Friedrich, M., 2000. Effect of soil aggregate size on methanogenesis and archaeal community structure in anoxic rice field soil. FEMS Microbiology Ecology 32, 261–270.

Raynaud, X., Nunan, N., 2014. Spatial ecology of bacteria at the microscale in soil. PLOS ONE 9.

Rillig, M.C., Muller, L.A., Lehmann, A., 2017. Soil aggregates as massively concurrent evolutionary incubators. The ISME Journal 11, 1943–1948.

Rodrigues, J.L.M., Pellizari, V.H., Mueller, R., Baek, K., Jesus, E. da C., Paula, F.S., Mirza, B., Hamaoui, G.S., Tsai, S.M., Feigl, B., Tiedje, J.M., Bohannan, B.J.M., Nüsslein, K., 2013. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proceedings of the National Academy of Sciences 110, 988–993.

Rousk, J., Brookes, P.C., Glanville, H.C., Jones, D.L., 2011. Lack of correlation between turnover of low-molecular-weight dissolved organic carbon and differences in microbial community composition or growth across a soil pH gradient. Applied and Environmental Microbiology 77, 2791–2795.

Samaddar, S., Karp, D.S., Schmidt, R., Devarajan, N., McGarvey, J.A., Pires, A.F.A., Scow, K., 2021. Role of soil in the regulation of human and plant pathogens: Soils' contributions to people. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20200179.

Schimel, J.P., 2018. Life in dry soils: Effects of drought on soil microbial communities and processes. Annual Review of Ecology, Evolution, and Systematics 49, 409–432.

Schutter, M.E., Dick, R.P., 2002. Microbial community profiles and activities among aggregates of winter fallow and cover-cropped soil. Soil Science Society of America Journal 66, 142–153.

Sexstone, A.J., Parkin, T.B., Tiedje, J.M., 1985. Temporal Response of Soil Denitrification Rates to Rainfall and Irrigation. Soil Science Society of America Journal 49, 99–103.

Sey, B.K., Manceur, A.M., Whalen, J.K., Gregorich, E.G., Rochette, P., 2008. Small-scale heterogeneity in carbon dioxide, nitrous oxide and methane production from aggregates of a cultivated sandy-loam soil. Soil Biology and Biochemistry 40, 2468–2473.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil and Tillage Research, Advances in Soil Structure Research 79, 7–31.

Six, J., Paustian, K., Elliott, E.T., Combrink, C., 2000. Soil structure and organic matter I. Distribution of aggregate-size classes and aggregate-associated carbon. Soil Science Society of America Journal 64, 681–689.

Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. Journal of Soil Science 33, 141–163.

Totsche, K.U., Amelung, W., Gerzabek, M.H., Guggenberger, G., Klumpp, E., Knief, C., Lehndorff, E., Mikutta, R., Peth, S., Prechtel, A., Ray, N., Kögel-Knabner, I., 2017. Microaggregates in soils. Journal of Plant Nutrition and Soil Science 181, 104–146.

Trivedi, P., Delgado-Baquerizo, M., Jeffries, T.C., Trivedi, C., Anderson, I.C., Lai, K., McNee, M., Flower, K., Pal Singh, B., Minkey, D., Singh, B.K., 2017. Soil aggregation and associated microbial communities modify the impact of agricultural management on carbon content. Environmental Microbiology 19, 3070–3086.

Trivedi, P., Rochester, I.J., Trivedi, C., Van Nostrand, J.D., Zhou, J., Karunaratne, S., Anderson, I.C., Singh, B.K., 2015. Soil aggregate size mediates the impacts of cropping regimes on soil carbon and microbial communities. Soil Biology and Biochemistry 91, 169–181.

Upton, R.N., Bach, E.M., Hofmockel, K.S., 2019. Spatio-temporal microbial community dynamics within soil aggregates. Soil Biology and Biochemistry 132, 58-68.

Vargas, R., Hattori, T., 1986. Protozoan predation of bacterial cells in soil aggregates. FEMS Microbiology Letters 38, 233–242.

Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G.A., 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiology Reviews 37, 936–954.

Wang, L., Luo, X., Liao, H., Chen, W., Wei, D., Cai, P., Huang, Q., 2018. Ureolytic microbial community is modulated by fertilization regimes and particle-size fractions in a black soil of northeastern China. Soil Biology and Biochemistry 116, 171–178.

Wang, Y., Hu, N., Ge, T., Kuzyakov, Y., Wang, Z.-L., Li, Z., Tang, Z., Chen, Y., Wu, C., Lou, Y., 2017. Soil aggregation regulates distributions of carbon, microbial community and enzyme activities after 23-year manure amendment. Applied Soil Ecology 111, 65–72.

Wertz, S., Degrange, V., Prosser, J.I., Poly, F., Commeaux, C., Freitag, T., Guillaumaud, N., Roux, X.L., 2006. Maintenance of soil functioning following erosion of microbial diversity. Environmental Microbiology 8, 2162–2169.

Wilpiszeski, R.L., Aufrecht, J.A., Retterer, S.T., Sullivan, M.B., Graham, D.E., Pierce, E.M., Zablocki, O.D., Palumbo, A.V., Elias, D.A., 2019. Soil aggregate microbial communities: Towards understanding microbiome interactions at biologically relevant scales. Applied and Environmental Microbiology 85, e00324-19.

Wilson, R.M., Tfaily, M.M., Kolton, M., Johnston, E.R., Petro, C., Zalman, C.A., Hanson, P.J., Heyman, H.M., Kyle, J.E., Hoyt, D.W., Eder, E.K., Purvine, S.O., Kolka, R.K., Sebestyen, S.D., Griffiths, N.A., Schadt, C.W., Keller, J.K., Bridgham, S.D., Chanton, J.P., Kostka, J.E., 2021. Soil metabolome response to whole-ecosystem warming at the Spruce and Peatland Responses under Changing Environments experiment. Proceedings of the National Academy of Sciences 118, e2004192118.

Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A.F., Hoelzle, R.D., Lamberton, T.O., McCalley, C.K., Hodgkins, S.B., Wilson, R.M., Purvine, S.O., Nicora, C.D., Li, C., Frolking, S., Chanton, J.P., Crill, P.M., Saleska, S.R., Rich, V.I., Tyson, G.W., 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560, 49–54.

Yang, C., Liu, N., Zhang, Y., 2019. Soil aggregates regulate the impact of soil bacterial and fungal communities on soil respiration. Geoderma 337, 444–452.

Yu, H., Ding, W., Chen, Z., Zhang, H., Luo, J., Bolan, N., 2015. Accumulation of organic C components in soil and aggregates. Scientific Reports 5, 13804.

CHAPTER 1

COMPOST AMENDMENT AND COVER CROPPING MAINTAINS SOIL STRUCTURE AND CARBON STORAGE BY INCREASING AVAILABLE CARBON AND MICROBIAL BIOMASS IN AN AGRICULTURAL SOIL – A SIX-YEAR FIELD STUDY

List of authors: Daoyuan Wang, Jonathan Y. Lin, Jordan M. Sayre, Radomir Schmidt, Steven J. Fonte, Jorge L.M. Rodrigues, Kate M. Scow

Published in *Geoderma*. 2022.
DOI: 10.1016/j.geoderma.2022.116117

ABSTRACT

Soil organic amendments in agricultural production can benefit crop production and a wide range of soil properties, including soil aggregation. Soil aggregate formation is largely driven by microbial activities and can in-turn influence microbial communities by generating distinct microbial habitats, as well as associated impacts on water and nutrient dynamics. We investigated the long-term effects of two fertilizer management strategies (poultry manure compost vs. mineral fertilizer) and biochar amendment (0 vs. 10 t ha^{-1} walnut shell biochar, 900 °C pyrolysis temperature) on soil aggregation, soil organic C, and microbial community dynamics in water-stable aggregate fractions in corn-tomato rotations. Using wet-sieving, soils (0-15 cm) were divided into four size fractions: large macroaggregates (2000-8000 μm), small macroaggregates (250-2000 μm), microaggregates (53-250 μm) and silt and clay (< 53 μm) for calculation of mean weight diameter in both 2014 and 2018. The total C and microbial community composition and abundance within each fraction were evaluated in 2018. Across all treatments, six years of continuous compost application maintained soil aggregate stability and C storage by increasing soil microbial biomass and associated dissolved organic C. Bacterial and fungal populations under compost treatments were significantly higher than under mineral fertilizer treatments based on 16S rRNA gene copy number and internal transcribed spacer (ITS) abundance, which likely contributed to the formation and maintenance of macroaggregates in

compost treatments. Interestingly, continuous application of manure compost may increase microbial available C sources by increasing the abundance of bacteria with the potential to degrade aromatic C as predicted from 16S sequences. Soil under the mineral fertilizer treatment showed decreases in the proportion of large macroaggregates, bulk soil C, and aggregate-associated C storage compared to the compost treatment. The application of highly recalcitrant walnut shell biochar had limited long-term impacts on soil aggregation and C dynamics, likely due to its lack of microbially-available C and limited interaction with the soil environment. Our results indicate that continuous compost inputs maintained soil structure and associated physical stabilization of SOM by enlarging soil microbial available C pool, higher soil microbial biomass, and increasing aggregate formation. The soil aggregate structure, in-turn, generated diverse habitats and altered soil microbial communities. Compost inputs, in addition to or in partial replacement of mineral fertilizer inputs, can provide valuable microbial-driven ecosystem services, such as carbon storage and soil structure, while still providing fertility for crop growth.

INTRODUCTION

Soil aggregation, a frequently overlooked property of agroecosystems, plays many crucial roles: maintaining agricultural productivity, promoting soil C storage, providing habitats for soil biology, and regulating soil water dynamics. Soil water-stable aggregates can contribute to infiltration and water retention (Karami et al., 2012), help control runoff and erosion, and physically protect soil organic matter (SOM) leading to increased soil C storage (Six et al., 2004). Soil aggregate fractions and intra-aggregate pores of different sizes contain distinct physicochemical properties and thus can provide unique habitats for diverse microbial communities (Bach et al., 2018; Bailey et al., 2013; Davinic et al., 2012). Aggregate structure in agricultural soils is influenced by a series of factors, including soil biota (both microbes and

macrofauna), plant root growth, soil mineralogy and texture, inorganic binding agents, and environmental conditions (Six et al., 2004).

Soil aggregate structure can help stabilize microbial communities and enhance interactions between microbes (Kuzyakov and Blagodatskaya, 2015; Raynaud and Nunan, 2014). Meanwhile, soil chemical conditions can drastically change over a short distance in soil (Raynaud and Nunan, 2014). Aggregation contributes to heterogeneity in soils by governing oxygen, water and nutrient availability and shapes microbial communities living inside or at the interface of aggregates (Briar et al., 2011). Furthermore, soil microbial communities can quickly respond to local habitat shifts associated with changes in soil aggregation dynamics (Blaud et al., 2012). A better understanding of the interaction between soil amendments, soil aggregate dynamics, and aggregate-associated soil microbial communities can inform management strategies to enhance soil biological activity and a range of desired soil functions.

Most large-scale agricultural systems have moved away from organic amendments with increasing reliance on synthetic fertilizers. Soil degradation, erosion, and soil organic matter loss are unfortunate consequences from the paucity of organic inputs that are common in conventional agricultural production (Lehman et al., 2015). For this reason, a variety of organic amendments such as compost (Diacono and Montemurro, 2011), animal manures (Mikha and Rice, 2004), and biochar (Atkinson et al., 2010) are being considered once again, and studies comparing their impacts on soils managed with synthetic fertilizer that dominate today are critically needed.

Soil organic amendments can enhance soil C storage through multiple mechanisms; these include increasing soil microbial biomass and activity (Liang et al., 2017), enhancing soil water-stable aggregation (Mpeketula and Snapp, 2019), and introducing recalcitrant C (Smith, 2016). Biochar is typically applied only once every few years (Major, 2010), while the application of manure-based compost tends to be more frequent (Larney and Angers, 2012), both to provide

sufficient macro- and micronutrients for plants (Diacono and Montemurro, 2011) and to support soil microbial communities (Kuzyakov and Blagodatskaya, 2015).

Biochar has received considerable attention regarding its potential to increase soil carbon pools (Smith, 2016). Among biochars produced from various feedstocks under different pyrolysis conditions, highly recalcitrant biochar was considered to have the greatest potential to increase stable soil carbon pool due to its long half-life (Leng and Huang, 2018). On the other hand, biochar has been observed to have a priming effect in both lab and field-scale studies, which may generate soil native organic matter loss (Zimmerman and Ouyang, 2019). Under field conditions, biochar particles experience a decrease in their bioavailablity through association with clay particles and soil organic matter (Lehmann and Joseph, 2015). However, biochar surface functional groups may become exposed during its aging, which would then increase biochar-microbe interactions (Wang et al., 2020). Unlike highly recalcitrant biochar, compost and cover crops contain a higher variety and concentration of biologically-available nutrients that can be metabolized by microbes (Diacono and Montemurro, 2010). Soil water-stable aggregate dynamics driven largely by soil microbes play an essential role in soil C sequestration (Joseph C Blankinship et al., 2016). In a meta-analysis, Islam et al. (2021) reported that biochar properties, soil and environmental conditions all contributed to the impact of biochar on soil aggregation, while wood-sourced and high-temperature biochar were found to have the greatest effect on aggregation. The long-term effects of the combined application of compost and biochar on soil water-stable aggregate dynamics and C sequestration under field conditions remain understudied, especially in annual Mediteranean agroecosystems

The objective of our study was to investigate the effects of two fertilizer treatments-poultry manure compost or mineral fertilizer-and biochar amendment (with or without application), on soil water-stable soil aggregation, aggregate-associated C storage, and soil microbial community composition and abundance. We hypothesized that: (1) multi-season

continuous and diverse C input in the form of compost will increase soil C storage, (2) one-time highly recalcitrant biochar-C input will not significantly increase soil C storage, (3) multiple and continuous compost addition will increase soil microbial biomass in bulk soil and alter microbial community composition across aggregate fractions, and (4) compost amendments will increase soil aggregation and associated soil C storage compared to unamended soil receiving only mineral fertilizer. To test these hypotheses, we compared the impacts of different fertility management practices and biochar amendment in a 6-year tilled, row crop field trial in northern California, representative of an annual Mediterranean agroecosystem.

MATERIALS AND METHODS

Long-term field trial setup

The field site is located at the Russell Ranch Sustainable Agricultural Research Facility, University of California, Davis (Davis, California, USA; 38° 32′ 47″ N 121° 52′ 28″ W). The region has a Mediterranean climate characterized by dry arid summers and wet winters. The soil is a Rincon silty clay loam (fine, smectitic, thermic Mollic Haploxeralfs, 20% sand, 49% silt, and 31% clay; 11 g C kg-1 C content; 1.30 g cm-3 bulk density).

A field experiment was initiated in May of 2012 to investigate the impacts of soil amendments (poultry manure compost, biochar, and mineral fertilizer) on soil aggregation, C storage, and microbial communities in aggregate size fractions. The field was kept fallow for 10 years before 2012, except for a season of Montezuma oats grown between October 2009 and March 2010. The cropping system consisted of a 2-yr rotation of processing tomatoes (Lycopersicon esculentum Mill.) and corn (Zea mays L.). The farm management was based on practices and equipment similar to that used by local commercial growers. Biochar was applied once to half of the plots at the start of the experiment at a rate of 10 Mg ha-1 and disked to a depth of 15 cm. The applied biochar was derived from walnut shells and produced by Dixon

Ridge Farms in Winters, CA (pyrolysis temperature of 900 °C, with 57.5 m2 g-1 surface area, 40.4% ash content, 33.4 cmol g-1 cation exchange capacity, pH of 9.7, 55.3 wt. % C, 0.47 wt. % N, 0.89 wt. % H, 1.6 wt. % O, 0.64 wt. % PO4-P, 9.32 wt. % K; see Mukome et al. (2013) for details). Additionally, two fertility management systems were tested with equivalent total N inputs, based on either: i) mineral fertilizer (27.6 kg N ha−1 as urea-ammonium-nitrate 32 (UAN-32), 36.2 kg P ha−1 P as phosphorus pentoxide, 17.2 kg K ha−1 as potassium oxide, and 1.7 kg ha−1 of zinc chelate as starter fertilizer applied before planting each season; UAN-32 was applied at a rate of 134.5 kg N ha−1 three weeks after tomato transplanting and at a rate of 207.4 kg N ha−1 at the four-leaf growth stage in corn), or ii) poultry manure compost (8.97 Mg ha-1 applied yearly, adding on average 225.4 kg ha−1 total N, 119.5 kg ha−1 total P, and 155.4 kg ha−1 total K; including an incorporated winter cover crop for the first four years) (Griffin et al., 2017). This resulted in four treatments: 1) mineral fertilizer without biochar; 2) mineral fertilizer with biochar; 3) compost without biochar; and 4) compost with biochar, arranged in a randomized complete block design with four replicate blocks per treatment and one treatment replicate per block, making a total of 16 plots (**Figure S2-1**). Each replicate plot was 4.6 m wide and 50 m long.

The average annual above-ground C input in the mineral fertilizer treatment without biochar was 4.30 Mg C ha-1 year-1 as crop residue. Based on calculations for the adjacent Century Experiment at Russell Ranch that includes crop rotations with identical compost and cover crop management, the compost without biochar treatment received an average annual input of 7.27 Mg C ha-1 year-1 for the first 4 years and 6.52 Mg C ha-1 year-1 for the remaining duration of the experiment, of which approximately 2.22 Mg C ha-1 year-1 was accounted for by the compost amendment (Tautges et al., 2019). The biochar-amended treatments both received 5.53 Mg C ha-1 as biochar-C (only in Year 1) in addition to the carbon inputs above.

Soil water-stable aggregate analysis

Water-stable aggregates were separated by size using a wet-sieving method adapted from (Elliott, 1986). In March 2014 and March 2018, three soil sub-samples were taken to a depth 15 cm from each field replicate using a soil knife and combined into a single representative soil sample. The field-moist soils were passed through an 8 mm sieve by gently breaking the soil clods by hand along natural planes of weakness. A 50 g sample of the moist, 8 mm sieved soil was then submerged in deionized water (at room temperature) on top of a 2000 µm sieve for 5 min. The sieve was then moved up and down (~3 cm) for 2 min (50 repetitions min-1). The soil and water passing through the sieve were transferred by gently rinsing the material with deionized water onto the next smaller size sieve, and the same sieving procedure was repeated. Three sieve sizes (2000 µm, 250 µm and 53 µm) were used to generate four aggregate size fractions: 1) > 2000 µm (large macroaggregates); 2) 250-2000 µm (small macroaggregates); 3) 53-250 µm (microaggregates); 4) < 53 µm (silt and clay fraction). Two independent rounds of sieving were performed. First, one set of samples were obtained to quantify water-stable aggregates and conduct physicochemical analyses, in which all the aggregate fractions retained on each sieve were rinsed off the sieve in pre-weighed aluminum pans, oven-dried at 60 °C, and then weighed. The other set of samples was obtained for analysis of microbial community composition and abundance. The large and small macroaggregates and microaggregates retained on each sieve were rinsed off the sieve into sterile 50 mL polypropylene tubes. The silt and clay fraction was allowed to settle for a few minutes, and then subsamples of both sediment and supernatant were collected in a 50 mL sterile tube. All the aggregate size fractions were immediately stored at -80 °C until DNA extraction.

Mean weight diameter (MWD), an index of aggregate stability based on a weighted average of the four aggregate size classes, was calculated according to the following equation (Van Bavel, 1950):

$$MWD = \sum_{i=1}^{4} P_i * S_i \quad (1)$$

where P_i is the weight percentage of the fraction in the whole soil and S_i is the average diameter (μm) for particles in its fraction.

Soil C content

The C content of bulk soils and of each aggregate size fraction was analyzed using a PDZ Europa ANCA-GSL elemental analyzer interfaced to a PDZ Europa 20-20 isotope ratio mass spectrometer (Sercon Ltd., Cheshire, UK) at the UC Davis Stable Isotope Facility.

Soil dissolved organic C and microbial biomass C

The fresh soil collected in March 2018 was also used to evaluate the impacts of different soil management treatments on soil dissolved organic C and soil microbial biomass-C. A representative bulk soil sample (8 g) was mixed with 40 mL 0.5 mol L-1 potassium sulfate in polypropylene tubes and placed on an orbital shaker (250 rev min-1, 30 min). After shaking, samples were centrifuged (relative centrifugal force of 7969 × g for 15 min) to remove suspended solids. Supernatant solutions were retained for dissolved organic C concentrations (mg L-1). The total microbial biomass-C was measured by chloroform fumigation (Joergensen, 1996; Yang et al., 2016). Both soil dissolved organic C and microbial biomass-C were measured using a TOC analyzer (Shimadzu TOC-VCSH analyzer, Kyoto, KYT, Japan).

Soil DNA extraction from aggregate fractions and amplicon sequencing

DNA from each soil aggregate size fraction was extracted using the Powerlyzer PowerSoil DNA Isolation kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions. The extracted DNA was quantified using the Qubit dsDNA HS Assay kit (Life Technologies, Carlsbad, CA, USA). The V4 hypervariable region of the bacterial 16S rRNA gene was amplified

from each sample in duplicate using the primer pair 505F/816R (Caporaso et al., 2012), which was designed to include Illumina adaptors and 12 bp barcode sequences. The resulting amplicons were inspected by gel electrophoresis on a 1% agarose gel, quantified by fluorimetry as above, pooled in equimolar concentrations, and sequenced on an Illumina MiSeq platform (paired-end 250 bp) at the UC Davis DNA Technologies core facility. The raw reads were processed using DADA2 (Callahan et al., 2016) implemented in R v.3.4.4. Briefly, paired-end fastq files were processed by quality-trimming forward and reverse reads to 190 and 150 bp lengths, respectively. After sequence dereplication, merging, error-correction, and chimera removal, Exact Sequence Variants (ESVs) were inferred and taxonomy was assigned using the SILVA database v. 132. After quality control, the number of sequences per sample varied from 14,596 to 44,786, with an average of 28,724. The resulting ESV abundance table was rarified to 14,000 sequences per sample to ensure equal sampling depth for statistical analysis. The Functional Annotation of Prokaryotic Taxa (FAPROTAX) pipeline (Louca et al., 2016) was used to predict the functional potential of bacterial taxa identified in our dataset. Raw sequences were deposited at the NCBI sequence read archive (SRA) under BioProject accession number PRJNA644905.

Quantitative PCR

To assess prokaryotic and fungal gene copy number as a proxy for absolute abundance, quantitative PCR (qPCR) was performed on each DNA sample using the universal primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') for the 16S rRNA gene (Rubin et al., 2014) and ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3') and ITS2 (5'-GCTGCGTTCTTCATCGATGC-3') for the fungal Internal Transcribed Spacer (ITS) region (De Beeck et al., 2014). qPCR was performed in 20 µL reaction mixtures containing 10 µL SsoAdvanced Universal SYBR Green Supermix (Biorad, Hercules, CA, USA), 0.5 µM each primer,

and 10 ng sample DNA. Reactions were carried out on a BioRad CFX Connect System (Biorad, Hercules, CA, USA) and amplification of the 16S rRNA gene consisted of an initial denaturation of 95 °C for 3 min, followed by 39 cycles of 95 °C for 10 s and 60 °C for 30 s. Amplification of the ITS region consisted of an initial denaturation of 95 °C for 3 min, followed by 39 cycles of 95 °C for 10 s, 50 °C for 30 s, and 72 °C for 10 s. Quantification was performed by comparing unknown samples to a standard curve (ranging from 102-109 copies for 16S rRNA; 101-108 copies for ITS) generated with the pCR Blunt II-TOPO vector (Invitrogen, Carlsbad, CA, USA) containing a PCR-amplified fragment of each target. R2 values for the standard curves ranged from 0.982-0.986 and 0.991-0.994 for the 16S rRNA gene and ITS region, respectively. Triplicate reactions were performed for each target per sample, and a melting curve analysis was performed after each assay to ensure specificity of the amplified products. The abundance of total 16S rRNA and ITS were normalized as copies per gram of soil aggregate fraction.

16S rRNA and ITS abundances of the bulk soil were estimated based on soil aggregation and copy numbers according to the following equation:

$$Gene\ abundance\ in\ bulk\ soil = \sum_{i=1}^{4} P_i * Q_i \quad (2)$$

where Q_i is the copy number of the target for each soil aggregate fraction, and P_i is the weight percentage of the soil aggregate fraction in the whole soil, respectively, as above.

Statistical analyses

All soil physicochemical data were analyzed in Microsoft Excel for Windows 2010 with XLSTAT Version 2019.1 (Addinsoft, 2019) and were tested for assumptions of normality and homogeneity of variance. Statistically significant differences between treatments were analyzed using a mixed model analysis of variance (ANOVA) with biochar and fertility management considered as fixed effects and block was a random effect followed by a Tukey's range test.

Non-metric multidimensional scaling (NMDS) plots were generated based on Bray-Curtis dissimilarities to visualize differences in bacterial community composition from our 16S rRNA sequences. Differences in bacterial community composition were tested by permutational multivariate analysis of variance (PERMANOVA) with the 'adonis' function (Oksanen et al., 2007) in R v.3.4.4 using Bray-Curtis dissimilarities and management practice, biochar treatment, and aggregate size fraction as predictor variables. All other microbial data, including FAPROTAX counts and qPCR values were tested for assumptions of normality and homogeneity of variance before performing ANOVA and a Tukey HSD post-hoc test to identify significant differences between treatments, with biochar and fertility management considered as main effects, block included as a random variable and aggregate fractions nested within each plot replicate. Natural log transformations were applied to meet the assumptions of ANOVA (normality, homoscedasticity) when necessary. For all analyses, statistically significant differences were defined at $P < 0.05$.

RESULTS

Soil water-stable aggregates

Compost amended soil had significantly higher aggregate stability compared to mineral fertilizer treatments after six years, such that in March 2018, the MWD of soils managed with compost was 140% higher than those receiving mineral fertilizer (**Figure 2-1**). Compost application maintained water-stable aggregation, in which the soil MWD in March 2018 remained the same as that in March 2014, while the mineral fertilizer treatments significantly decreased by ~ 48% over the four years (**Figure 2-1**). The observed loss in aggregation under mineral fertilizer was primarily due to a significant loss in large macroaggregates (**Table 2-1**). Neither fertility management nor biochar amendment had a significant short-term (two years, in March 2014)

effect on soil structure (**Figure 2-1a**). Biochar amendment had no long-term effect on soil MWD under either fertilizer treatment (**Figure 2-1b**).

Soil C content and C in aggregate fractions

Compost application resulted in significantly higher bulk soil C content compared to the mineral fertilizer treatment. After six years of compost amendment in 2018, the C content in the top 15 cm of soil was 12.4 g C kg-1whole soil, which was 17% higher than the mineral fertilizer treatment (**Figure 2-2**) and due largely to an increase in the large macroaggregate associated C (**Table 2-2**). This was in contrast to early in the trial, where two years of compost addition did not significantly affect bulk soil C content in 2014 (**Figure 2-2a**). Surprisingly, the 10 Mg ha-1 biochar amendment had no long-term effect on soil C content or C distribution across aggregate fractions under either fertility management practice after six years (**Figure 2-2** and **Table 2-2**).

Soil dissolved organic C and microbial biomass C

Compost addition significantly increased both soil dissolved organic C and soil microbial biomass C compared to the mineral fertilizer treatment. Soil dissolved organic C contents in compost treatments were 174.6 mg C kg-1 in the whole soil after 6 years, which was 54% higher than the mineral fertilizer treatment (**Figure 2-3a**). Similar differences were also observed in soil microbial biomass C, which was 315.2 mg C kg-1 whole soil in compost, approximately two times as in mineral fertilizer treatments (**Figure 2-3b**). Similar to total soil C, dissolved organic C and microbial biomass C were not influenced by an initial application of biochar under either treatment in 2018.

Soil microbial community composition and abundance in water-stable aggregate fractions

Compost management significantly increased both bacterial 16S rRNA gene and fungal ITS abundances across soil water-stable aggregate fractions, while biochar had no effect in year

2018 (**Figures 2-4 and 2-5**). The weighted average gene abundance in bulk soil also indicated that the abundance of the 16S rRNA gene was 76% higher under compost than in soils receiving mineral fertilizer (**Figure 2-6a**). The weighted average ITS copy number in bulk soil under compost addition was two orders of magnitudes higher than in soil under mineral fertilizer treatment (**Figure 2-6b**). Nonmetric multidimensional scaling (NMDS) analysis of the 16s rRNA gene at the ESV level revealed that soil prokaryotic community composition was significantly distinct in aggregate fractions of different sizes (**Figure 2-7**). These differences were reflected by higher proportions of bacteria from the orders *Micrococcales*, *Streptomycetales*, *Propionibacteriales*, and *Sphingomonadales* in the microaggregate and silt and clay fractions, and higher proportions of the *Gaiellales*, *Gemmatimonadales*, *Nitrososphaerales*, and Nitrospirales in the large and small aggregate fractions. Across all aggregate size fractions, both compost and mineral fertilizer treatments had significant but limited effects on bacterial community composition, whereas biochar had no effect (**Figure S2-2**).

Predictive assignment of soil microbial functions based on the 16S rRNA gene sequencing results revealed that the abundance of bacteria potentially capable of degrading aromatic C compounds was higher in compost than mineral fertilizer treatments in 2018 (**Figure 2-8**). No differences were detected for predicted functions related to many major C (methanotrophy, methylotrophy, chitinolysis, cellulolysis, xylanolysis, non-methane aliphatic hydrocarbon degradation, hydrocarbon degradation) and N (aerobic ammonia oxidation, aerobic nitrite oxidation, nitrification, nitrate reduction, ureolysis) cycling pathways between the two treatments.

DISCUSSION

Impacts of external C inputs on soil C storage

Our findings suggest that continuous compost application is more effective at increasing soil C sequestration by maintaining soil structure and stabilizing SOM than in soils not receiving compost. In contrast, without compost inputs, there were significant decreases in aggregate structure, which in turn was associated with significant C loss. Poultry manure compost is rich in multivalent ions, such as Ca and P (Griffin et al., 2017), which can generate a bridging effect to enhance sorption of SOM to clay minerals (Feng et al., 2005) and increase soil aggregation (Bronick and Lal, 2005). The lack of sufficient C input and associated aggregate structure in the mineral fertilizer treatments could lead to greater exposure and more rapid decay of native SOM (Dungait et al., 2012).

Surprisingly, walnut shell biochar had little effect on soil C in the top 15 cm. Previous research has hypothesized that recalcitrant C compounds are the major contributors to C sequestration from biochar (Cheng et al., 2008; Harvey et al., 2012), especially high-temperature biochars like the type applied in our study, which has a high proportion of recalcitrant to labile C (Zimmerman, 2010). The O/C and H/C atom ratios of the walnut shell biochar were 0.0217 and 0.193, respectively (Mukome et al., 2013), which indicated that the walnut shell biochar-C was relatively stable and potentially had a longer than 1000-year half-life (Spokas, 2010). A 14-month lab incubation study using similar soil and the same biochar revealed that this walnut shell biochar (with application rates equivalent or doubled as in the field study), can enhance soil C storage under lab conditions (Wang et al., 2017), but no evidence of this was observed in our field study. Additionally, we speculate that in the field, some biochar particles may migrate from the point of application and leave the surface soil through irrigation, wind erosion (Gelardi et al., 2019), or vertical movement (Singh et al., 2015), thus decreasing their impact on soil properties. The walnut shell biochar contains some fine particles, which can be readily mobilized during irrigation events in the first growing season. We observed some fine biochar particles in the surface runoff during furrow irrigation at the beginning of the field trial. However, a large amount

of biochar particles (with a diameter of around 2-4 mm) was easily observed and recovered in the surface soil at the end of the field trial.

Interactions between agricultural management, soil structure and soil aggregate-associated microbial communities

Our findings suggest that continuous compost addition can potentially generate a positive feedback loop for enhancing soil C storage by increasing microbially-available substrates, in turn increasing and maintaining soil microbial biomass, and thereby increasing aggregation through microbial activities. Soils received ~60% more C in the compost-amended than mineral fertilizer treatments. Compost also contains a diverse range of C sources for soil microbes (Barker, 1997), and its addition helped maintain higher soil labile C (**Figure 2-3a**) and microbial biomass (**Figure 2-3b**) compared to in the mineral fertilizer treatments. The compost-induced increase in soil microbial biomass also corroborated our findings of increased 16S rRNA gene and ITS copy numbers across aggregate size fractions (**Figure 2-4 and 2-5**) and in weighted bulk soil averages (**Figure 2-6**). Compost amendment enhanced soil aggregation by providing a large amount of labile C (Amlinger et al., 2003) as feedstock for microbes to produce extracellular polymeric substances, which can serve as binding agents for soil aggregates (J C Blankinship et al., 2016; Miltner et al., 2012). The increased fungal biomass (**Figures 2-5 and 2-6**) can also promote soil aggregation by binding and entangling soil particles to form macroaggregate structure (Van Der Heijden et al., 2006). Enhanced soil aggregation provides more diverse habitats for organisms, which can lead to increases in soil microbial diversity (Briar et al., 2011). Furthermore, the soil aggregate structure can enhance the interactions within microbial consortia responsible for metabolizing complex organic compounds (Wilpiszeski et al., 2019). The shift we observed in bacterial communities in different aggregate size fractions (**Figure 2-7**) provides evidence for the paradigm that unique aggregate microenvironments can

provide niches to support the development of distinct microbial communities, which can further benefit soil and ecosystem properties (Bach et al., 2018; Wilpiszeski et al., 2019).

Across all aggregate size fractions, we also observed distinct differences in prokaryotic community compositions between compost and mineral fertilizer treatments (**Figure S2-2**). Previous research investigating a similar soil (Yolo silt loam) from a field adjacent to this trial indicated that the dissolved organic matter from the soil in our field trial was highly aromatic (Wang et al., 2016). Our FAPROTAX predictive results also indicated a higher abundance of bacteria potentially capable of degrading aromatic C compounds in compost treatments (**Figure 2-8**), which suggests that microbial communities under continuous compost application are potentially capable of utilizing a wider range of C sources from both soil amendments and native soil organic matter, which further enriched soil microbially-available nutrients. Such feedback can potentially enlarge the active C pool in the agroecosystem which, in turn, could maintain a higher level of soil aggregation associated microbial activities. Other soil C and N cycling pathways did not show differences since the sampling time was 300 days after the yearly compost amendment.

Compost application may have increased macroaggregate formation by maintaining higher fungal populations across aggregate fractions. Although both bacteria and fungi populations were higher in the compost than mineral fertilizer treatments (**Figure 2-5 and 2-6**), the difference in fungal abundance was two orders of magnitudes higher in compost than mineral fertilizer treatments (**Figure 2-6b**). Higher fungal biomass density has been linked with increased soil aggregate formation (Lehmann et al., 2020) and fungi have been suggested to contribute more than prokaryotic communities to macroaggregate formation due to the enmeshing properties of their hyphae (Lehmann et al., 2017). The major difference in aggregation between the mineral fertilizer and compost treatments was due to the loss of large-macroaggregates under mineral fertilizer (**Table 2-1**), which suggested that the differences in fungal community

abundance could have been a major contributor to the observed differences in soil structure. Similar to our finding, Li et al. (2019) in a long-term field study in Guizhou, China, reported that soil fungal abundances across aggregate fractions were maintained when mineral fertilizer was supplemented with manure but decreased in the absence of amendments; aggregate-associated C content also declined. Increased macroaggregate structure in compost treatments can, in turn, enhance the resistance of soil microbial habitats to environmental disturbances (Rillig et al., 2017). Our results support the hypothesis that impacts of management practices on soil C are dependent on changes in soil aggregation and aggregate-associated microbial communities (Trivedi et al., 2017). As shown in **Table S2-1** and summarized in Griffin et al. (2017), despite the benefits associated with soil aggregation and C dynamics, crop yield in compost treatments were significantly lower than those in mineral fertilizer treatments due to the uncertainty of nitrogen availability in compost amended soil and the asynchronous nitrogen supply and demand, which is a common tradeoff for similar practices (Seufert et al., 2012).

Interstingly, we found no effect of walnut shell biochar on soil microbial community composition or any soil parameters in our study, unlike what has been observed in previous studies investigating at other types of biochar (Jiang et al., 2016; Khodadad et al., 2011; Zhu et al., 2019). This may be due to the differences in biochar feedstock, pyrolysis conditions, soil and other environmental factors, or the fact that biochar under field conditions behaves differently than in the lab (Islam et al., 2021). It is possible that biochars produced at high temperatures (900 °C) contain little labile organic C compared to low temperature biochar and hence are not capable of supporting growth of microbial populations. Another explanation for the lack of impact may be the limited accessibility of biochar pores to microbes, despite its relatively high surface area and pore volume (Mukome et al., 2013). The interparticle pore structure in biochar is rapidly filled by soil particles after addition to soil (Lehmann and Joseph, 2015). Joseph et al. (2010) found that biochar internal pores started to fill in with organic and mineral matter after 1

year and most pores were filled 2 years after application in a field experiment. Ameloot et al. (2014) also found that biochar stopped serving as microbial substrate soon after application (~60 days) and had little long-term impact on soil microbial biomass and activities under field conditions. While previous studies have shown that biochar can alter soil microbial communities and potentially increase microbial interactions, most of these studies have been conducted in highly weathered oxisols with limited impacts observed in soils with high fertility (Yu et al., 2018). We speculate that soil amendments with higher amounts of non-pyrolyzed and/or microbially-available organic C can help achieve agricultural management targets, such as enhancing soil aggregation and C storage more effectively.

CONCLUSION

Our findings indicate that continuous compost amendment can potentially generate a positive feedback loop for soil aggregate formation and associated C storage by maintaining higher dissolved organic C content, increasing microbial biomass, and supporting large-macroaggregate formation. These processes, in turn, promote the ability of the microbial community to utilize more diverse C sources. In fine-textured soil, biochar had a limited impact on the soil microbial community, soil aggregation, and C dynamics, likely due to low microbial available C and limited interaction with the environment. Long-term continuous diversified C source amendment, such as adding compost, could be an effective agricultural management practice to not only maintain but increase soil microbial biomass, aggregate formation, and soil C storage, and also replace some mineral fertilizer which could reduce greenhouse gas emissions associated with fertilizer production. An integrated management practice that carefully balances compost and mineral fertilizer composition can potentially balance the benefits and tradeoffs.

ACKNOWLEDGEMENTS

We thank Israel Herrera, Deirdre Griffin-LaHue, Nicole Tautges, members of the Scow Soil Microbial Ecology Lab, and staff at Russell Ranch Sustainable Agriculture Facility in University of California, Davis, for their valuable support in our field study. This publication was made possible by the United States Department of Agriculture (USDA), National Institute of Food and Agriculture (NIFA) through Hatch Formula Funding CA 2122-H and multistate regional project W-2082. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the National Institute of Food and Agriculture (NIFA) or the United States Department of Agriculture (USDA).

REFERENCES

Addinsoft, 2019. XLSTAT statistical and data analysis solution.

Ameloot, N., Sleutel, S., Case, S.D.C., Alberti, G., McNamara, N.P., Zavalloni, C., Vervisch, B., Vedove, G. delle, De Neve, S., 2014. C mineralization and microbial activity in four biochar field experiments several years after incorporation. Soil Biology and Biochemistry. 78, 195–203.

Amlinger, F., Götz, B., Dreher, P., Geszti, J., Weissteiner, C., 2003. Nitrogen in biowaste and yard waste compost: dynamics of mobilisation and availability—a review. European Journal of Soil Biology. 39, 107–116.

Atkinson, C.J., Fitzgerald, J.D., Hipps, N.A., 2010. Potential mechanisms for achieving agricultural benefits from biochar application to temperate soils: a review. Plant and Soil 337, 1–18.

Bach, E.M., Williams, R.J., Hargreaves, S.K., Yang, F., Hofmockel, K.S., 2018. Greatest soil microbial diversity found in micro-habitats. Soil Biology and Biochemistry. 118, 217–226.

Bailey, V.L., McCue, L.A., Fansler, S.J., Boyanov, M.I., DeCarlo, F., Kemner, K.M., Konopka, A., 2013. Micrometer-scale physical structure and microbial composition of soil macroaggregates. Soil Biology and Biochemistry. 65, 60–68.

Barker, A. V, 1997. Composition and uses of compost, in: Agricultural uses of by-products and wastes. American Chemical Society, pp. 140–162.

Blankinship, Joseph C, Fonte, S.J., Six, J., Schimel, J.P., 2016. Plant versus microbial controls on soil aggregate stability in a seasonally dry ecosystem. Geoderma 272, 39–50.

Blankinship, J C, Fonte, S.J., Six, J., Schimel, J.P., 2016. Plant versus microbial controls on soil aggregate stability in a seasonally dry ecosystem. Geoderma 272, 39–50.

Blaud, A., Lerch, T.Z., Chevallier, T., Nunan, N., Chenu, C., Brauman, A., 2012. Dynamics of bacterial communities in relation to soil aggregate formation during the decomposition of ^{13}C-labelled rice straw. Applied Soil Ecology. 53, 1–9.

Briar, S.S., Fonte, S.J., Park, I., Six, J., Scow, K., Ferris, H., 2011. The distribution of nematodes and soil microbial communities across soil aggregate fractions and farm management systems. Soil Biology and Biochemistry. 43, 905–914.

Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. Geoderma 124, 3–22.

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

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal. 6, 1621.

Cheng, C., Lehmann, J., Thies, J.E., Burton, S.D., 2008. Stability of black carbon in soils across a climatic gradient. Journal of Geophysical Research: Biogeosciences 113.

Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak, J.C., Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy

reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. Soil Biology and Biochemistry 46, 63–72.

De Beeck, M.O., Lievens, B., Busschaert, P., Declerck, S., Vangronsveld, J., Colpaert, J. V, 2014. Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. PLOS ONE 9, e97629.

Diacono, M., Montemurro, F., 2011. Long-term effects of organic amendments on soil fertility, in: Sustainable Agriculture Volume 2. Springer, pp. 761–786.

Diacono, M., Montemurro, F., 2010. Long-term effects of organic amendments on soil fertility. A review. Agronomy for Sustainable Development 30, 401–422.

Dungait, J.A.J., Hopkins, D.W., Gregory, A.S., Whitmore, A.P., 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. Global Change Biology 18, 1781–1796. https://doi.org/10.1111/j.1365-2486.2012.02665.x

Elliott, E.T., 1986. Aggregate structure and carbon, nitrogen, and phosphorus in native and cultivated soils. Soil Science Society of America Journal 50, 627–633.

Feng, X., Simpson, A.J., Simpson, M.J., 2005. Chemical and mineralogical controls on humic acid sorption to clay mineral surfaces. Organic Geochemistry 36, 1553–1566.

Gelardi, D.L., Li, C., Parikh, S.J., 2019. An emerging environmental concern: Biochar-induced dust emissions and their potentially toxic properties. Science of the Total Environment 678, 813–820.

Griffin, D.E., Wang, D., Parikh, S.J., Scow, K.M., 2017. Short-lived effects of walnut shell biochar on soils and crop yields in a long-term field experiment. Agriculture, Ecosystems & Environment 236.

Harvey, O.R., Kuo, L.-J., Zimmerman, A.R., Louchouarn, P., Amonette, J.E., Herbert, B.E., 2012. An index-based approach to assessing recalcitrance and soil carbon sequestration potential of engineered black carbons (biochars). Environmental Science and Technology 46, 1415–1421.

Islam, M.U., Jiang, F., Guo, Z., Peng, X., 2021. Does biochar application improve soil aggregation? A meta-analysis. Soil and Tillage Research 209, 104926.

Jiang, X., Denef, K., Stewart, C.E., Cotrufo, M.F., 2016. Controls and dynamics of biochar decomposition and soil microbial abundance, composition, and carbon use efficiency during long-term biochar-amended soil incubations. Biology and Fertility of Soils 52, 1–14.

Joergensen, R.G., 1996. The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEC value. Soil Biology and Biochemistry. 28, 25–31.

Joseph, S.D., Camps-Arbestain, M., Lin, Y., Munroe, P., Chia, C.H., Hook, J., van Zwieten, L., Kimber, S., Cowie, A., Singh, B.P., Lehmann, J., Foidl, N., Smernik, R.J., Amonette, J.E., 2010. An investigation into the reactions of biochar in soil. Australian Journal of Soil Research. 48, 501–515.

Karami, A., Homaee, M., Afzalinia, S., Ruhipour, H., Basirat, S., 2012. Organic resource management: Impacts on soil aggregate stability and other soil physico-chemical properties. Agriculture, Ecosystems & Environment 148, 22–28.

Khodadad, C.L.M., Zimmerman, A.R., Green, S.J., Uthandi, S., Foster, J.S., 2011. Taxa-specific changes in soil microbial community composition induced by pyrogenic carbon amendments. Soil Biology and Biochemistry. 43, 385–392.

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry. 83, 184–199.

Larney, F.J., Angers, D.A., 2012. The role of organic amendments in soil reclamation: A review. Canadian Journal of Soil Science. 92, 19–38.

Lehman, R.M., Cambardella, C.A., Stott, D.E., Acosta-Martinez, V., Manter, D.K., Buyer, J.S., Maul, J.E., Smith, J.L., Collins, H.P., Halvorson, J.J., 2015. Understanding and enhancing soil biological health: the solution for reversing soil degradation. Sustainability 7, 988–1027.

Lehmann, A., Zheng, W., Rillig, M.C., 2017. Soil biota contributions to soil aggregation. Nature Ecology and Evolution. 1, 1828–1835.

Lehmann, A., Zheng, W., Ryo, M., Soutschek, K., Roy, J., Rongstock, R., Maaß, S., Rillig, M.C., 2020. Fungal traits important for soil aggregation. Frontiers in Microbiology 10.

Lehmann, J., Joseph, S., 2015. Biochar for environmental management: Science, technology and implementation. Routledge.

Leng, L., Huang, H., 2018. An overview of the effect of pyrolysis process parameters on biochar stability. Bioresource Technology 270, 627–642.

Li, F., Qiu, P., Shen, B., Shen, Q., 2019. Soil aggregate size modifies the impacts of fertilization on microbial communities. Geoderma 343, 205–214.

Liang, C., Schimel, J.P., Jastrow, J.D., 2017. The importance of anabolism in microbial control over soil carbon storage. Nature Microbiology 2, 17105.

Louca, S., Parfrey, L.W., Doebeli, M., 2016. Decoupling function and taxonomy in the global ocean microbiome. Science 353, 1272.

Major, J., 2010. Guidelines on practical aspects of biochar application to field soil in various soil management systems. International Biochar Initiative 8.

Mikha, M.M., Rice, C.W., 2004. Tillage and manure effects on soil and aggregate-associated carbon and nitrogen contribution. Soil Science Society of America Journal 68, 809–816.

Miltner, A., Bombach, P., Schmidt-Brücken, B., Kästner, M., 2012. SOM genesis: Microbial biomass as a significant source. Biogeochemistry 111, 41–55.

Mpeketula, P.M.G., Snapp, S.S., 2019. Structural stability conditions soil carbon gains from compost management and rotational diversity. Soil Science Society of America Journal 83, 203–211.

Mukome, F.N.D., Zhang, X., Silva, L.C.R., Six, J., Parikh, S.J., 2013. Use of chemical and physical characteristics to investigate trends in biochar feedstocks. Journal of Agricultural and Food Chemistry 61, 2196–2204.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., Oksanen, M.J., Suggests, M., 2007. Vegan: Community ecology package.

Raynaud, X., Nunan, N., 2014. Spatial ecology of bacteria at the microscale in soil. PLOS ONE 9, e87217.

Rillig, M.C., Muller, L.A.H., Lehmann, A., 2017. Soil aggregates as massively concurrent evolutionary incubators. The ISME Journal 11, 1943–1948.

Rubin, B.E.R., Sanders, J.G., Hampton‐Marcell, J., Owens, S.M., Gilbert, J.A., Moreau, C.S., 2014. DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure. Microbiologyopen 3, 910–921.

Seufert, V., Ramankutty, N., Foley, J.A., 2012. Comparing the yields of organic and conventional agriculture. Nature 485, 229–232.

Singh, B.P., Fang, Y., Boersma, M., Collins, D., Van Zwieten, L., Macdonald, L.M., 2015. *In situ* persistence and migration of biochar carbon and its impact on native carbon emission in contrasting soils under managed temperate pastures. PLOS ONE 10.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil and Tillage Research. 79, 7–31.

Smith, P., 2016. Soil carbon sequestration and biochar as negative emission technologies. Global Change Biology 22, 1315–1324.

Spokas, K.A., 2010. Review of the stability of biochar in soils: Predictability of O:C molar ratios. Carbon Management 1, 289–303.

Tautges, N.E., Chiartas, J.L., Gaudin, A.C.M., O'Geen, A.T., Herrera, I., Scow, K.M., 2019. Deep soil inventories reveal that impacts of cover crops and compost on soil carbon sequestration differ in surface and subsurface soils. Global Change Biology 25, 3753–3766.

Trivedi, P., Delgado-Baquerizo, M., Jeffries, T.C., Trivedi, C., Anderson, I.C., Lai, K., McNee, M., Flower, K., Pal Singh, B., Minkey, D., Singh, B.K., 2017. Soil aggregation and associated microbial communities modify the impact of agricultural management on carbon content. Environmental Microbiology. 19, 3070–3086.

Van Bavel, C.H.M., 1950. Mean weight-diameter of soil aggregates as a statistical index of aggregation 1. Soil Science Society of America Journal 14, 20–23.

Van Der Heijden, M.G.A., Streitwolf‐Engel, R., Riedl, R., Siegrist, S., Neudecker, A., Ineichen, K., Boller, T., Wiemken, A., Sanders, I.R., 2006. The mycorrhizal contribution to plant productivity, plant nutrition and soil structure in experimental grassland. New Phytologist 172, 739–752.

Wang, D., Fonte, S.J., Parikh, S.J., Six, J., Scow, K.M., 2017. Biochar additions can enhance soil structure and the physical stabilization of C in aggregates. Geoderma 303, 110-117.

Wang, D., Griffin, D.E., Parikh, S.J., Scow, K.M., 2016. Impact of biochar amendment on soil water soluble carbon in the context of extreme hydrological events. Chemosphere 160, 287-292.

Wang, L., O'Connor, D., Rinklebe, J., Ok, Y.S., Tsang, D.C.W., Shen, Z., Hou, D., 2020. Biochar aging: Mechanisms, physicochemical changes, assessment, and implications for field applications. Environment Science and Technology. 54, 14797–14814.

Wilpiszeski, R.L., Aufrecht, J.A., Retterer, S.T., Sullivan, M.B., Graham, D.E., Pierce, E.M., Zablocki, O.D., Palumbo, A. V, Elias, D.A., 2019. Soil aggregate microbial communities:

Towards understanding microbiome interactions at biologically relevant scales. Applied and Environmental Microbiology 85, e00324-19.

Yang, L., Zhang, L., Geisseler, D., Wu, Z., Gong, P., Xue, Y., Yu, C., Juan, Y., Horwath, W.R., 2016. Available C and N affect the utilization of glycine by soil microorganisms. Geoderma 283, 32–38.

Yu, J., Deem, L.M., Crow, S.E., Deenik, J.L., Penton, C.R., 2018. Biochar application influences microbial assemblage complexity and composition due to soil and bioenergy crop type interactions. Soil Biology and Biochemistry. 117, 97–107.

Zhu, X., Mao, L., Chen, B., 2019. Driving forces linking microbial community structure and functions to enhanced carbon stability in biochar-amended soil. Environment International 133, 105211.

Zimmerman, A.R., 2010. Abiotic and microbial oxidation of laboratory-produced black carbon (biochar). Environmental Science and Technology 44, 1295–1301.

Zimmerman, A.R., Ouyang, L., 2019. Priming of pyrogenic C (biochar) mineralization by dissolved organic matter and vice versa. Soil Biology and Biochemistry. 130, 105–112.

TABLES AND FIGURES

Table 2-1. Distribution of water stable aggregate size fractions (% of whole soil mass) under mineral fertilizer and compost management, with and without biochar amendment (0 and 10 Mg ha^{-1}) after 2 (March 2014) and 6 years (March 2018). The numbers to the right of each value represent the standard error of the mean. Significant differences based on Tukey test between two different time points are indicated by different letters in parentheses to the right of each value.

Time	Treatment	**Large macroaggregates (2000-8000 µm)**	**Small macroaggregates (250-2000 µm)**	**Microaggregates (53-250 µm)**	**Silt and clay (< 53 µm)**
		% of whole soil mass			
March 2014	**Mineral Fertilizer**	26.56 ± 2.39 (a)	33.52 ± 3.24 (a)	17.95 ± 1.15 (a)	21.97 ± 4.84 (b)
	Mineral Fertilizer + Biochar	26.87 ± 2.69 (a)	31.6 ± 3.78 (a)	18.01 ± 0.72 (a)	23.53 ± 2.55 (b)
	Compost	28.11 ± 1.68 (a)	37.61 ± 3.27 (ab)	19.92 ± 0.94 (b)	14.36 ± 2.39 (a)
	Compost + Biochar	25.39 ± 3.58 (a)	41.38 ± 2.59 (b)	21.06 ± 1.39 (b)	12.17 ± 1.91 (a)
	ANOVA	*P*-values			
	Fertilizer	0.985	**0.019**	**0.010**	**0.004**
	Biochar	0.552	0.646	0.530	0.879
	Fertilizer x Biochar	0.211	0.292	0.286	0.976
March 2018	**Mineral Fertilizer**	6.40 ± 3.75 (a)	46.80 ± 1.74 (a)	33.80 ± 3.36 (b)	13.01 ± 1.86 (ab)
	Mineral Fertilizer + Biochar	3.92 ± 1.34 (a)	44.02 ± 2.66 (a)	37.72 ± 4.11 (b)	14.33 ± 0.38 (b)
	Compost	29.80 ± 6.64 (b)	41.01 ± 4.95 (a)	18.53 ± 1.85 (a)	10.66 ± 1.12 (a)
	Compost + Biochar	30.27 ± 10.32 (b)	39.29 ± 5.13 (a)	19.98 ± 4.59 (a)	10.46 ± 1.59 (a)
	ANOVA	*P*-values			
	Fertilizer	**0.003**	**0.039**	**0.002**	**0.034**
	Biochar	0.822	0.484	0.187	0.279
	Fertilizer x Biochar	0.689	0.804	0.550	0.410

Table 2-2. Soil C distribution in soil aggregate fractions (g C kg^{-1} whole soil) after 6 years (March 2018) under mineral fertilizer and compost management, with and without biochar amendment (0 and 10 Mg ha^{-1}). The numbers to the right of each value represent the standard error of the mean. Significant differences based on Tukey test between all four treatments are indicated by different letters in parentheses to the right of each value.

Fertilizer treatment	**Biochar application rate** (Mg ha^{-1})	**Large macroaggregate C** (2000-8000 µm)	**Small macroaggregate C** (250-2000 µm)	**Microaggregate C** (53-250 µm)	**Silt and clay C** (< 53 µm)
		--- g C kg^{-1} whole soil ---			
Mineral fertilizer	0	0.80 ± 0.40 (a)	5.10 ± 0.15 (a)	2.88 ± 0.28 (bc)	1.61 ± 0.23 (ab)
	10	0.55 ± 0.17 (a)	4.92 ± 0.46 (a)	3.27 ± 0.41 (c)	1.91 ± 0.25 (b)
Compost	0	3.64 ± 0.82 (b)	5.26 ± 0.58 (a)	1.88 ± 0.14 (a)	1.49 ± 0.11 (a)
	10	3.63 ± 1.25 (b)	5.16 ± 1.02 (a)	2.24 ± 0.65 (ab)	1.42 ± 0.17 (a)
	ANOVA	*P*-values			
	Fertilizer	**0.004**	**0.005**	**0.008**	0.070
	Biochar	0.777	0.251	0.122	0.320
	Fertilizer x Biochar	0.769	0.698	0.953	0.063

Figure 2-1. (a) Soil aggregate stability (mean weight diameter) after 2 years (March 2014) and (b) soil aggregate stability (mean weight diameter) after 6 years (March 2018) under different management practices (compost and mineral fertilizer), with and without walnut shell biochar amendment (0 and 10 tons ha^{-1}); The error bars represent standard errors and bars with different letters indicate statistically significant ($P < 0.05$) differences.

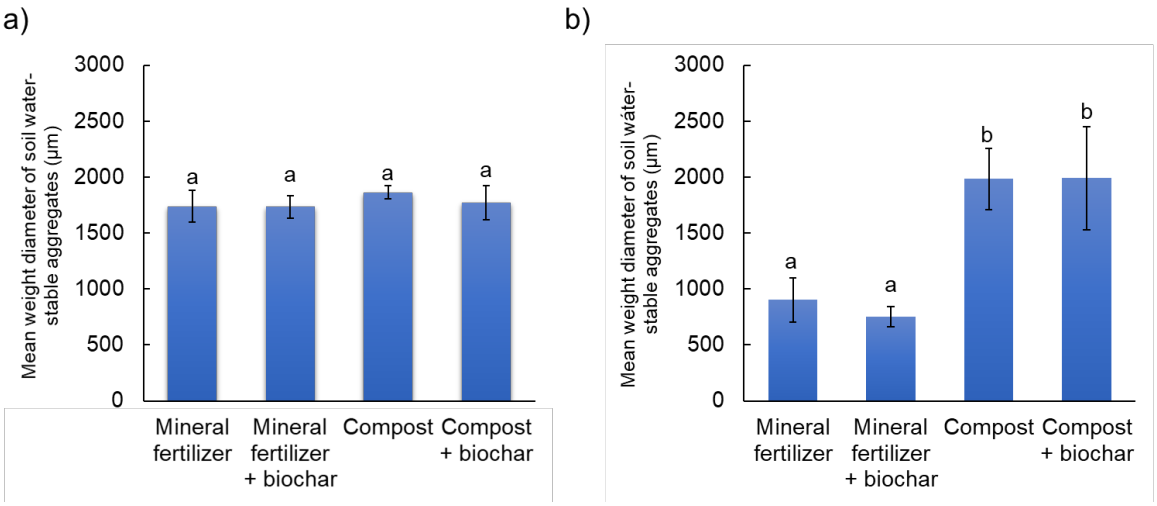


Figure 2-2. (a) Soil bulk carbon content after 2 years (March 2014) and (b) soil bulk carbon content after 6 years (March 2018) under different management practices (compost and mineral fertilizer), with and without walnut shell biochar amendment (0 and 10 tons ha^{-1}); The error bars represent standard errors and bars with different letters indicate statistically significant ($P < 0.05$) differences.

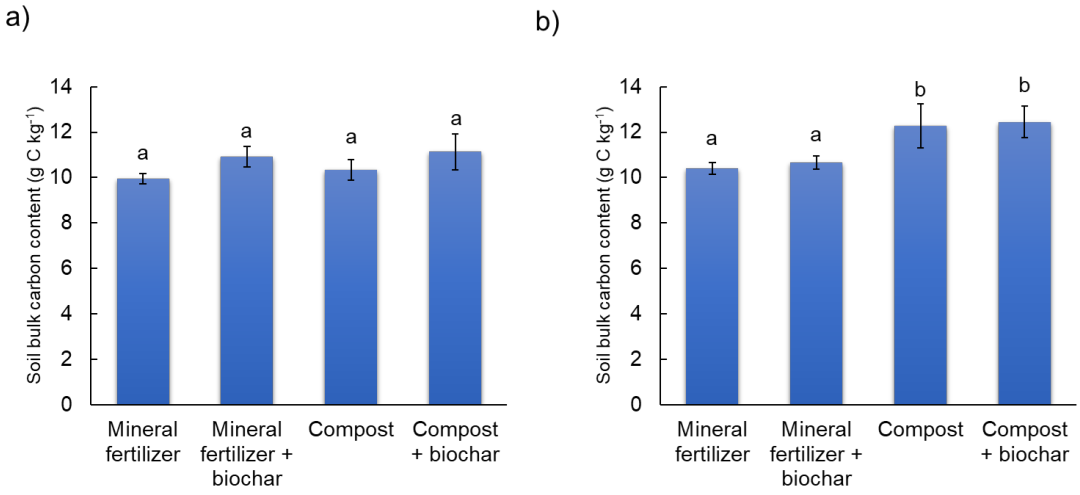


Figure 2-3. Soil dissolved organic carbon and soil microbial biomass carbon after 6 years (March 2018) under different management practices (compost and mineral fertilizer), with and without walnut shell biochar amendment (0 and 10 tons ha^{-1}). The error bars represent standard errors and bars with different letters indicate statistically significant ($P < 0.05$) differences.

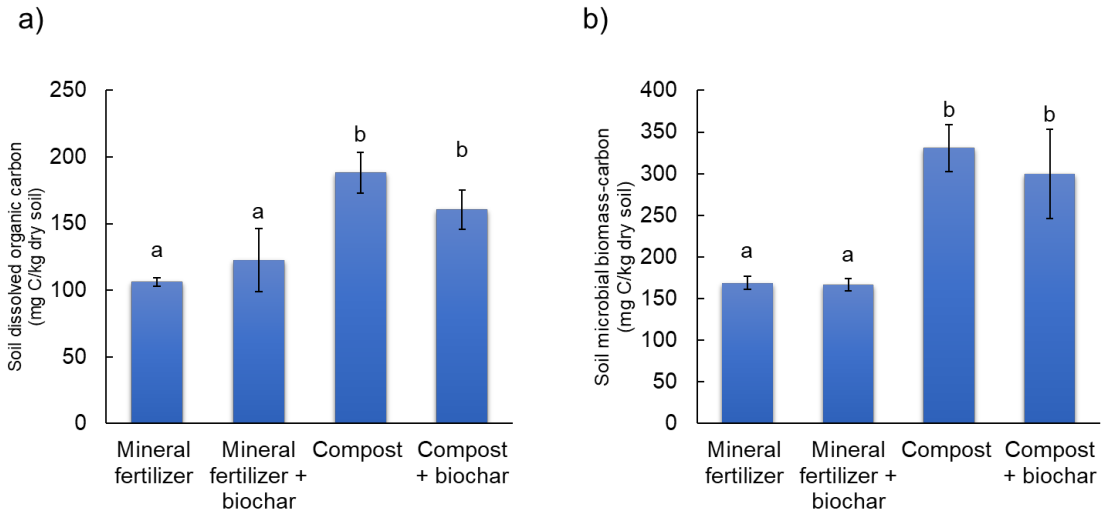


Figure 2-4. 16S rRNA gene copy number after 6 years (March 2018) under different management practices. In Figure 4(b), the with/without biochar treatments under the same fertilizer treatment were combined and compared since there was no significant difference between with/without biochar treatments and there was no interaction between fertilizer treatment and biochar treatment. The error bars represent standard errors and bars with different letters indicate statistically significant ($P < 0.05$) differences.

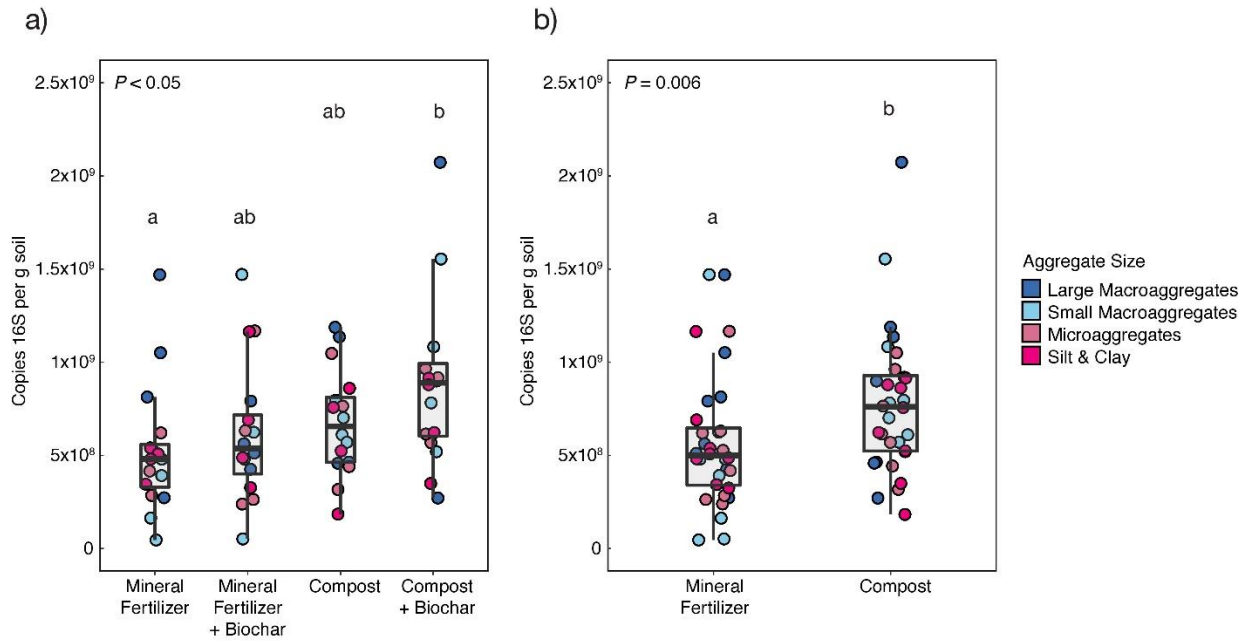


Figure 2-5. Internal transcribed spacer (ITS) copy number after 6 years (March 2018) under different management practices. In Figure 5b, the with/without biochar treatments under the same fertilizer treatment were combined and compared since there was no significant difference between with/without biochar treatments and there was no interaction between fertilizer treatment and biochar treatment. The error bars represent standard errors and bars with different letters indicate statistically significant (P < 0.05) differences.

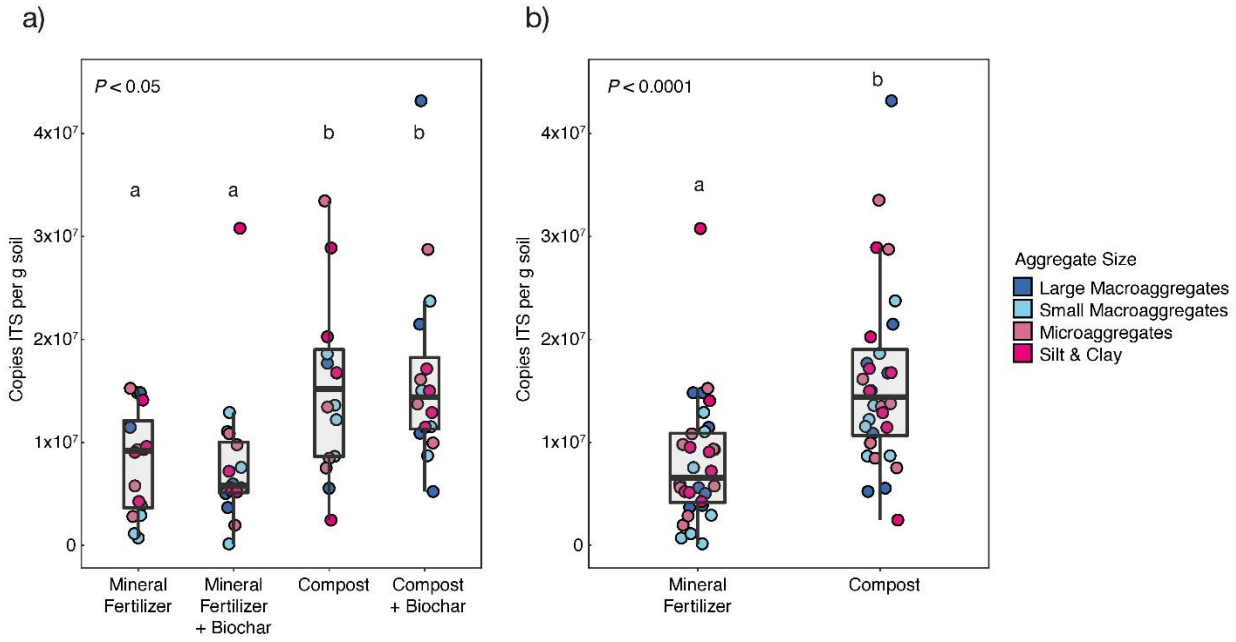


Figure 2-6. Weighted average 16S rRNA gene (a) and internal transcribed spacer (ITS) copy numbers (b) in bulk soil after 6 years (March 2018) under different management practices. The with/without biochar treatments under the same fertilizer treatment were combined and compared since there was no significant difference between with/without biochar treatments and no interaction between fertilizer treatment and biochar treatment. The error bars represent standard errors and bars with different letters indicate statistically significant (P < 0.05) differences.

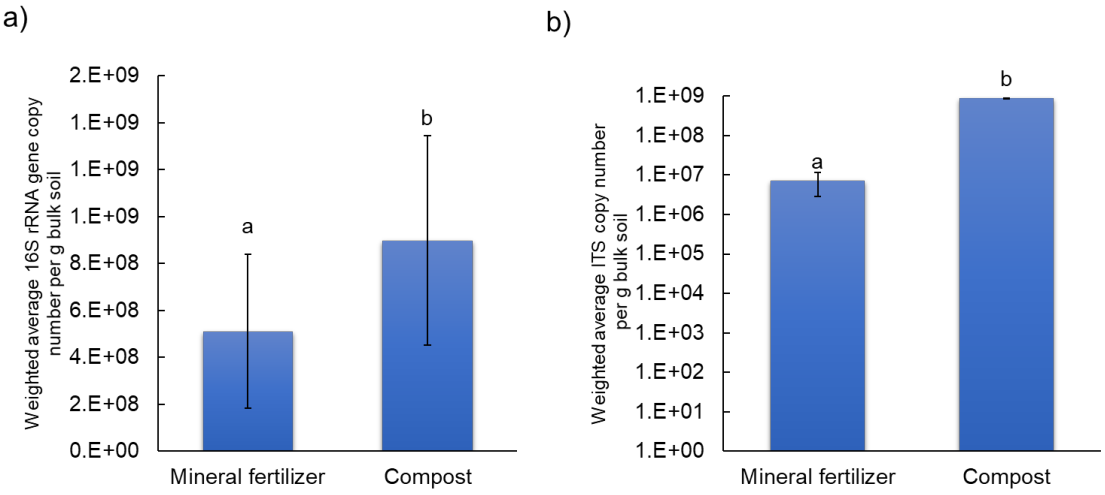


Figure 2-7. Nonmetric multidimensional scaling (NMDS) analysis of soil microbial communities with ESVs (triangles indicated samples from mineral fertilizer treatments, circles indicated samples from compost treatments) after 6 years (March 2018). Different colors and symbols represent different water-stable aggregate fractions.

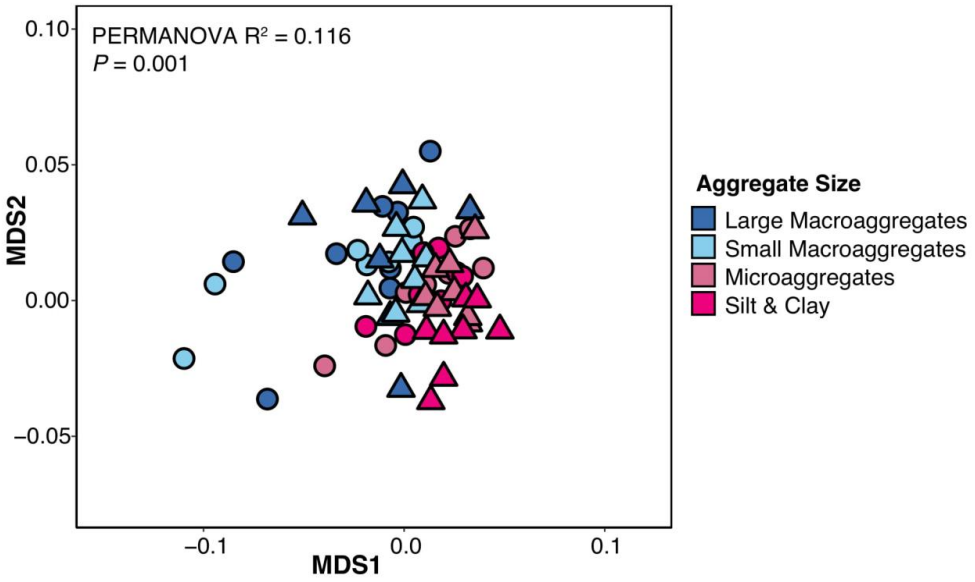
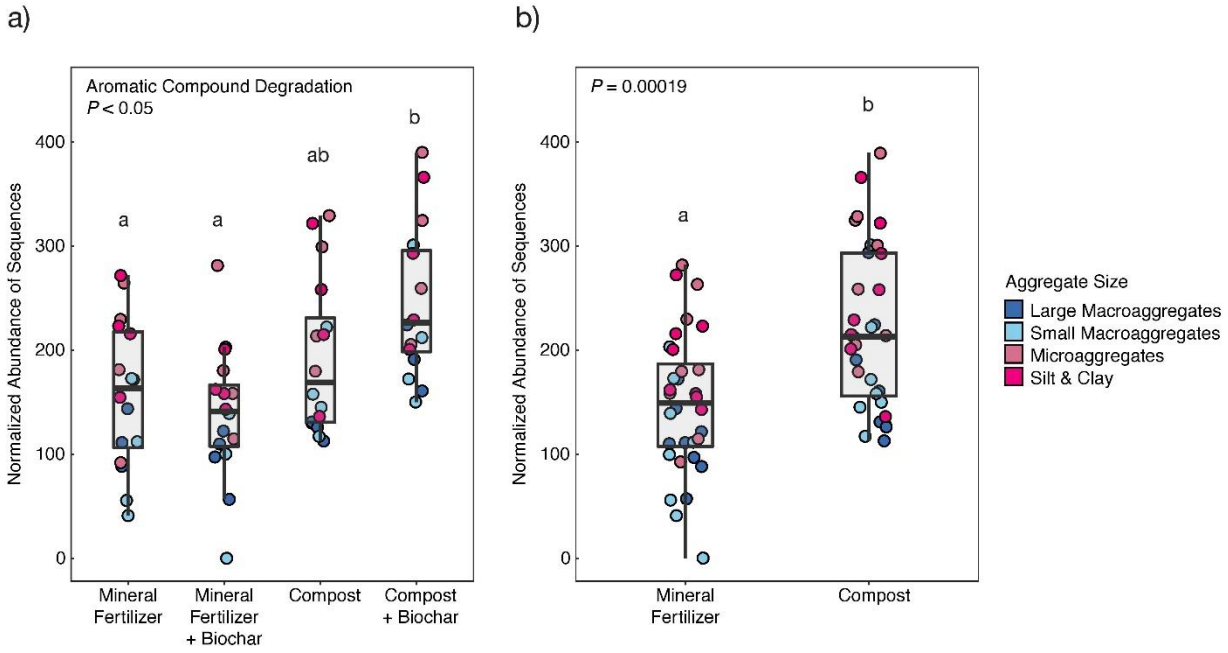


Figure 2-8. FAPROTAX predictive assignment of abundance of bacteria capable of degrading aromatic C compounds based on the 16S sequencing after 6 years (March 2018). In Figure 8(b), the with/without biochar treatments under the same fertilizer treatment were combined and compared since there was no significant difference between with/without biochar treatments and no interaction between fertilizer treatment and biochar treatment. The error bars represent standard errors and bars with different letters indicate statistically significant ($P < 0.05$) differences.



CHAPTER 1

SUPPLEMENTARY MATERIAL

Table S2-1. Tomato fruit yield in year 5 (2016) and corn grain yield (dry grain weight) in year 6 (2017) under different management practices (compost and mineral fertilizer), with and without walnut shell biochar amendment (0 and 10 Mg ha^{-1}); The numbers to the right of each value represent the standard error about the mean. Significant differences ($P < 0.05$) between treatments are indicated by different letters in parentheses to the right of each value differences.

	Mineral fertilizer	Mineral fertilizer + biochar	Compost	Compost + biochar
Tomato fruit yield, year 5 (Mg ha^{-1} fresh weight)	121.80 ± 2.91 (b)	131.90 ± 15.00 (b)	92.07 ± 6.05 (a)	93.5 ± 12.21 (a)
Corn yield, year 6 (Mg ha^{-1} dry weight)	11.48 ± 0.48 (b)	11.67 ± 0.89 (b)	8.50 ± 0.44 (a)	9.07 ± 0.44 (a)

Figure S2-1. Plot layout of the field trial. Four treatments: 1) mineral fertilizer without biochar; 2) mineral fertilizer with biochar; 3) compost without biochar; and 4) compost with biochar, arranged in a randomized complete block design with four replicate blocks per treatment and one treatment replicate per block, making a total of 16 plots. Each replicate plot was 4.6 m wide and 50 m long.

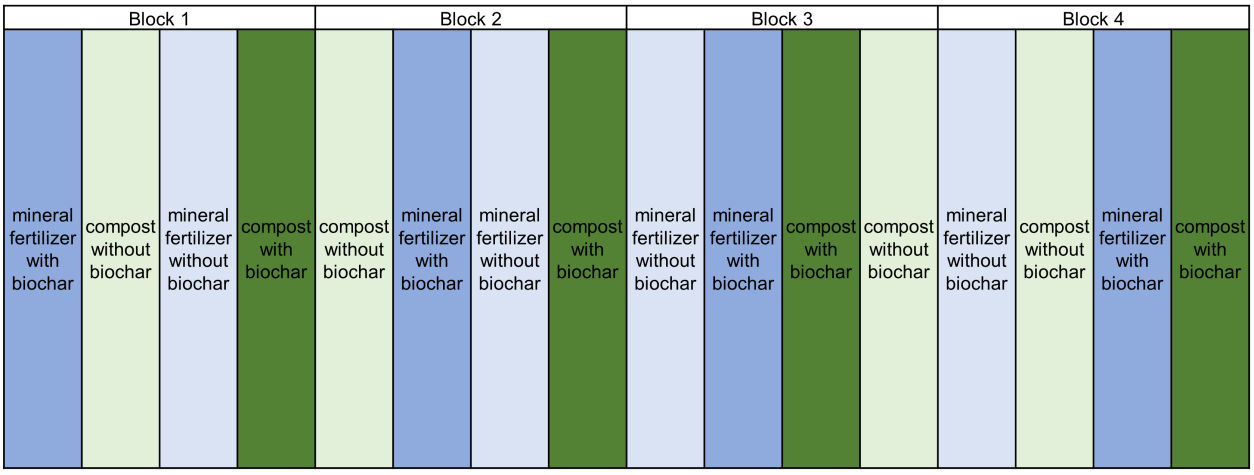
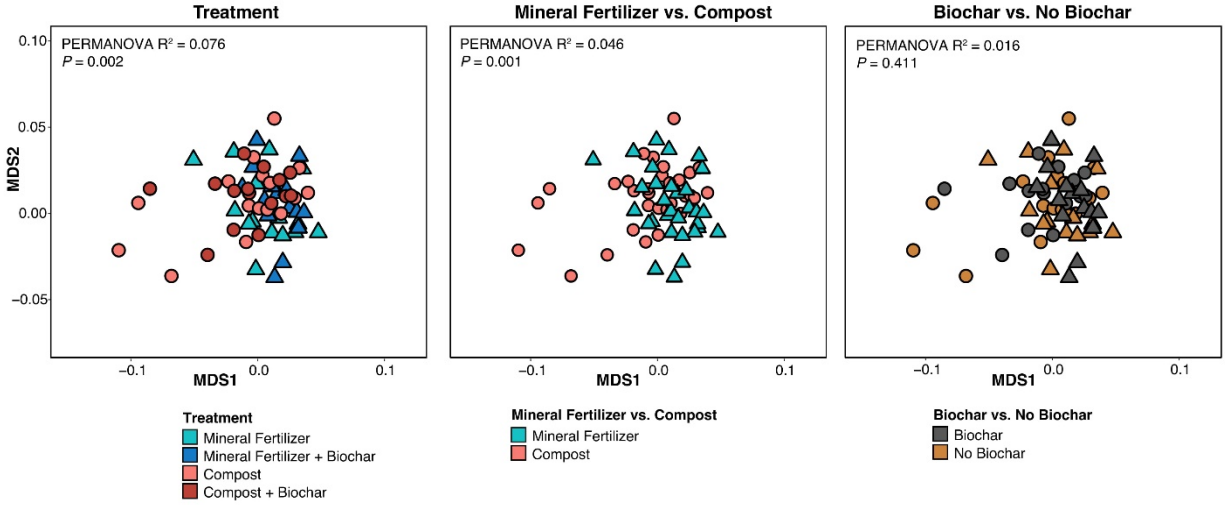


Figure S2-2. Nonmetric multidimensional scaling (NMDS) analysis of soil bacterial communities at the level of ESVs based on Bray-Curtis distances. Different colors and symbols represent (a) all different treatments, (b) compost vs. mineral fertilizer, and (c) biochar treatments.



CHAPTER 2

DIFFERENTIAL RESPONSES OF PROKARYOTIC AND FUNGAL COMMUNITIES IN SOIL MICROENVIRONMENTS TO DRYING AND WETTING AS AFFECTED BY SOIL AGGREGATE ISOLATION METHOD

List of authors: Jonathan Y. Lin, Daoyuan Wang, King C. Law, Kate M. Scow, Jorge L.M. Rodrigues

Under review at *Soil Biology and Biochemistry*

ABSTRACT

Most soil microorganisms live in communities within and on the surface of soil aggregates of varying sizes. A growing body of evidence suggests that different size fractions of aggregates are habitats for different microbial communities. A challenge in arriving at generalizations, however, is that a variety of different methods are used to isolate aggregates from bulk soil, thus making comparisons difficult. We compared two common aggregate isolation methods, dry and wet sieving, to investigate their effects on the prokaryotic and fungal communities in four aggregate size fractions (large macroaggregates (> 2000 μm), small macroaggregates (250-2000 μm), microaggregates (53-250 μm), and silt & clay (< 53 μm)) originating from field moist and dried soils. While prokaryotic community composition was different among treatments in each of the four size fractions, the composition and alpha diversity for fungi were more resistant to change in large and small macroaggregates than in the microaggregate and silt & clay fractions. A decrease in ammonia-oxidizing archaea and increase in potential spore-forming bacteria was detected when soils were dried prior to sieving based on the 16S rRNA gene sequences, while a decrease in free-living saprotrophs with an increase in fungal pathogens was observed when aggregates were separated by wet sieving based on the ITS amplicons. Compared to aggregates isolated from field moist soils, the average prokaryotic *rrn* copy number and genome size were increased in all aggregate size fractions when soils were dried before dry sieving, and these traits decreased in the microaggregate and silt & clay fractions when dried soils were wet sieved.

Together with our quantification of total C & N in each aggregate size fraction, for which both elements decreased in the microaggregate and silt & clay fractions with wet sieving, our results show that dry and wet sieving soils with different starting moisture conditions can affect the microbial communities in aggregates through drying and wetting dynamics.

INTRODUCTION

The soil is a heterogeneous matrix made up of aggregates – three-dimensional complexes composed of organic materials and mineral particles (Wilpiszeski et al., 2019). These complexes are held together by a combination of electrostatic forces and biotic factors and are classified by size, ranging from micro- (< 250 μm) to macroaggregates (> 250 μm) (Tisdall and Oades, 1982; Six et al., 2004; Jastrow et al., 2007). Aggregates are believed to be assembled in a hierarchical pattern where charged minerals, carbonates, and other particles are bound together to form microaggregates (Tisdall and Oades, 1982; Totsche et al., 2017). Microaggregates are then enmeshed by organic material such as bacterial polysaccharides, fungal hyphae, and plant roots to form macroaggregates (Tisdall and Oades, 1982; Six et al., 2000; Totsche et al., 2017). Together, soil aggregates and the pores within and between them comprise the physical structure of the soil (Tisdall and Oades, 1982; Wilpiszeski et al., 2019).

Soil aggregation plays many crucial roles in agroecosystems including promoting soil carbon (C) storage, contributing to soil water infiltration, and maintaining agricultural productivity (Totsche et al., 2017; Wilpiszeski et al., 2019). Aggregates can also determine the niche availability for resident soil microbes, where their physicochemical properties can affect the microbial communities that dwell inside or on the surface of aggregates (Rillig et al., 2017; Wilpiszeski et al., 2019). For instance, microelectrode profiles have revealed the presence of sharp O_2 concentrations inside water-saturated macroaggregates, potentially stratifying the distribution of oxygen-sensitive microbial activities in soil (Sexstone et al., 1985). Furthermore,

macroaggregates contain more particulate organic matter that may be more susceptible to microbial decomposition, whereas microaggregates have more phenolic C adsorbed to mineral particles that may be protected from microbial access (Davinic et al., 2012; Yu et al., 2015). These findings provide evidence that the unique environments in aggregates can serve as microhabitats for soil microbial communities (Foster, 1988; Hattori, 1988; Vos et al., 2013; Rillig et al., 2017).

Previous studies have shown that micro- and macroaggregate soil fractions harbor distinct communities of prokaryotes (Mummey et al., 2006; Davinic et al., 2012; Trivedi et al., 2017; Bach et al., 2018; Wang et al., 2022) and fungi (Bach et al., 2018; Li et al., 2019; Upton et al., 2019), with a few studies finding an overall higher diversity of microbes in the microaggregates compared to the macroaggregates (Bach et al., 2018; Upton et al., 2019). In addition, different aggregate size fractions have different total abundances of bacteria determined by cultivation-based methods (Kanazawa and Filip, 1986; Gupta and Germida, 1988), phospholipid-derived fatty acid (PLFA) profiles (Helgason et al., 2010; Wang et al., 2017), or quantitative PCR (Kong et al., 2010; Trivedi et al., 2015, 2017). By contrast, fungi are believed to be more important for macroaggregate formation and hence may be more abundant in macroaggregates than microaggregates (Lehmann et al., 2017). Soil aggregates also exhibit distinct extracellular enzyme activities in different size fractions (Fansler et al., 2005; Bach and Hofmockel, 2014; Trivedi et al., 2017; Wang et al., 2017), with high variation observed among individual macroaggregates (Bailey et al., 2012).

Despite numerous studies, relationships between aggregate size and microbial community dynamics have been highly variable (Blaud et al., 2017). For example, while clear differences in microbial community composition and relative abundance between aggregate size fractions were found in several studies (Davinic et al., 2012; Trivedi et al., 2017; Bach et al., 2018), two different reports found that the bacterial (Blaud et al., 2014) and archaeal community

compositions (Ramakrishnan et al., 2000) were highly similar across different aggregate sizes and variable in individual macroaggregates (Bailey et al., 2013). Although more bacterial and fungal PLFAs were detected in macro- compared to microaggregates from a long-term agricultural plot (Wang et al., 2017), others have found no differences in the abundance of total bacteria (Blaud et al., 2018; Wang et al., 2022) or nitrogen cycling genes (Blaud et al., 2018) by aggregate size. This lack of consensus makes it difficult to generalize the relationship between the soil physical structure and the microbial community, a critical component for predicting contributions of the soil microbiome to nutrient cycling, C dynamics, and soil health.

One major factor contributing to this variation is the use of different methods across studies to separate aggregates from bulk soil (Blaud et al., 2017). Soil aggregates are typically collected using one of two commonly employed isolation methods. The first method, dry sieving, involves shaking bulk soil on top of a stack of sieves with decreasing mesh sizes (Blaud et al., 2017). Shaking by manual movement or a mechanical rotary shaker allows aggregates of different sizes to fall into pre-determined fractions (Bach et al., 2018). Because the smaller particles such as the microaggregates (53-250 μm) and silt & clay (< 53 μm) can stick to the sieve or with each other, soils are often air-dried to desiccation or a predetermined “optimal” moisture prior to dry sieving to improve recovery of the smallest fractions (Bach and Hofmockel, 2014). However, this process can cause dehydration stress for soil microbes. On the other hand, wet sieving uses water to separate aggregates that remain stable under hydrostatic pressure (Elliott, 1986). Soils are first immersed in water for several minutes, followed by repeated vertical strokes in water to separate the aggregates on top of a nest of submerged sieves (Elliott, 1986; Blaud et al., 2017). The wet sieving method is more effective than dry sieving in retrieving material from the smallest size fractions (< 250 μm) but can alter nutrient levels and impart osmotic and mechanical pressure on microbes from water saturation and breakdown of macroaggregates into smaller aggregates (Bach and Hofmockel, 2014; Blaud et al., 2017). Some protocols

additionally dry soils before wet-sieving (Cambardella and Elliott, 1993), constituting a dry-rewetting cycle that can subject microbes to further desiccation and hypotonic stress.

Dry and wet sieving can differentially impact microbial communities within soil aggregates and affect the concentration of nutrients that can be metabolized by microbes. Yet, few studies have compared effects of different sieving methods (Blaud et al 2017, Bach and Hofmockel 2014) and none have investigated how prior moisture conditions of the bulk soil affect microbial communities in aggregates after dry or wet sieving. A better understanding of how these methods may shift ecologically important traits for microbial growth, resource use, and trophic strategies is necessary to provide a foundation to predict microbial community responses to moisture that may be associated with changing precipitation patterns or irrigation management.

The objective of our study was to investigate the effect of dry and wet sieving of field moist and dried soils on microbial communities in four size classes of soil aggregates. We hypothesized that 1) the prokaryotic and fungal communities in different aggregate size fractions will have different responses in their composition and ecological traits between the different sieving treatments and 2) microbial diversity in aggregates will be impacted more by wet than dry sieving.

MATERIALS AND METHODS

Long term field trial and soil sampling

Soils used in this study were collected from the Russell Ranch Sustainable Agricultural Research Facility, University of California, Davis (UC Davis), USA (38°32′47″N, 121°52′28″W). Details on the management, amendments, and soil type are described elsewhere (Griffin et al., 2017; Wang et al., 2022). Briefly, the field experiment was initiated in May 2012 to investigate the long-term impacts of different soil amendments on soil carbon storage, microbial

communities, and plant production in a Mediterranean climate characterized by dry arid summers and wet winters. The cropping system consisted of a 2-year rotation of processing tomatoes (*Lycopersicon esculentum* Mill.) and corn (*Zea mays* L.) managed using practices and equipment similar to those used by local commercial growers. The experiment was arranged in a randomized block design with four replicate plots per sample. Soils were collected in April 2019 from plots receiving either conventional synthetic mineral fertilizer or manure compost as fertilizer treatments (**Figure 3-1a**). A total of six soil cores were taken from a depth of 15 cm from each plot, combined into a single composite sample, and transported to the lab for subsequent analysis.

Soil aggregate sieving treatments

We compared two commonly used methods to isolate soil aggregates based on size: dry-sieving and wet-sieving. Freshly collected bulk soils were first passed through an 8 mm sieve by breaking soil clods by hand along natural planes of weakness. For dry-sieving, the field-most samples were first air-dried at room temperature until they reached a gravimetric water content of ~ 6.7% to obtain enough material from the smallest size fractions for analysis as determined by previous tests. After air drying, 100 g of soil was placed on top of a stack of three sieves of with different mesh sizes (2000 μm, 250 μm, and 53 μm) and separated using a rotary shaker for 3 min at ~ 200 rpm. These three sieve sizes (2000 μm, 250 μm, and 53 μm) were used to generate four aggregate size fractions (**Figure 3-1b**): 1) large macroaggregates (> 2000 μm); 2) small macroaggregates (250-2000 μm); 3) microaggregates (53-250 μm); and 4) silt & clay (< 53 μm). For each fraction, subsamples of the aggregates that were retained on each sieve were collected in sterile 50 mL polypropylene tubes and immediately stored at -80°C until DNA extraction. The remaining material in each fraction was then oven-dried at 60°C for analysis of total C and N content as described below.

For wet-sieving, 100 g of the field-moist soil was submerged in deionized water at room temperature on top of a 2000 μm sieve for 5 min. The sieve was then moved up and down (~ 3 cm) for 2 min at a rate of 50 repetitions per min (Elliott, 1986). The soil and water passing through the sieve were then transferred by gently rinsing the material with deionized water onto the next smaller size sieve as listed above, and the same procedure was repeated. Subsamples of aggregates collected from each fraction were stored at -80°C until DNA extraction or rinsed off into aluminum pans and oven-dried at 60°C for chemical analysis. To simulate a prolonged drought event, subsamples of the original field-moist soils were air-dried at room temperature for approximately one month where they reached an average gravimetric water content of 4.77% (**Table S3-1**). Afterwards, 100 g subsamples of these dried soils were subjected to dry or wet-sieving as above to obtain aggregate samples after a simulated drying or dry-rewetting event, respectively. This resulted in four sieving treatments (hereafter referred to the names within parentheses): 1) Dry-sieving of field-moist soils (Field + DS); 2) wet-sieving of field-moist soils (Field + WS); 3) dry-sieving of dried soils (Dry + DS); and 4) wet-sieving of dried soils (Dry + WS) (**Figure 3-1a**).

Soil aggregate C and N content and bulk soil physicochemical parameters

The total C and N content for each aggregate size fraction was analyzed using a PDZ Europa ANCA-GSL elemental analyzer interfaced with a PDZ Europa 20-20 isotope ratio mass spectrometer (Sercon, Cheshire, UK) at the UC Davis Stable Isotope Facility. Bulk soil samples were sent to A & L Western Laboratories (Modesto, CA, USA) to measure soil physicochemical parameters, which are listed in **Table S3-2**.

Soil DNA extraction and amplicon sequencing

DNA from each soil aggregate size fraction (32 per fraction) was extracted using the Powerlyzer PowerSoil DNA Isolation kit (Qiagen, Germantown, MD, USA) using a vortex adaptor

according to the manufacturer's instructions. The extracted DNA was quantified using the Qubit High-Sensitivity dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA). To identify prokaryotic communities, the V4 hypervariable region of the 16S rRNA gene was amplified from each sample in duplicate using the primer pair 505F/816R (Caporaso et al., 2012), which was designed to include Illumina adaptors and 12 bp barcode sequences. Fungal communities were analyzed by amplifying the internal transcribed spacer (ITS) region in duplicate using the ITS1F/ITS2 primer pair designed by Smith and Peay (2014), which incorporates the same barcoding and indexing strategy as the primers for the 16S rRNA gene. Amplification reactions were carried out in 20 µL volumes containing 10 µL Phusion Hot-Start II High-Fidelity Master Mix (ThermoFisher, Waltham, MA, USA), 0.5 µM each primer, 10 ng sample DNA, and 4 µL sterile ddH_2O. The resulting amplicons were inspected by gel electrophoresis on a 1% agarose gel and pooled in equimolar concentrations for each amplicon. Pools comprising 128 samples for each amplicon were cleaned using the QIAquick PCR Purification kit (Qiagen, Germantown, MD, USA) and sequenced using an Illumina MiSeq platform (1 sequencing lane per amplicon, paired-end 250 bp) at the UC Davis DNA Technologies core facility using the custom sequencing and indexing primers described for each protocol (Caporaso et al., 2012; Smith and Peay, 2014).

Sequence data processing

Raw reads for the 16S rRNA gene and ITS amplicon sequences were processed using DADA2 v.1.12.1 (Callahan et al., 2016) implemented in R v.3.6.2. Briefly, paired-end fastq files were processed by quality-trimming forward and reverse reads to 240 and 200 bp and 230 and 160 bp lengths for the 16S rRNA gene and ITS reads, respectively. After sequence dereplication, merging, error correction, and chimera removal, Exact Sequence Variants (ESVs) were inferred, and taxonomic identification was performed using the SILVA database v.132 for the 16S rRNA gene and UNITE database (general FASTA release 04.02.2020) for the ITS sequences. The

FUNGuild annotation pipeline (Nguyen et al., 2016) was used to assign fungal taxa into ecologically relevant groups representing different trophic strategies and functional guilds.

Phylogenetic estimation of microbial traits

We used a phylogenetic method to estimate the ribosomal RNA gene (*rrn*) copy number and genome size, two ecologically important traits for prokaryotes that predict their ability to proliferate after rapid nutrient input (Klappenbach et al., 2000) and their range of resource use, respectively. Representative 16S rRNA gene sequences for each ESV (totaling 12,677 sequences) were first aligned to each other using MUSCLE (Edgar, 2004). Then, a profile-profile alignment was performed in MUSCLE to align the ESV sequences with the "broad reference alignment" of 16S rRNA gene sequences from reference genomes curated by Grauver and Eskelinen (2017). These alignments were then used in pplacer (v1.1.alpha19-0-g807f6f3) (Matsen et al., 2010) to place the query ESV sequences onto the reference tree. The resulting placements were exported using the '-tog' option in Guppy to produce a tree in newick format, where each ESV placed onto the reference tree is represented as a pendant edge (Matsen et al., 2010). The tree was then imported into R studio as a 'phylo' object using the package 'ape' (Paradis et al., 2004) and traits were estimated using the 'phyEstimate' function in the package 'picante' (Kembel et al., 2010), which uses ancestral state estimation techniques to predict each ESV's trait value based on known trait values of neighboring reference taxa. For both traits, the 'phyEstimate' command was used with the dataset of reference trait values for *rrn* copy number and genome size provided previously (Gravuer and Eskelinen, 2017). Community-weighted averages for each estimated trait were calculated for each sample using the 'functcomp' function within the 'FD' package (Laliberté and Legendre, 2010).

Statistical analysis

All univariate data were first tested for assumptions of normality and homogeneity of variance before comparison using a one-way Analysis of Variance (ANOVA) test and a Tukey Honestly Significant Difference post-hoc test to identify significant differences between treatments. Data that failed to meet assumptions for ANOVA were compared by using a nonparametric Kruskal-Wallis test followed by a Dunn's post-hoc test.

Bray-Curtis distance matrices were calculated for the 16S rRNA and ITS amplicon datasets for Non-Metric Multidimensional Scaling (NMDS) analysis to visualize overall differences in the prokaryotic and fungal community composition among treatments. Differences in microbial community composition were tested by permutational multivariate analysis of variance (PERMANOVA) running 999 permutations with the 'adonis' function (Oksanen et al., 2019) in R v.3.6.2. using Bray-Curtis dissimilarities and treatment, aggregate size, fertilizer management, and their interaction as predictor variables. Within-group dispersions were calculated and their differences between treatments were tested using the 'betadisper' and 'permutest' functions (Oksanen et al., 2019). All statistical analyses were performed in the R environment using the 'vegan' package (Oksanen et al., 2019) and graphics were generated using 'ggplot2' (Wickham et al., 2020). For all analyses, a value of $P < 0.05$ was considered statistically significant.

Data availability

The 16S rRNA and ITS amplicon sequencing data were deposited to the NCBI sequence read archive (SRA) under BioProject PRJNA794192 with the following BioSample accession numbers: SAMN24618211 (16S rRNA) and SAMN24618224 (ITS).

RESULTS

Sequencing results and normalization

After quality filtering, error correction, and chimera removal, the total number of paired-end reads was 3,441,795 for the 16S rRNA gene and 2,418,680 for the ITS region. The number of 16S rRNA gene sequences per sample ranged from 8,161 to 88,429 with an average of 27,101, whereas the number of ITS reads per sample ranged from 8,768 to 41,675 with an average of 19,664. The 16S rRNA gene and ITS reads were rarefied to 8,161 and 8,768 sequences per sample, respectively, to ensure equal depth for statistical analysis. After rarefaction, the total number of unique ESVs was 12,677 and 3,204 for the 16S rRNA gene and ITS region, respectively. Due to low sequencing depth, 1 sample from the 16S dataset and 5 samples from the ITS dataset were removed prior to analysis.

Consistent responses of the prokaryotic community to sieving treatments

Across all samples, the prokaryotic community composition at the ESV level differed by sieving treatment as well as by aggregate size, and fertilizer treatment (**Table 3-1a**). The community composition was also significantly different from the interaction between sieving treatment and aggregate size, sieving treatment and fertilizer regime, but not from the interaction between aggregate size and fertilizer or from all three variables (**Table 3-1a**). These differences are most apparent with the separation of prokaryotic communities by sieving treatment across all samples observed in the NMDS ordination (**Figure 3-2a**). A permutational test of the multivariate homogeneity of variances revealed that within-group dispersions differed by sieving treatment (BETADISPER, $F = 4.06$, $P = 0.01$), with a lower microbial community variation observed in Dry + WS compared to Dry + DS and Field + DS (**Figure S3-1a**).

To identify the effects of our sieving treatments on microbial community dynamics within each aggregate size fraction, we separated our dataset to perform individual analyses on each aggregate size. We found that the sieving treatment significantly affected prokaryotic community

composition in each aggregate size fraction (**Figure 3-3a-d**), with larger effect sizes observed in the microaggregate and silt & clay fractions than in large and small macroaggregates.

Among all samples, higher numbers of prokaryotic ESVs (Richness) were observed in Field + WS compared to the other treatments (**Figure S3-2a**), and similar results were observed by using the Shannon index (**Figure S3-2c**). This pattern of higher alpha diversity in Field + WS was consistent in each aggregate size fraction using both the Richness and Shannon indices except in the small macroaggregates, where no differences between any of the treatments were found (**Figure 3-4a-h**).

Overall, the 16S rRNA gene was represented by the phyla *Actinobacteria* (30.23%), *Proteobacteria* (19.98%), *Acidobacteria* (10.28%), *Firmicutes* (10.08%), *Chloroflexi* (9.50%), *Thaumarcheota* (6.23%), *Gemmatimonadetes* (4.64%), *Planctomycetes* (2.93%), *Bacteroidetes* (2.88%), and *Verrucomicrobia* (1.36%) (**Figure S3-3**). Of the 20 most abundant taxa at the genus level, notable groups that significantly responded to the sieving treatments included Candidatus *Nitrososphaera*, which was lower in Dry + DS; *Bacillus*, *Blastococcus*, and *Nocardioides*, which were higher in Dry + DS and Dry + WS, *Microvirga*, which was higher in Dry + WS; and other groups from within the phyla *Actinobacteria* and *Proteobacteria* (**Table S3-3**).

Fungal community responses to sieving treatments differed by aggregate size

Similar to the 16S rRNA gene results, the composition of the fungal community at the ESV level was significantly different by sieving treatment, aggregate size, and fertilizer regime in all samples (**Table 3-1b**). The fungal community composition also differed by the interaction between sieving treatment and aggregate size, but not by interactions between any of the other variables (**Table 3-1b**). Visualization of the NMDS ordination showed a strong effect of sieving in shifting fungal community composition (**Figure 3-2b**), with an apparent effect of the sieving treatment on decreasing within-group dispersions. Indeed, the BETADISPER analysis showed

that fungal community dispersions were significantly different by the sieving treatments (F = 17.14, P = 0.001). Both wet sieving treatments had reduced fungal community dispersion compared to the two dry sieving treatments, respectively, whereas Dry + DS also had lower community dispersion compared to Field + DS (**Figure S3-1b**). Interestingly, the sieving treatments had no effect on the fungal community composition within the large or small macroaggregates (**Figure 3-3e & f**). In contrast, significant effects of the sieving treatments on fungal community composition were observed in the microaggregate and silt & clay fractions (**Figure 3-3g & h**).

Overall, Dry + WS had lower fungal alpha diversity compared to all other treatments when measured using the Richness and the Shannon indices (**Figure S3-2b & d**). Dry + WS was also lower in fungal alpha diversity compared to the three other treatments in the small macroaggregates, microaggregates, and silt & clay fractions (**Figure 3-5b-d, f-h**), whereas no differences were observed between treatments in the large macroaggregate fraction (**Figure 3-5a & e**). Among all samples, the fungal community was dominated by the phyla *Basidiomycota* (73.93%), *Mortierellomycota* (14.20%), *Ascomycota* (9.69%), and *Olpidiomycota* (1.75%), with *Rozellomycota* (0.39%), *Basidiobolomycota* (0.02%), and *Glomeromycota* (0.01%) comprising the main phyla at under 1% (**Figure S3-4**). Notable groups that shifted significantly in relative abundance among the top 20 genera included *Solicoccozyma* and *Olpidium*, which were increased with wet sieving, and *Mortierella*, *Cystofilobasidium*, *Filobasidium*, *Lepidosphaeria*, and *Articulospora*, all of which were decreased with wet sieving (**Table S3-4**). The relative abundances of other major groups from within the phyla *Ascomycota* and *Basidiomycota* by treatment are reported in **Table S3-4**.

Sieving treatment effects on prokaryotic traits and fungal guilds

The sieving treatment significantly affected the distribution of microbial traits calculated from the 16S rRNA gene sequences. Across all samples, both Dry + DS and Dry + WS had higher community-weighted averages for *rrn* copy number and genome size compared to Field + DS and Field + WS (**Figure S3-5**). Notably, Dry + WS had a lower average *rrn* copy number compared to Dry + DS (**Figure S3-5a**). Different effects of the sieving treatment on microbial traits were observed in different aggregate sizes. First, Dry + DS and Dry + WS generally had higher community-weighted averages for *rrn* copy number and genome size compared to Field + DS and Field + WS in each aggregate size fraction (**Figure 3-6a-h**). Interestingly, the wet sieving treatments decreased the weighted averages of both traits, but this trend was only observed in the microaggregate and silt & clay fractions and not in the large and small macroaggregates (**Figure 3-6**). Specifically, Dry + WS had a lower average *rrn* copy number compared to Dry + DS in microaggregates (**Figure 3-6c**), and both wet sieving treatments had a decreased average *rrn* copy number and genome size compared to their respective dry sieving treatments in the silt & clay fraction (**Figure 3-6d & h**). Together, these results show that drying soils can lead to an overall increase the average *rrn* copy number and genome size, while wetting soils can decrease the average values of these traits in the microaggregate and silt & clay fractions.

The sieving treatment shifted the abundance of fungal taxa among different trophic modes and functional guilds. The wet sieving treatments significantly decreased the relative abundance of fungi classified as Saprotrophs (**Figure 3-7a**) and fungi classified as either Saprotrophs or Symbiotrophs (**Figure 3-7b**) compared to dry sieving. In contrast, the relative abundance of Pathotrophs increased with wet sieving compared to the dry sieving treatments (**Figure 3-7c**). These results were matched by changes in the fungal guilds found within each trophic group. Both wet sieving treatments decreased the relative abundance of the guilds

classified as Undefined Saprotrophs and Leaf Saprotrophs (**Figure 3-7d & e**) and increased the relative abundance of the Plant Pathogen guild compared to the dry sieving treatments (**Figure 3-7f**).

Sieving treatment effects on total aggregate C and N

The wet sieving treatments decreased the concentration of total C compared to dry sieving and had no effect on total N across all samples (**Figure S3-6a & b**). However, different patterns were observed when the treatment effects were analyzed in each aggregate size fraction. First, within each aggregate size and treatment group, samples from fields receiving compost generally had higher concentrations of total C and N compared to samples from fields receiving mineral fertilizer (**Figure 3-8**), which was consistent with a previous report by Wang et al. (2022). Second, the sieving treatment had no effect on total C in the large and small macroaggregate fractions, but the total N was increased in Dry + WS in both fractions of macroaggregates (**Figure 3-8a-b, e-f**). Finally, in the microaggregate and silt & clay fractions, both wet sieving treatments had lower total C and N in the microaggregates compared to dry sieving (**Figure 3-8c & g**), whereas only the total C was significantly decreased by wet sieving in the silt & clay fraction (**Figure 3-8d & h**). The average *rrn* copy number but not genome size was positively correlated with aggregate total C and N (**Figure S3-7**).

DISCUSSION

It has been proposed that the vast variation in sizes and composition of soil aggregates with their unique physicochemical properties provides a diverse range of habitats for different soil microorganisms (Hattori, 1988; Vos et al., 2013; Rillig et al., 2017). Yet, despite numerous studies spanning more than two decades (Chenu and Cosentino, 2011), it is still difficult to pull out generalizations about relationships between soil aggregate size and microbial community

composition and diversity. In our study, we propose that a major source of this variation in the literature is methodological. Overall, we found that different soil aggregate isolation methods significantly influenced the microbial community composition and diversity within aggregates, but the specific effects varied for each aggregate size and for prokaryotes versus fungi.

Effects of sieving treatments on prokaryotes

While microbes can undergo dehydration stress when soils are dried prior to dry sieving, the slaking step in wet sieving breaks up unstable aggregates and can expose microbes to strong osmotic stress (Cambardella and Elliott, 1993; Bach and Hofmockel, 2014). The Prokaryotic communities within aggregates in our study were sensitive to the different sieving treatments within each aggregate size. Variation in prokaryotic communities among replicates also differed between treatments, with a significantly lower dispersion in aggregates from Dry + WS than Dry + DS. Soils that are dried before wet sieving show greater slaking, due to higher volumes of air that become trapped in aggregates after submersion in water (Chenu and Cosentino, 2011). This increase in pressure gradient stimulates breakdown of macroaggregates (Chenu and Cosentino, 2011). The fact that within-group dispersions of the prokaryotic communities were decreased in Dry + WS may be because the dried macroaggregates were disrupted into smaller particles during wet sieving (Blaud et al., 2017), causing communities in each sample to be more similar to each other.

The addition of water in wet sieving acts as a rapid pulse of moisture in aggregates that can resuscitate phylogenetically diverse bacteria from a dormant state (Placella et al., 2012), leading to increased microbial activity (Schimel, 2018). Field + WS had the greatest impact on increasing prokaryotic alpha diversity in most aggregate size fractions except for the small macroaggregates. The wetting of soil is usually followed by an immediate flush of microbial respiration that is several orders of magnitude higher than in field-moist soil. It is not yet known

whether an increase in microbial diversity plays an important role in driving this phenomenon known as the “Birch effect” (Birch, 1958), especially in soils that undergo seasonal cycles of drying and wetting (Niederberger et al., 2019). Wetting increased 16S rRNA gene richness in one soil (Meisner et al., 2021), but decreased alpha diversity in another soil (Meisner et al., 2018). We did not find a consistent pattern in prokaryote diversity within each aggregate size fraction in our study, suggesting that clear patterns between wetting and microbial diversity responses are not yet evident.

Drying soil can lead to cell death, leaving behind only the strains with the ability to produce osmolytes to maintain cellular water potential or synthesize impermeable membranes (Filippidou et al., 2016). However, remnant DNA from dead bacteria can persist in soil and still be detected in DNA-based analyses (Carini et al., 2016). We found that Dry + DS and Dry + WS shifted community composition but had no effect on prokaryotic alpha diversity. Furthermore, Dry + DS and Dry + WS reduced the relative abundance of the genus Candidatus *Nitrososphaera* and increased the relative abundance of genus *Bacillus* compared to Field + DS and Field + WS. The *Nitrososphaera* are ammonia-oxidizing archaea (AOA) that are integral to nitrification in agricultural soils (Huang et al., 2021). Nitrification has been shown to be sensitive to soil drying due to the limited diffusion of ammonia (NH_4^+) in dried soils (Stark and Firestone, 1995). On the other hand, many *Bacillus* species form endospores as a survival strategy during stress conditions, remaining dormant until favorable conditions return (Abel-Santos, 2015). We surmise that the drying of soil prior to sieving limited nutrient mobility in aggregates and killed cells that were susceptible to desiccation, thereby resulting in a decrease in the *Nitrososphaera* and a proportional increase in the endospore-forming members of the *Bacillus*. However, some DNA from deceased cells may have persisted and been captured in our PCR analysis, potentially explaining the discordance between compositional shifts and lack of changes in prokaryotic alpha diversity in the Dry + DS and Dry + WS treatments.

Effects of sieving treatments on fungi

Previous studies have shown that fluctuations in moisture have small or limited effects on soil fungal communities. Drying and rewetting had little effect on fungal growth measured through PLFAs across three different soil types (Bapiri et al., 2010), whereas drying only resulted in small shifts in soil fungal community composition when manipulated in the field (Cregger et al., 2012) or in laboratory-based microcosms using DNA and RNA-based analyses (Barnard et al., 2013). These studies suggest that fungi are particularly tolerant in recovering from drought stress (Hawkes et al., 2011). We found that the magnitude of the effect of moisture fluctuation on soil fungal communities differed considerably by aggregate size. The fungal communities were more resistant to changes in composition and alpha diversity in large and small macroaggregates than in microaggregate and silt & clay fractions. These findings underscore the importance of fungi for the formation and stabilization of macroaggregates (Lehmann et al., 2017). Arbuscular mycorrhizal fungi (AMF) are symbiotrophs that grow in close association with plant roots and have been shown to stabilize macroaggregates through their hyphae and secretion of extracellular polymeric substances (EPS) (Rillig et al., 2010). Increasing evidence suggests that filamentous free-living saprotrophs are also significant contributors to the formation and maintenance of macroaggregates (Caesar-Tonthat, 2002; Daynes et al., 2012; Tisdall et al., 2012; Lehmann and Rillig, 2015). The enmeshing properties of hyphae from AMF and biomass of filamentous saprotrophs (Lehmann et al., 2020) likely explains the lack of detectable changes in fungal community dynamics in the macroaggregates across the different sieving treatments.

Compared to the large and small macroaggregates, fungal community composition and alpha diversity in the microaggregate and silt & clay fractions were more sensitive to changes in moisture. The response patterns in fungal alpha diversity identified in these size fractions were

similar to that observed in bulk soils by Meisner et al (2018) when soils in microcosms were exposed to drought and extreme re-wetting conditions. This suggests that unlike the fungi that inhabit the larger pore spaces within macroaggregates, the fungal species that reside in or on the surface of the microaggregates and silt & clay particles are more susceptible to water, likely through direct contact with washing and erosive forces. In addition, the proportion of microaggregate and silt & clay particles that resulted from macroaggregate breakdown or erosion from macroaggregate surfaces likely increased as soils were dried and then rewet, further homogenizing the fungal communities in these fractions.

Wet sieving decreased the relative abundance of saprotrophs and symbiotrophs and increased the relative abundance of pathotrophs in aggregates. These changes were mainly driven by decreases in the genus *Mortierella* and *Articulospora* and increases in *Olpidium* in the wet sieving treatments. The *Mortierella* are filamentous soils dwellers categorized as saprotroph-symbiotrophs due to their ability to grow on decaying organic material (Yadav et al., 2015), whereas *Articulospora* species are aquatic saprotrophs that can also grow as endophytes in plant roots (Seena et al., 2018). On the other hand, *Olpidium* species are fungal parasites that are commonly found in the roots of field crops (Lay et al., 2018). A previous study showed that conditions created by extreme rainfall events can decrease populations of symbiotrophs and increase the proportion and richness of fungal pathogens in soil (Barnes et al., 2018). The saturation of soil with water results in hypoxic or anoxic conditions that can reduce total fungal biomass (Wagner et al., 2015). Accordingly, the decrease in the proportion of saprotrophs and symbiotrophs may be due to depletion of O_2 in the soil matrix needed by many fungal groups for growth (Ivarsson et al., 2016) or increased decomposition of substrates from wetting resulting to depleted nutrient stocks and cell death. These conditions may thereby favor the expansion of opportunistic fungal pathogens. However, the specific abiotic conditions and biological

mechanisms important for the survival of fungi with different trophic strategies during soil wetting has not yet been determined.

Effects of sieving treatments on prokaryotic traits and total C and N

Traits are phenotypic characteristics that affect the fitness or function of an organism (Krause et al., 2014). Understanding how traits for microbes are structured at the community level is important for linking microbial survival strategies to environmental changes (Krause et al., 2014; Malik et al., 2020). Two commonly studied prokaryotic traits that are ecologically important are the ribosomal RNA gene (*rrn*) copy number and genome size. A higher *rrn* copy number per genome determines a cell's ability to synthesize more ribosomes for a higher maximum growth rate (Roller and Schmidt, 2015), predicting its ability to rapidly increase growth in response to nutrient pulses (Klappenbach et al., 2000; Gravuer and Eskelinen, 2017). While species with low *rrn* copy numbers per genome have lower maximum growth rates, they are better adapted to use substrates more efficiently under resource-limited (oligotrophic) conditions than microbes with higher *rrn* copy numbers (Roller and Schmidt, 2015). Prokaryotes with large genomes are predicted to be able to use a wide variety of substrates, enabling them to thrive in environments with a high diversity of resources (Barberán et al., 2014; Krause et al., 2014). By contrast, microbes with small genomes are more streamlined to dominate in oligotrophic or constant environments (Krause et al., 2014; Leff et al., 2015; Gravuer and Eskelinen, 2017). Using a phylogenetic method to estimate community-weighted abundances of these traits, we found that the average *rrn* copy number and genome size was greater in Dry + DS and Dry + WS, while both traits were lower in Dry + WS than in Dry + DS in the two smallest size fractions.

An increase in the average *rrn* copy number and genome size from drying suggests that microorganisms with these traits are better adapted to the changes in nutrient conditions that occur from soil drying. Patel et al. (2021) recently found significantly higher concentrations of

water-extractable C in soils air dried to 5% saturation than in moist soils, suggesting that a substantial amount of C can become available during the drying process. This flush of C compounds consisted of aliphatic and alpha-H amide groups that likely originated from microbes that had lysed (Patel et al., 2021). However, these compounds quickly disappeared (Patel et al., 2021), likely because they were rapidly metabolized by fast-growing strains with high *rrn* copy numbers and large genome sizes (Fierer et al., 2007). Second, species with higher *rrn* copy numbers per genome may have an increased fitness advantage compared to species with lower *rrn* copy numbers during stress conditions. Bacteria with higher *rrn* copy numbers per genome have greater ribosomal translational power enabling higher protein synthesis rates (Dethlefsen and Schmidt, 2007), and some prokaryotes with high *rrn* copies per genome such as the *Actinobacteria* can invest in ribosome synthesis and accumulation during soil drying (Barnard et al., 2013), potentially giving them a competitive edge for nutrient acquisition once conditions become favorable again for growth. Furthermore, some evidence suggests that species with multiple *rrn* copies, of which many are genetically distinct from each other within the same genome (Pei et al., 2010), can selectively express different *rrn* operons as a survival strategy under stress conditions. For example, *Vibrio vulnificus* CMCP67, a gram-negative bacterium that possesses 9 copies of the *rrn* operon within its genome, preferentially transcribes *rrnI* under nutrient-limited conditions (Song et al., 2019). The *rrnI* operon of *V. vulnificus* is genetically divergent from its other *rrn* operons, and the expression of this operon led to the translation of specific mRNAs that allowed the strain to rapidly adapt to temperature and nutrient shifts (Song et al., 2019). This illustrates that, in principle, prokaryotes possessing higher *rrn* copies per genome can have additional genetic mechanisms enabling them to survive during stress conditions compared to species with only one or two *rrn* copies per genome. However, the relationship between *rrn* copy number and tolerance to specific stressors is not yet known.

Wet sieving of dried soil likely further released substrates during drying. As soils dry, the increasing ionic strength of shrinking water films stimulates the desorption of aromatic compounds from mineral surfaces which can then be released when soils are re-wet (Bailey et al., 2019). In the same study by Patel et al. (2021), re-wetted soils had a greater proportion of aromatic and lignin-like molecules in the water-extractable organic C pool, suggesting an increasing bioavailability of formerly adsorbed C that became solubilized with wetting. Moreover, the total concentration of water-extractable organic C decreased as dried soils were rewet, corresponding with an increase in CO_2 production and indicating that the eluted C was consumed by microbes as hydrologic pore connectivity was reestablished (Patel et al., 2021). This burst and subsequent consumption of nutrients likely explained the decrease in C and N with wet sieving we observed in this study. As the *rrn* copy numbers were correlated with aggregate C and N, the decrease in the average *rrn* copy number with wet sieving likely signifies a succession towards microbes that are adapted to oligotrophic environments as available nutrients become depleted (Nemergut et al., 2016).

Soil microaggregates are mainly formed by the interaction of mineral particles with organic matter (Totsche et al., 2017). As macroaggregates are composed of microaggregates (Six et al., 2000), it has been proposed that the stability of C in soil is driven by the physical occlusion of organic matter within macroaggregates (Dungait et al., 2012). Hence, compared to the microaggregate and silt & clay fractions that were originally free in the soil or resulted from aggregate breakdown, the C and N within the large and small macroaggregates remaining stable despite submersion in water are likely protected from microbial enzymes associated with macroaggregate surfaces (Bailey et al., 2012), potentially explaining the lack of changes observed in the macroaggregates compared to the two smallest size fractions. Overall, our findings are consistent with previous reports on the effects of water on *rrn* copy number and genome size (Gravuer and Eskelinen, 2017) and C and N (Sainju, 2006; Borken and Matzner,

2009), indicating that the consequences of dry and wet sieving on microbial communities in aggregates are similar to the effects of drying and wetting on bulk soil (Schimel, 2018).

Practical considerations

The wet sieving method was originally developed to obtain water-stable aggregates (Six et al., 2000) and has since become the standard for quantifying pools of soil C protected from turnover during slaking events (Six et al., 2000; Chenu and Cosentino, 2011; Bach and Hofmockel, 2014). However, soil microorganisms and their activities are sensitive to the fluctuations in moisture during preparation of soils for wet sieving. Recently, Bach and Hofmockel proposed drying soils to an "optimal moisture" threshold followed by dry sieving is the best strategy for interpreting *in situ* biological activities in aggregates (Bach and Hofmockel, 2014). This method, which involves first drying the soil to around 10% gravimetric water content, was designed to increase the reproducibility of aggregate separation and minimize impacts on the microbial community (Dorodnikov et al., 2009; Bach and Hofmockel, 2014). Our results provide additional support for this approach. When we quantified the separation of prokaryotic communities by aggregate size nested within each sieving treatment, we found that the effect sizes (R^2 values that indicate the proportion of variance in community composition explained) were lower in the wet sieving treatments (**Table S3-5**). A more pronounced effect was observed for the fungi with wet sieving, where the effect sizes were decreased and no differences in community composition were observed between aggregates when dried soils were wet sieved (**Table S3-5**). In other words, we found that while both dry and wet sieving were sufficient to identify differences in microbial communities between aggregates in our soils, the submersion of soil in water during the slaking process can reduce the differences in microbial communities between each aggregate size fraction. This may be due to erosion of smaller particles from aggregate surfaces, breakdown of unstable aggregates, or loss of microorganisms in the water

during wet sieving (Blaud et al., 2017), which can introduce artefacts to the smallest aggregate size fractions and potentially misrepresent the microbial communities detected in aggregates. Drying soils to desiccation prior to wet sieving can further add to this homogenizing effect and hence is not recommended, particularly for analyzing fungal communities in aggregates.

It is important to highlight that not all soils may benefit from an optimal moisture approach, and dry and wet sieving may affect the interpretation of microbial communities in aggregates differently from what we observed in our soils. For example, wet but not dry sieving differentiated bacterial community composition between aggregates of different size in a grassland soil but neither method showed any differences in an adjacent cropland soil (Blaud et al., 2017). Similarly, neither of the methods revealed differences in protistan communities within macro- and microaggregates in a fluvo-aquic soil (Liao et al., 2021). Furthermore, while wet sieving showed differences in microbial functional gene composition and diversity between aggregate size fractions in soils receiving inorganic mineral fertilizer, no differences between aggregates were found in soils amended with organic manure (Han et al., 2021). These results show that the decision to use dry or wet sieving, as well as the identification of the “optimal” moisture most useful for isolating aggregates from bulk soil will vary considerably based on soil texture, soil inputs, and land use history.

CONCLUSION

In conclusion, we found that dry and wet sieving had significant impacts on the microbial community in soil aggregates, with effects for prokaryotes and fungi in each aggregate size fraction reflecting different life strategies. While drying soils favored fast-growing generalist prokaryotes and fungal saprotrophs in aggregates, rewetting soils through wet sieving resulted in more slow-growing specialists and fungal pathogens in aggregates in tandem with decreasing aggregate C and N in the smallest aggregate size fractions. As soil microbiologists are becoming

increasingly interested in aggregates as the ecologically relevant scale to study microbial communities in the soil environment (Vos et al., 2013), we recommend that researchers consider their particular research question in choosing between methods as each is representative of the microbial community under strikingly different conditions of moisture saturation and disturbance. Dry sieving can be used to study microbial dynamics prior to natural events or management practices that can cause aggregate turnover such as rain, flooding, or tillage; wet sieving may be most useful for investigating microbial community shifts or enzymatic responses in aggregates to rapid wetting events. When combined, both dry and wet sieving can be used in experiments to simulate extreme fluctuations in moisture to investigate microbial community resistance and resilience in soil microenvironments.

ACKNOWLEDGEMENTS

We thank the staff of the Russell Ranch Sustainable Agriculture Facility at UC Davis for their support in the field study. We thank Daniel Rath for permission to use his illustrations in figure 1 and Vanessa Bailey for helpful comments on an earlier version of this manuscript. This research was supported in part by the USDA National Institute of Food and Agriculture (NIFA) through Hatch Formula Funding CA-D-LAW-2616-H, and a USDA NIFA AFRI Predoctoral Fellowship (2020-67034-31937) to JYL.

REFERENCES

Abel-Santos, E., 2015. Endospores, sporulation and germination, in: Tang, Y.-W., Sussman, M., Liu, D., Poxton, I., Schwartzman, J. (Eds.), Molecular Medical Microbiology (Second Edition). Academic Press, Boston, pp. 163–178.

Bach, E.M., Hofmockel, K.S., 2014. Soil aggregate isolation method affects measures of intra-aggregate extracellular enzyme activity. Soil Biology and Biochemistry 69, 54–62.

Bach, E.M., Williams, R.J., Hargreaves, S.K., Yang, F., Hofmockel, K.S., 2018. Greatest soil microbial diversity found in micro-habitats. Soil Biology and Biochemistry 118, 217–226.

Bailey, V.L., Bilskis, C.L., Fansler, S.J., McCue, L.A., Smith, J.L., Konopka, A., 2012. Measurements of microbial community activities in individual soil macroaggregates. Soil Biology and Biochemistry 48, 192–195.

Bailey, V.L., McCue, L.A., Fansler, S.J., Boyanov, M.I., DeCarlo, F., Kemner, K.M., Konopka, A., 2013. Micrometer-scale physical structure and microbial composition of soil macroaggregates. Soil Biology and Biochemistry 65, 60–68.

Bailey, V.L., Pries, C.H., Lajtha, K., 2019. What do we know about soil carbon destabilization? Environmental Research Letters 14, 083004.

Bapiri, A., Bååth, E., Rousk, J., 2010. Drying–rewetting cycles affect fungal and bacterial growth differently in an arable soil. Microbial Ecology 60, 419–428.

Barberán, A., Ramirez, K.S., Leff, J.W., Bradford, M.A., Wall, D.H., Fierer, N., 2014. Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. Ecology Letters 17, 794–802.

Barnard, R.L., Osborne, C.A., Firestone, M.K., 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. The ISME Journal 7, 2229–2241.

Barnes, C.J., van der Gast, C.J., McNamara, N.P., Rowe, R., Bending, G.D., 2018. Extreme rainfall affects assembly of the root-associated fungal community. New Phytologist 220, 1172–1184.

Birch, H.F., 1958. The effect of soil drying on humus decomposition and nitrogen availability. Plant and Soil 10, 9–31.

Blaud, A., Chevallier, T., Virto, I., Pablo, A.-L., Chenu, C., Brauman, A., 2014. Bacterial community structure in soil microaggregates and on particulate organic matter fractions located outside or inside soil macroaggregates. Pedobiologia 57, 191–194.

Blaud, A., Menon, M., van der Zaan, B., Lair, G.J., Banwart, S.A., 2017. Effects of dry and wet sieving of soil on identification and interpretation of microbial community composition, in: Sparks, S.A.B. and D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 119–142.

Blaud, A., van der Zaan, B., Menon, M., Lair, G.J., Zhang, D., Huber, P., Schiefer, J., Blum, W.E.H., Kitzler, B., Wei, E.H., van Gaans, P., Banwart, S., 2018. The abundance of nitrogen cycle genes and potential greenhouse gas fluxes depends on land use type and little on soil aggregate size. Applied Soil Ecology 125, 1–11.

Borken, W., Matzner, E., 2009. Reappraisal of drying and wetting effects on C and N mineralization and fluxes in soils. Global Change Biology 15, 808–824.

Caesar-Tonthat, T.C., 2002. Soil binding properties of mucilage produced by a basidiomycete fungus in a model system. Mycological Research 106, 930–937.

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

Cambardella, C.A., Elliott, E.T., 1993. Methods for physical separation and characterization of soil organic matter fractions. Geoderma 56, 449–457.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 6, 1621–1624.

Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., Fierer, N., 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nature Microbiology 2, 1–6.

Chenu, C., Cosentino, D., 2011. Microbial regulation of soil structural dynamics., in: Ritz, K., Young, I. (Eds.), The Architecture and Biology of Soils: Life in Inner Space. CABI, Wallingford, pp. 37–70.

Cregger, M.A., Schadt, C.W., McDowell, N.G., Pockman, W.T., Classen, A.T., 2012. Response of the soil microbial community to changes in precipitation in a semiarid ecosystem. Applied and Environmental Microbiology 78, 8587–8594.

Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak, J.C., Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. Soil Biology and Biochemistry 46, 63–72.

Daynes, C.N., Zhang, N., Saleeba, J.A., McGee, P.A., 2012. Soil aggregates formed in vitro by saprotrophic *Trichocomaceae* have transient water-stability. Soil Biology and Biochemistry 48, 151–161.

Dethlefsen, L., Schmidt, T.M., 2007. Performance of the translational apparatus varies with the ecological strategies of Bacteria. Journal of Bacteriology.

Dorodnikov, M., Blagodatskaya, E., Blagodatsky, S., Marhan, S., Fangmeier, A., Kuzyakov, Y., 2009. Stimulation of microbial extracellular enzyme activities by elevated CO_2 depends on soil aggregate size. Global Change Biology 15, 1603–1614.

Dungait, J.A.J., Hopkins, D.W., Gregory, A.S., Whitmore, A.P., 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. Global Change Biology 18, 1781–1796.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

Elliott, E.T., 1986. Aggregate structure and carbon, nitrogen, and phosphorus in native and cultivated soils. Soil Science Society of America Journal 50, 627–633.

Fansler, S.J., Smith, J.L., Bolton, H., Bailey, V.L., 2005. Distribution of two C cycle enzymes in soil aggregates of a prairie chronosequence. Biology and Fertility of Soils 42, 17–23.

Fierer, N., Bradford, M.A., Jackson, R.B., 2007. Toward an ecological classification of soil bacteria. Ecology 88, 1354–1364.

Filippidou, S., Wunderlin, T., Junier, T., Jeanneret, N., Dorador, C., Molina, V., Johnson, D.R., Junier, P., 2016. A combination of extreme environmental conditions favor the prevalence of endospore-forming *Firmicutes*. Frontiers in Microbiology 7, 1707.

Foster, R.C., 1988. Microenvironments of soil microorganisms. Biology and Fertility of Soils 6, 189–203.

Gravuer, K., Eskelinen, A., 2017. Nutrient and rainfall additions shift phylogenetically estimated traits of soil microbial communities. Frontiers in Microbiology 8, 1271.

Griffin, D.E., Wang, D., Parikh, S.J., Scow, K.M., 2017. Short-lived effects of walnut shell biochar on soils and crop yields in a long-term field experiment. Agriculture, Ecosystems & Environment 236, 21–29.

Gupta, V.V.S.R., Germida, J.J., 1988. Distribution of microbial biomass and its activity in different soil aggregate size classes as affected by cultivation. Soil Biology and Biochemistry 20, 777–786.

Han, S., Delgado-Baquerizo, M., Luo, X., Liu, Y., Van Nostrand, J.D., Chen, W., Zhou, J., Huang, Q., 2021. Soil aggregate size-dependent relationships between microbial functional diversity and multifunctionality. Soil Biology and Biochemistry 154, 108143.

Hattori, T., 1988. Soil aggregates as microhabitats of microorganisms. Reports of the Institute for Agricultural Research - Tohoku University (Japan).

Hawkes, C.V., Kivlin, S.N., Rocca, J.D., Huguet, V., Thomsen, M.A., Suttle, K.B., 2011. Fungal community responses to precipitation. Global Change Biology 17, 1637–1645.

Helgason, B.L., Walley, F.L., Germida, J.J., 2010. No-till soil management increases microbial biomass and alters community profiles in soil aggregates. Applied Soil Ecology 46, 390–397.

Huang, L., Chakrabarti, S., Cooper, J., Perez, A., John, S.M., Daroub, S.H., Martens-Habbena, W., 2021. Ammonia-oxidizing archaea are integral to nitrogen cycling in a highly fertile agricultural soil. ISME Communications 1, 1–12.

Ivarsson, M., Schnürer, A., Bengtson, S., Neubeck, A., 2016. Anaerobic fungi: A potential source of biological H_2 in the oceanic crust. Frontiers in Microbiology 7, 674.

Jastrow, J.D., Amonette, J.E., Bailey, V.L., 2007. Mechanisms controlling soil carbon turnover and their potential application for enhancing carbon sequestration. Climatic Change 80, 5–23.

Kanazawa, S., Filip, Z., 1986. Distribution of microorganisms, total biomass, and enzyme activities in different particles of brown soil. Microbial Ecology 12, 205–215.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464.

Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M., 2000. rRNA operon copy number reflects ecological strategies of bacteria. Applied and Environmental Microbiology. 66, 1328–1333.

Kong, A.Y.Y., Hristova, K., Scow, K.M., Six, J., 2010. Impacts of different N management regimes on nitrifier and denitrifier communities and N cycling in soil microenvironments. Soil Biology and Biochemistry 42, 1523–1533.

Krause, S., Le Roux, X., Niklaus, P.A., Bodegom, V., M, P., Lennon, J.T., Bertilsson, S., Grossart, H.-P., Philippot, L., Bodelier, P.L.E., 2014. Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. Frontiers in Microbiology 5.

Laliberté, E., Legendre, P., 2010. A distance-based framework for measuring functional diversity from multiple traits. Ecology 91, 299–305.

Lay, C.-Y., Hamel, C., St-Arnaud, M., 2018. Taxonomy and pathogenicity of *Olpidium brassicae* and its allied species. Fungal Biology 122, 837–846.

Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L., Harpole, W.S., Hobbie, S.E., Hofmockel, K.S., Knops, J.M.H., McCulley, R.L., Pierre, K.L., Risch, A.C., Seabloom, E.W., Schütz, M., Steenbock, C., Stevens, C.J., Fierer, N., 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proceedings of the National Academy of Sciences 112, 10967–10972.

Lehmann, A., Rillig, M.C., 2015. Understanding mechanisms of soil biota involvement in soil aggregation: A way forward with saprobic fungi? Soil Biology and Biochemistry 88, 298–302.

Lehmann, A., Zheng, W., Rillig, M.C., 2017. Soil biota contributions to soil aggregation. Nature Ecology & Evolution 1, 1828–1835.

Lehmann, A., Zheng, W., Ryo, M., Soutschek, K., Roy, J., Rongstock, R., Maaß, S., Rillig, M.C., 2020. Fungal traits important for soil aggregation. Frontiers in Microbiology 10, 2904.

Li, F., Qiu, P., Shen, B., Shen, Q., 2019. Soil aggregate size modifies the impacts of fertilization on microbial communities. Geoderma 343, 205–214.

Liao, H., Gao, S., Hao, X., Qin, F., Ma, S., Chen, W., Huang, Q., 2021. Soil aggregate isolation method affects interpretation of protistan community. Soil Biology and Biochemistry 161, 108388.

Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K., Allison, S.D., 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. The ISME Journal 14, 1–9.

Matsen, F.A., Kodner, R.B., Armbrust, E.V., 2010. pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11, 538.

Meisner, A., Jacquiod, S., Snoek, B.L., ten Hooven, F.C., van der Putten, W.H., 2018. Drought legacy effects on the composition of soil fungal and prokaryote communities. Frontiers in Microbiology 9.

Meisner, A., Snoek, B.L., Nesme, J., Dent, E., Jacquiod, S., Classen, A.T., Priemé, A., 2021. Soil microbial legacies differ following drying-rewetting and freezing-thawing cycles. The ISME Journal 15, 1207–1221. doi:10.1038/s41396-020-00844-3

Mummey, D., Holben, W., Six, J., Stahl, P., 2006. Spatial stratification of soil bacterial populations in aggregates of diverse soils. Microbial Ecology 51, 404–411.

Nemergut, D.R., Knelman, J.E., Ferrenberg, S., Bilinski, T., Melbourne, B., Jiang, L., Violle, C., Darcy, J.L., Prest, T., Schmidt, S.K., Townsend, A.R., 2016. Decreases in average bacterial community rRNA operon copy number during succession. The ISME Journal 10, 1147–1156.

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., Schilling, J.S., Kennedy, P.G., 2016. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. Fungal Ecology 20, 241–248.

Niederberger, T.D., Bottos, E.M., Sohm, J.A., Gunderson, T., Parker, A., Coyne, K.J., Capone, D.G., Carpenter, E.J., Cary, S.C., 2019. Rapid microbial dynamics in response to an induced wetting event in Antarctic dry valley soils. Frontiers in Microbiology 10, 621.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: Community ecology package.

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290.

Patel, K.F., Myers-Pigg, A., Bond-Lamberty, B., Fansler, S.J., Norris, C.G., McKever, S.A., Zheng, J., Rod, K.A., Bailey, V.L., 2021. Soil carbon dynamics during drying vs. rewetting: Importance of antecedent moisture conditions. Soil Biology and Biochemistry 156, 108165.

Pei, A.Y., Oberdorf, W.E., Nossa, C.W., Agarwal, A., Chokshi, P., Gerz, E.A., Jin, Z., Lee, P., Yang, L., Poles, M., Brown, S.M., Sotero, S., DeSantis, T., Brodie, E., Nelson, K., Pei, Z., 2010. Diversity of 16S rRNA genes within individual prokaryotic genomes. Applied and Environmental Microbiology 76, 3886–3897.

Placella, S.A., Brodie, E.L., Firestone, M.K., 2012. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. Proceedings of the National Academy of Sciences 109, 10931–10936.

Ramakrishnan, B., Lueders, T., Conrad, R., Friedrich, M., 2000. Effect of soil aggregate size on methanogenesis and archaeal community structure in anoxic rice field soil. FEMS Microbiology Ecology 32, 261–270.

Rillig, M.C., Mardatin, N.F., Leifheit, E.F., Antunes, P.M., 2010. Mycelium of arbuscular mycorrhizal fungi increases soil water repellency and is sufficient to maintain water-stable soil aggregates. Soil Biology and Biochemistry 42, 1189–1191.

Rillig, M.C., Muller, L.A., Lehmann, A., 2017. Soil aggregates as massively concurrent evolutionary incubators. The ISME Journal 11, 1943–1948.

Roller, B.R., Schmidt, T.M., 2015. The physiology and ecological implications of efficient growth. The ISME Journal 9, 1481–1487.

Sainju, U.M., 2006. Carbon and nitrogen pools in soil aggregates separated by dry and wet sieving methods. Soil Science 171, 937.

Schimel, J.P., 2018. Life in dry soils: Effects of drought on soil microbial communities and processes. Annual Review of Ecology, Evolution, and Systematics 49, 409–432.

Seena, S., Marvanová, L., Letourneau, A., Bärlocher, F., 2018. *Articulospora* – phylogeny vs morphology. Fungal Biology 122, 965–976.

Sexstone, A.J., Revsbech, N.P., Parkin, T.B., Tiedje, J.M., 1985. Direct measurement of oxygen profiles and denitrification rates in soil aggregates. Soil Science Society of America Journal 49, 645–651.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. Soil and Tillage Research, Advances in Soil Structure Research 79, 7–31.

Six, J, Elliott, E.T., Paustian, K., 2000. Soil macroaggregate turnover and microaggregate formation: A mechanism for C sequestration under no-tillage agriculture. Soil Biology and Biochemistry 32, 2099–2103.

Six, J., Paustian, K., Elliott, E.T., Combrink, C., 2000. Soil structure and organic matter I. Distribution of aggregate-size classes and aggregate-associated carbon. Soil Science Society of America Journal 64, 681–689.

Smith, D.P., Peay, K.G., 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLOS ONE 9, e90234.

Song, W., Joo, M., Yeom, J.-H., Shin, E., Lee, M., Choi, H.-K., Hwang, J., Kim, Y.-I., Seo, R., Lee, J.E., Moore, C.J., Kim, Y.-H., Eyun, S., Hahn, Y., Bae, J., Lee, K., 2019. Divergent rRNAs as regulators of gene expression at the ribosome level. Nature Microbiology 4, 515–526.

Stark, J.M., Firestone, M.K., 1995. Mechanisms for soil moisture effects on activity of nitrifying bacteria. Applied and Environmental Microbiology 61, 218–221.

Tisdall, J.M., Nelson, S.E., Wilkinson, K.G., Smith, S.E., McKenzie, B.M., 2012. Stabilisation of soil against wind erosion by six saprotrophic fungi. Soil Biology and Biochemistry 50, 134–141.

Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. Journal of Soil Science 33, 141–163.

Totsche, K.U., Amelung, W., Gerzabek, M.H., Guggenberger, G., Klumpp, E., Knief, C., Lehndorff, E., Mikutta, R., Peth, S., Prechtel, A., Ray, N., Kögel-Knabner, I., 2017. Microaggregates in soils. Journal of Plant Nutrition and Soil Science 181, 104–146.

Trivedi, P., Delgado-Baquerizo, M., Jeffries, T.C., Trivedi, C., Anderson, I.C., Lai, K., McNee, M., Flower, K., Pal Singh, B., Minkey, D., Singh, B.K., 2017. Soil aggregation and associated microbial communities modify the impact of agricultural management on carbon content. Environmental Microbiology 19, 3070–3086.

Trivedi, P., Rochester, I.J., Trivedi, C., Van Nostrand, J.D., Zhou, J., Karunaratne, S., Anderson, I.C., Singh, B.K., 2015. Soil aggregate size mediates the impacts of cropping regimes on soil carbon and microbial communities. Soil Biology and Biochemistry 91, 169–181.

Upton, R.N., Bach, E.M., Hofmockel, K.S., 2019. Spatio-temporal microbial community dynamics within soil aggregates. Soil Biology and Biochemistry.

Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G.A., 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiology Reviews 37, 936–954.

Wagner, D., Eisenhauer, N., Cesarz, S., 2015. Plant species richness does not attenuate responses of soil microbial and nematode communities to a flood event. Soil Biology and Biochemistry 89, 135–149.

Wang, D., Lin, J.Y., Sayre, J.M., Schmidt, R., Fonte, S.J., Rodrigues, J.L.M., Scow, K.M., 2022. Compost amendment and cover cropping maintains soil structure and carbon storage by

increasing available carbon and microbial biomass in agricultural soil – a six-year field study. Under Review.

Wang, Y., Hu, N., Ge, T., Kuzyakov, Y., Wang, Z.-L., Li, Z., Tang, Z., Chen, Y., Wu, C., Lou, Y., 2017. Soil aggregation regulates distributions of carbon, microbial community and enzyme activities after 23-year manure amendment. Applied Soil Ecology 111, 65–72.

Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., RStudio, 2020. ggplot2: Create elegant data visualisations using the grammar of graphics.

Wilpiszeski, R.L., Aufrecht, J.A., Retterer, S.T., Sullivan, M.B., Graham, D.E., Pierce, E.M., Zablocki, O.D., Palumbo, A.V., Elias, D.A., 2019. Soil aggregate microbial communities: Towards understanding microbiome interactions at biologically relevant scales. Applied and Environmental Microbiology 85, e00324-19.

Yadav, D.R., Kim, S.W., Adhikari, M., Um, Y.H., Kim, H.S., Kim, C., Lee, H.B., Lee, Y.S., 2015. Three new records of *Mortierella* species isolated from crop field soil in Korea. Mycobiology 43, 203–209.

Yu, H., Ding, W., Chen, Z., Zhang, H., Luo, J., Bolan, N., 2015. Accumulation of organic C components in soil and aggregates. Scientific Reports 5, 13804.

TABLES AND FIGURES

Table 3-1. Permutational multivariate analysis of variance (PERMANOVA) results for prokaryotic (a) and fungal (b) communities in soil aggregates.

a)

Prokaryotes					
Variable	**Df**	**SumSqs**	**R^2**	**F**	**P-value**
Sieving Treatment	3	1.445	0.124	6.818	***0.001**
Aggregate Size	3	0.762	0.065	3.595	***0.001**
Fertilizer	1	0.664	0.057	9.396	***0.001**
Sieving Treatment:Aggregate Size	9	0.928	0.079	1.460	***0.001**
Sieving Treatment:Fertilizer	3	0.281	0.024	1.326	***0.024**
Aggregate Size:Fertilizer	3	0.265	0.023	1.250	0.063
Sieving Treatment:Aggregate Size:Fertilizer	9	0.638	0.055	1.004	0.407
Residuals	95	6.712	0.574		
Total	126	11.695	1.000		

b)

Fungi					
Variable	**Df**	**SumSqs**	**R^2**	**F**	**P-value**
Sieving Treatment	3	1.403	0.098	5.428	***0.001**
Aggregate Size	3	1.755	0.122	6.789	***0.001**
Fertilizer	1	0.368	0.026	4.272	***0.003**
Sieving Treatment:Aggregate Size	9	1.526	0.106	1.968	***0.001**
Sieving Treatment:Fertilizer	3	0.311	0.022	1.203	0.184
Aggregate Size:Fertilizer	3	0.332	0.023	1.282	0.164
Sieving Treatment:Aggregate Size:Fertilizer	9	0.804	0.056	1.037	0.361
Residuals	91	7.842	0.547		
Total	122	14.342	1.000		

Figure 3-1. Conceptual diagram showing experimental manipulations (a). Conceptual illustrations of the different aggregate size fractions analyzed in this study (b). Aggregate illustrations used with permission from D. Rath.

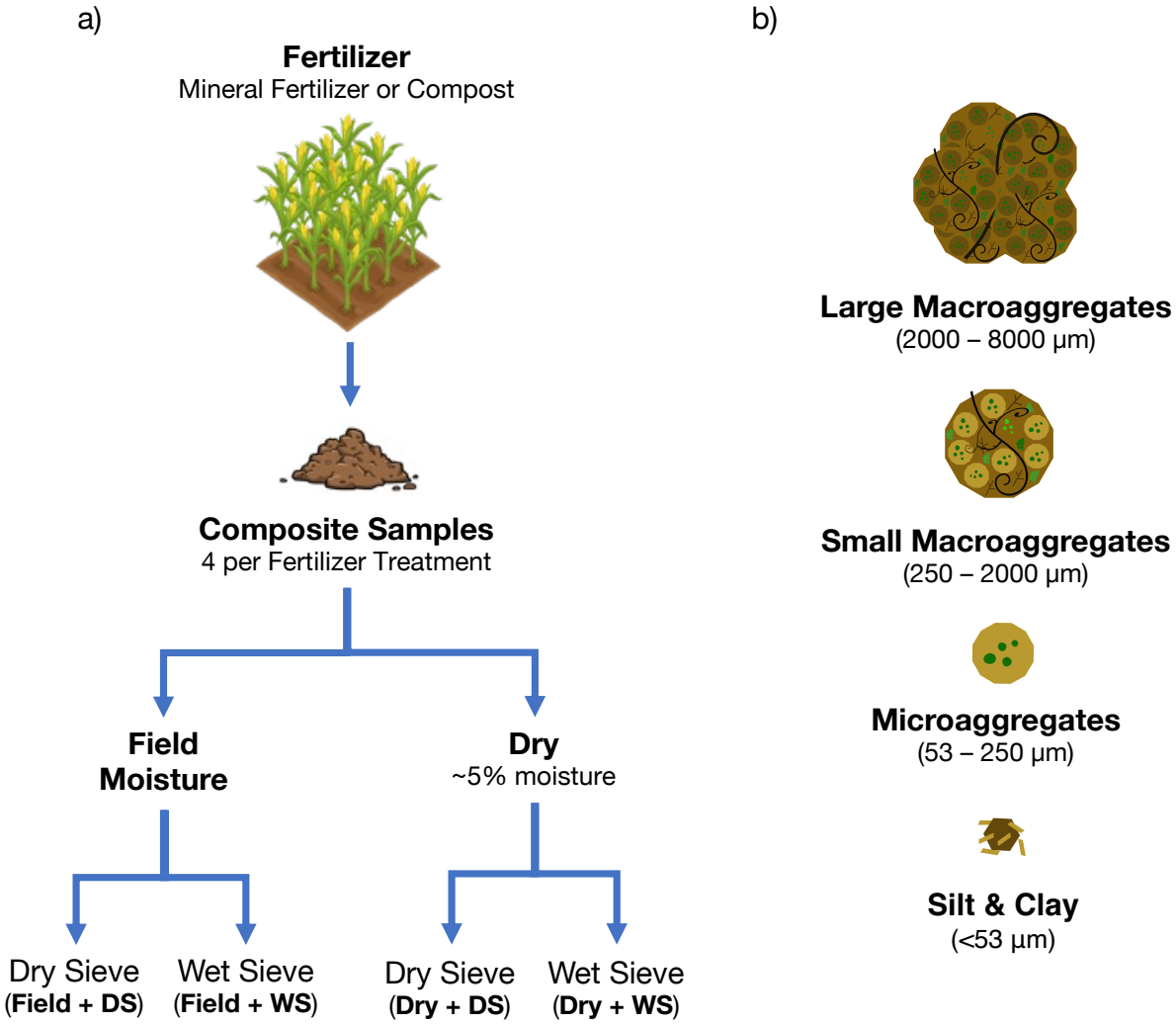


Figure 3-2. Non-metric multidimensional scaling (NMDS) plots based on Bray-Curtis distances of prokaryotic (a) and fungal (b) communities in soil aggregates. Samples are colored by sieving treatment and shapes correspond with different aggregate size fractions. Ellipses represent 95% confidence intervals for each treatment. Ordination stress values are 0.187 and 0.115 for prokaryotes and fungi, respectively. DS, dry sieving; WS, wet sieving.

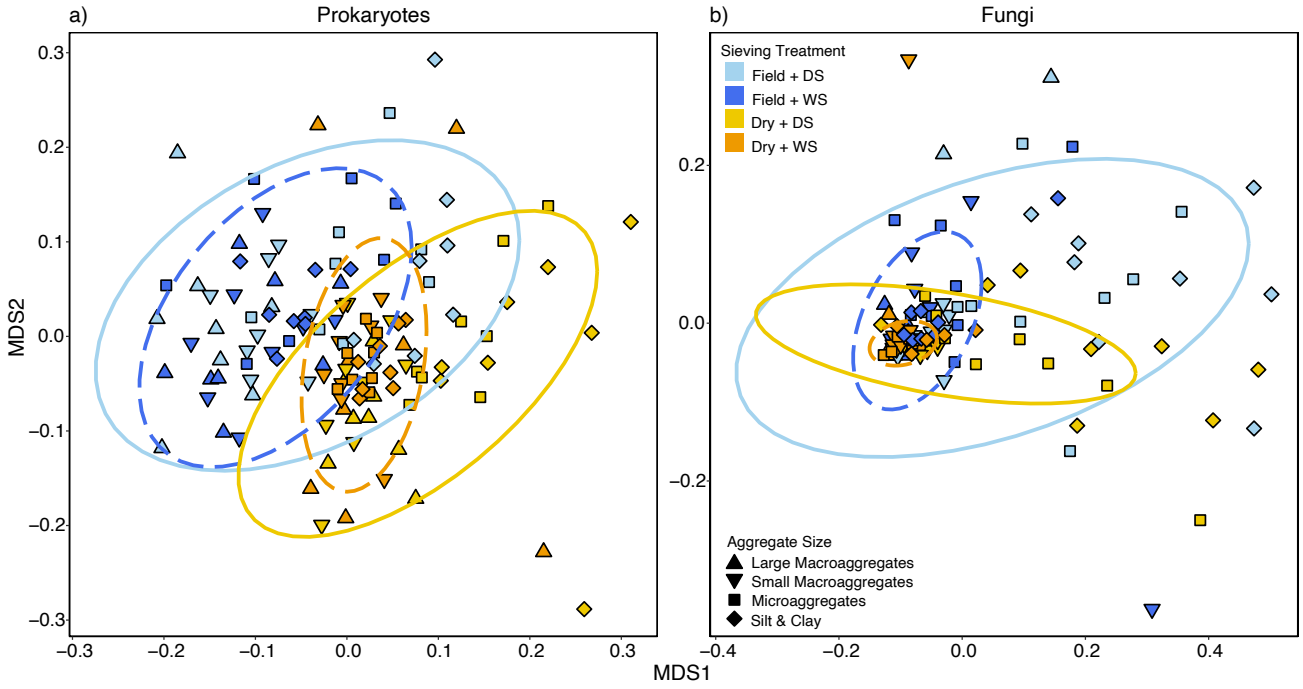


Figure 3-3. Non-metric multidimensional scaling (NMDS) plots based on Bray-Curtis distances of prokaryotic (a-d) and fungal (e-h) communities in different aggregate size fractions. Samples are colored by sieving treatment and shapes correspond with different aggregate size fractions. PERMANOVA R^2 and *P*-values for treatment are presented in each panel, and statistically significant ($P < 0.05$) results are indicated in bold. DS, dry sieving; WS, wet sieving.

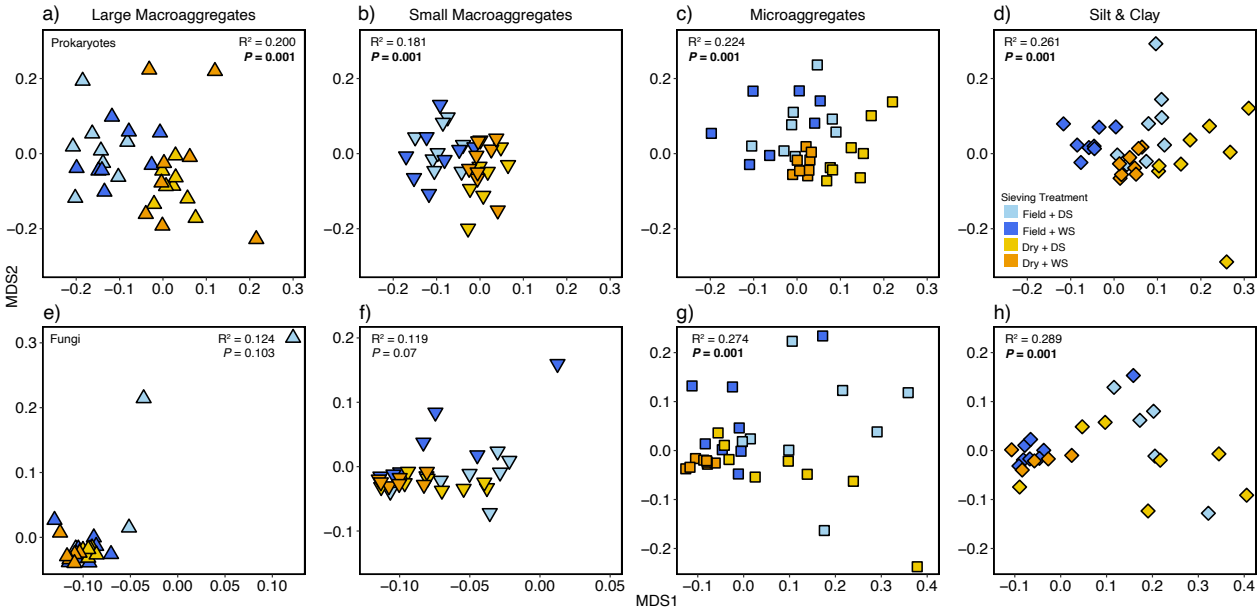


Figure 3-4. Alpha diversity based on Richness (a-d) and Shannon index (e-h) metrics of prokaryotic communities in different soil aggregate size fractions. Samples are colored by sieving treatment and different letters represent significant differences ($P < 0.05$) between treatments. DS, dry sieving; WS, wet sieving.

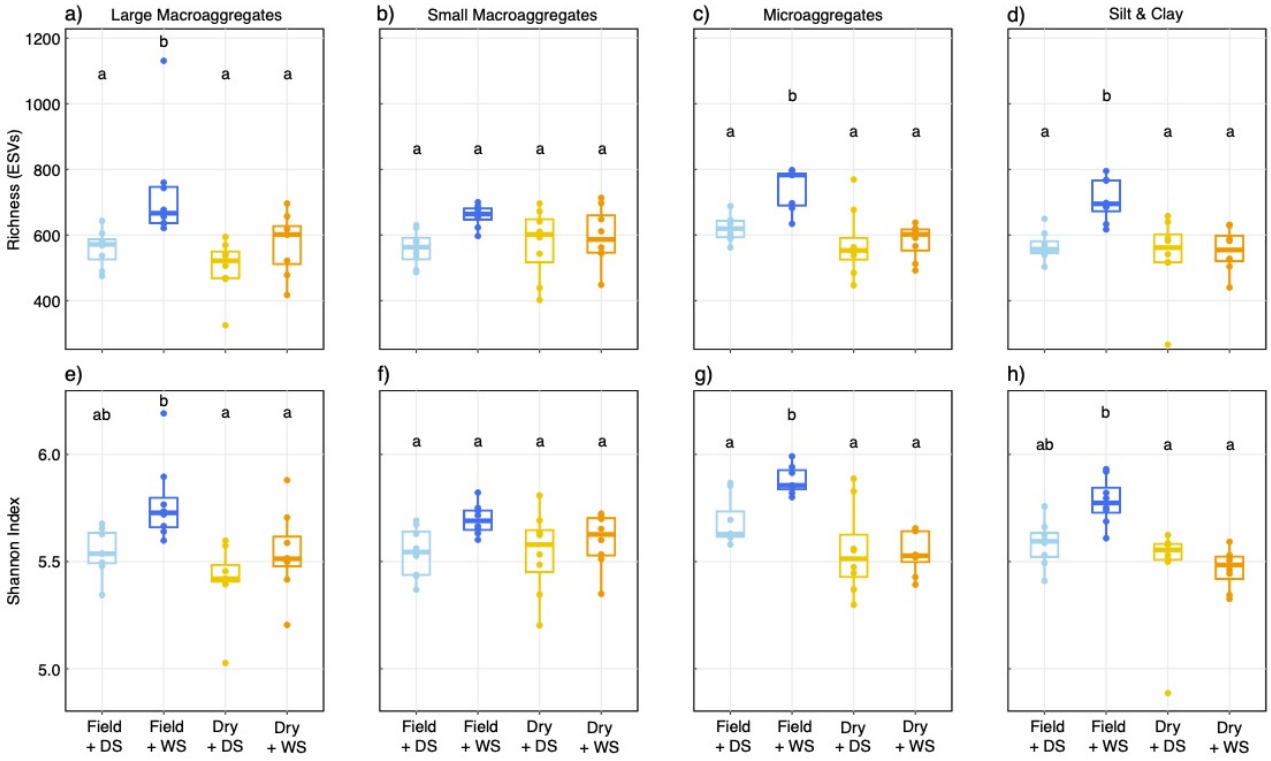


Figure 3-5. Alpha diversity based on Richness (a-d) and Shannon index (b-h) metrics of fungal communities in different soil aggregate size fractions. Samples are colored by sieving treatment and different letters represent significant differences ($P < 0.05$) between treatments. DS, dry sieving; WS, wet sieving.

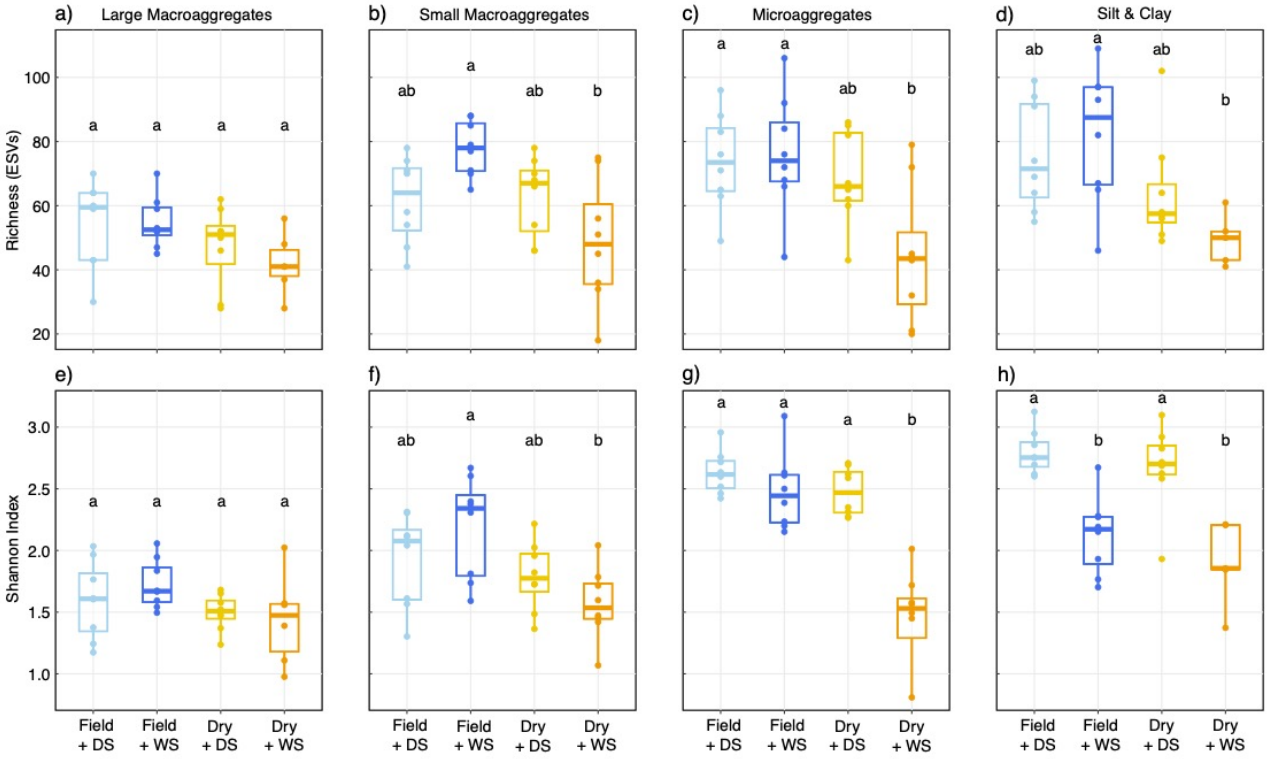


Figure 3-6. Community-weighted mean estimated traits for ribosomal RNA gene (*rrn*) copy number (a-d) and genome size (e-h) in different soil aggregate size fractions. Traits were estimated from 16S rRNA gene sequences and samples are colored by sieving treatments. Different letters represent significant differences ($P < 0.05$) between treatments. DS, dry sieving; WS, wet sieving.

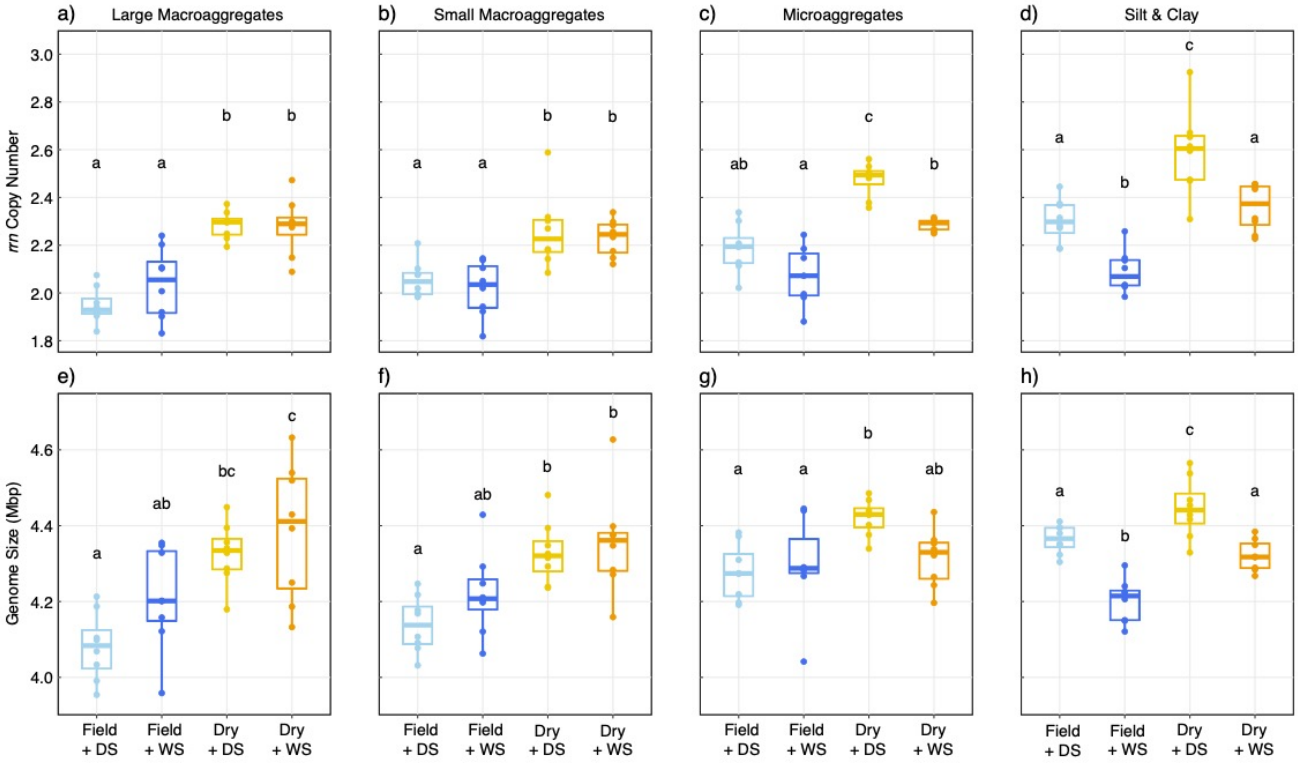


Figure 3-7. Relative abundance of fungal taxa assigned into different trophic groups (a-c) and ecological guilds (d-f) by treatment. Functional groups were assigned from ITS sequences and samples are colored by sieving treatment. Different letters represent significant differences ($P < 0.05$) between treatments. DS, dry sieving; WS, wet sieving.

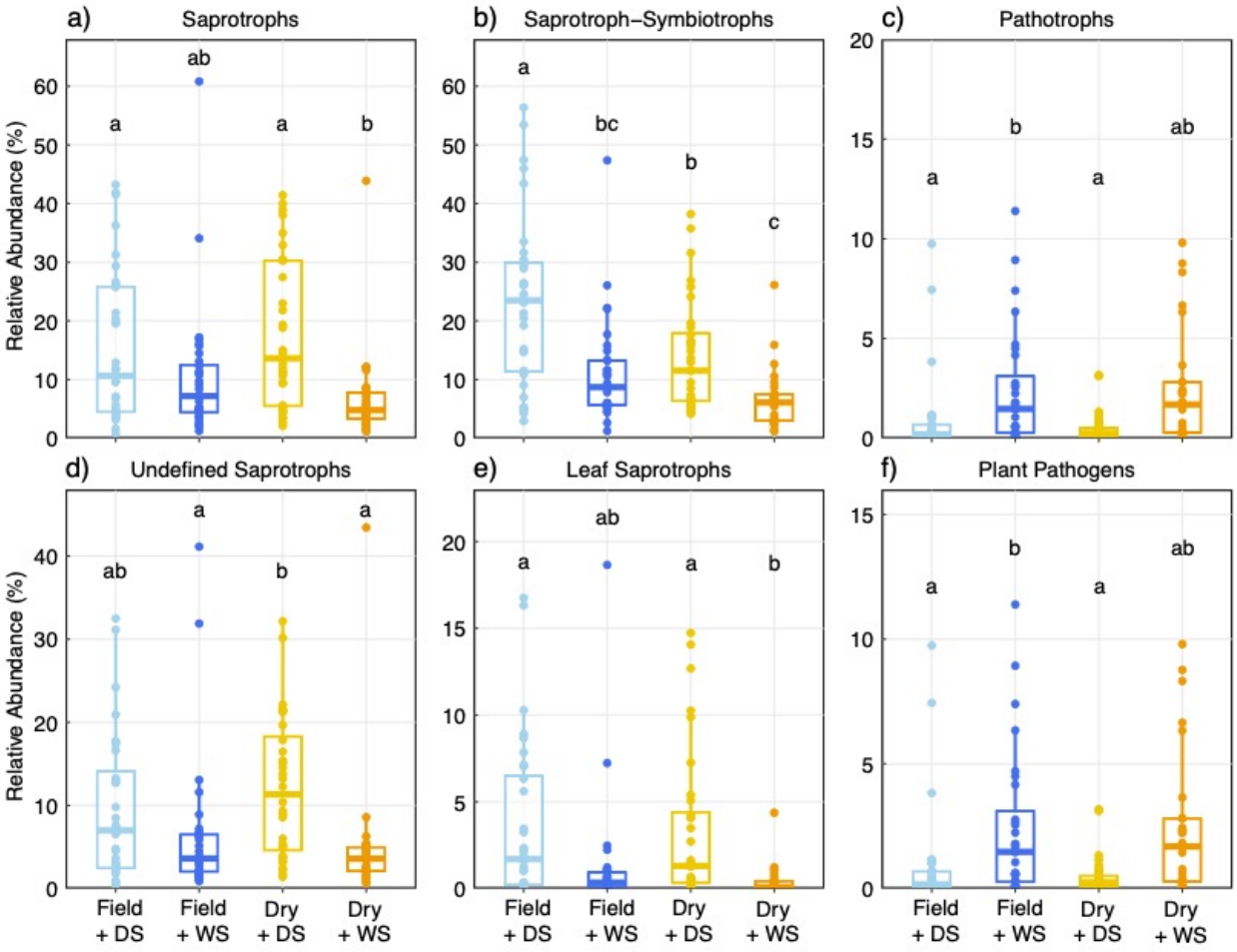
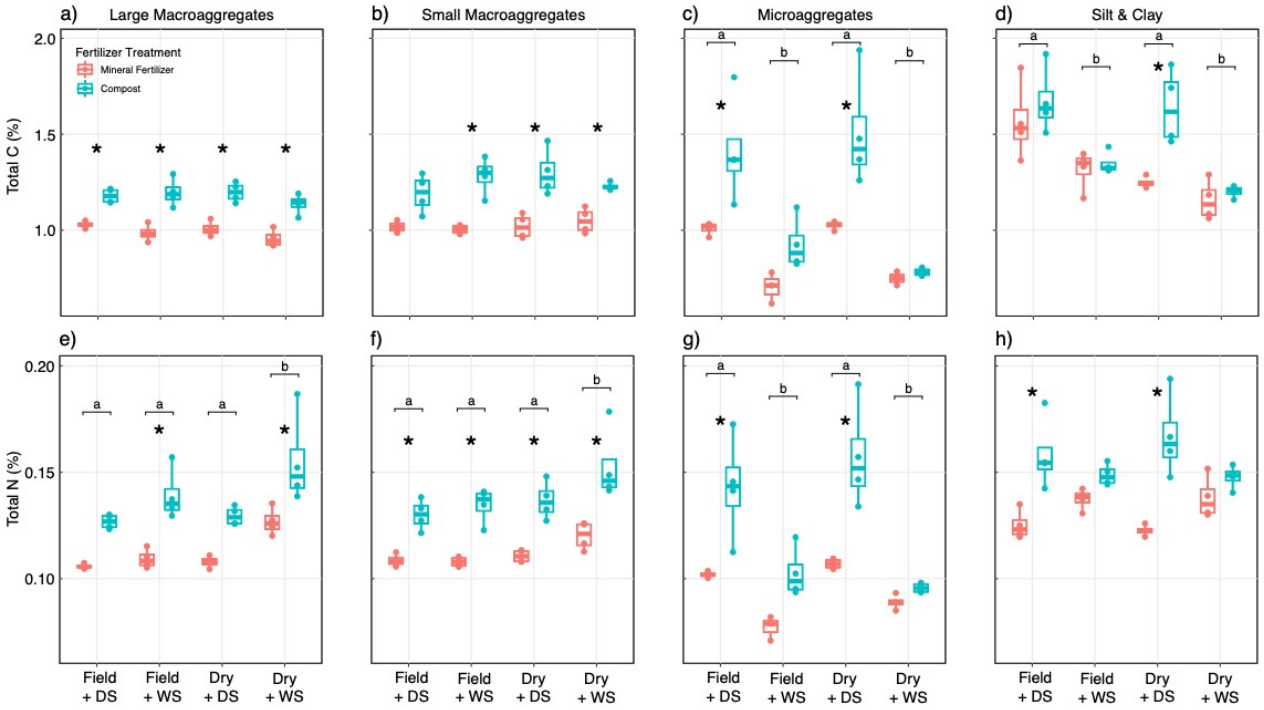


Figure 3-8. Total C and N concentrations in soil aggregate size fractions under different treatments. Samples are colored by fertilizer treatments. Asterisks indicate statistically significant ($P < 0.05$) differences between fertilizer regimes within each treatment group and different letters indicate significant differences between sieving treatments. DS, dry sieving; WS, wet sieving.



CHAPTER 2

SUPPLEMENTARY MATERIAL

Table S3-1. Gravimetric water content (%) of bulk soils under field-moisture conditions and after drying used in this study. Plot numbers indicate the block within the field trial at the UC Davis Russell Ranch Sustainable Agriculture Facility.

Plot	Fertilizer	Gravimetric Water Content (%)	
		Field	Dry
2	Compost	17.46	4.72
3	Mineral Fertilizer	16.83	4.78
5	Compost	17.97	4.71
7	Mineral Fertilizer	17.53	4.80
9	Mineral Fertilizer	17.40	4.72
12	Compost	17.36	4.81
13	Mineral Fertilizer	17.65	4.80
14	Compost	17.37	4.85

Table S3-2. Characteristics of bulk soils used this study. Values represent the mean ± standard error of the measured soil parameter and were obtained from 4 field replicates per treatment at the UC Davis Russell Ranch Sustainable Agriculture Facility. *P*-values indicating statistically significant ($P < 0.05$) differences between fertilizer treatments are in bold.

Physicochemical Parameter	Fertilizer		
	Mineral Fertilizer	Compost	*P*-value
Organic Matter (%)	2.43 ± 0.14	2.70 ± 0.06	0.13
Estimated Nitrogen Release (kg per hectare)	14.36 ± 0.00	15.33 ± 0.16	0.13
P Weak Bray (ppm)	27.25 ± 1.11	34.00 ± 3.92	0.15
$NaHCO_3$-P Olsen (ppm)	25.75 ± 1.44	30.25 ± 1.49	0.07
K (ppm)	177.50 ± 6.85	227.00 ± 10.51	**<0.01**
K (% cation saturation)	2.15 ± 0.05	2.85 ± 0.13	**<0.01**
Mg (ppm)	1576.25 ± 37.16	1426.75 ± 26.80	**<0.05**
Mg (% cation saturation)	60.93 ± 0.13	57.63 ± 0.70	**<0.01**
Ca (ppm)	1531.50 ± 36.77	1563.00 ± 21.21	0.49
Ca (% cation saturation)	35.93 ± 0.13	38.30 ± 0.62	**<0.01**
Na (ppm)	48.75 ± 0.25	55.75 ± 1.49	**<0.01**
Na (% cation saturation)	1.00 ± 0.0	1.20. ± 0.04	**<0.01**
pH	7.23 ± 0.05	7.30 ± 0.04	0.28
Cation Exchange Capacity (meq per 100 g)	21.25 ± 0.05	21.50 ± 0.21	0.14
NO_3-N (ppm)	16.50 ± 0.96	21.50 ± 1.50	**<0.05**
SO_4-S (ppm)	3.00 ± 0.41	5.25 ± 0.48	**<0.05**
Soluble Salts (mmhos per cm)	0.30 ± 0.0	0.43 ± 0.02	**<0.01**

Table S3-3. Relative abundances (%) of the 20 most abundant prokaryotic taxa at the genus level based on the 16S rRNA marker gene. Different letters indicate statistically significant ($P < 0.05$) differences between sieving treatments. DS, dry sieving; WS, wet sieving.

Taxonomy			Sieving Treatment			
Phylum	**Order**	**Genus**	**Field + DS**	**Field + WS**	**Dry + DS**	**Dry + WS**
Thaumarchaeota	*Nitrososphaerales*	*Candidatus Nitrososphaera*	7.27 a	8.18 ab	4.37 c	5.64 a
Actinobacteria	*Micrococcales*	*Pseudarthrobacter*	6.15 a	5.80 a	6.62 a	5.76 a
Proteobacteria	*Sphingomonadales*	*Sphingomonas*	6.66 ab	5.68 bc	5.16 c	6.76 a
Actinobacteria	*Rubrobacterales*	*Rubrobacter*	6.34 a	5.81 a	5.14 a	6.01 a
Firmicutes	*Bacillales*	*Bacillus*	3.55 a	3.09 a	6.38 b	6.55 b
Proteobacteria	*Hyphomicrobiales*	*Microvirga*	4.44 a	4.62 a	4.40 a	5.37 b
Acidobacteria	*Pyrinomonadales*	*Pyrinomonadaceae RB41*	4.47 a	3.99 a	3.35 a	3.49 a
Actinobacteria	*Geodermatophilales*	*Blastococcus*	2.36 a	2.71 a	3.86 b	3.72 b
Actinobacteria	*Solirubrobacterales*	*Solirubrobacter*	2.49 a	3.44 b	3.07 ab	2.93 ab
Actinobacteria	*Streptomycetales*	*Streptomyces*	2.63 ab	2.42 b	2.95 a	2.80 ab
Actinobacteria	*Gaiellales*	*Gaiella*	2.37 a	2.70 ab	2.63 ab	2.94 b
Actinobacteria	*Propionibacteriales*	*Nocardioides*	1.81 a	1.81 a	3.27 b	2.47 b
Proteobacteria	*Xanthomonadales*	*Lysobacter*	1.71 a	1.62 a	1.55 a	1.69 a
Bacteroidetes	*Cytophagales*	*Adhaeribacter*	1.52 a	1.33 a	2.14 b	1.46 a
Acidobacteria	*Geodermatophilales*	*Geodermatophilus*	1.00 a	1.11 ab	1.34 ab	1.43 b
Proteobacteria	*Burkholderiales*	*Massilia*	1.36 ab	0.92 bc	1.75 a	0.83 c
Proteobacteria	*Hyphomicrobiales*	*Psychroglaciecola*	1.15 ab	1.23 ab	1.03 a	1.27 b
Actinobacteria	*Solirubrobacterales*	*Conexibacter*	0.88 a	1.28 b	1.31 b	1.19 ab
Verrucomicrobia	*Cthoniobacterales*	*Chthoniobacter*	1.09 a	1.08 a	1.09 a	1.16 a
Proteobacteria	*Steroidobacterales*	*Steroidobacter*	1.23 a	1.50 a	0.66 c	0.92 d

Table S3-4. Relative abundances (%) of the 20 most abundant fungal taxa at the genus level based on the ITS region. Different letters indicate statistically significant ($P < 0.05$) differences between sieving treatments. DS, dry sieving; WS, wet sieving.

Taxonomy			Sieving Treatment			
Phylum	**Order**	**Genus**	**Field + DS**	**Field + WS**	**Dry + DS**	**Dry + WS**
Basidiomycota	*Filobasidiales*	*Solicoccozyma*	53.83 a	68.87 a	61.79 a	79.68 b
Mortierellomycota	*Mortierellales*	*Mortierella*	24.28 a	12.09 bc	14.14 b	6.98 c
Basidiomycota	*Cystofilobasidiales*	*Cystofilobasidium*	3.76 ab	1.39 bc	3.41 a	0.44 c
Basidiomycota	*Filobasidiales*	*Filobasidium*	2.44 ab	1.27 c	3.90 a	0.44 bc
Olpidiomycota	*Olpidiales*	*Olpidium*	0.91 a	3.29 b	0.44 a	2.51 b
Ascomycota	*Pleosporales*	*Lepidosphaeria*	1.67 ab	0.93 b	2.23 a	1.46 ab
Ascomycota	*Helotiales*	*Articulospora*	2.94 a	0.95 ab	1.84 a	0.26 b
Basidiomycota	*Tremellales*	*Papiliotrema*	0.82 a	1.16 ab	1.23 b	1.22 b
Ascomycota	*Thelebolales*	*Pseudogymnoascus*	0.98 a	1.76 b	0.72 a	0.49 a
Ascomycota	*Sordariales*	*Humicola*	0.67 a	0.82 a	1.01 a	1.17 a
Basidiomycota	*Agaricales*	*Agrocybe*	1.08 a	0.75 a	1.31 a	0.21 a
Ascomycota	*Pleosporales*	*Preussia*	1.04 a	0.16 b	1.58 a	0.35 b
Ascomycota	*Helotiales*	*Cyathicula*	0.35 a	0.52 a	0.66 a	1.68 a
Ascomycota	*Helotiales*	*Tetracladium*	0.51 ab	0.57 ab	0.95 a	0.19 b
Ascomycota	*Chaetothyriales*	*Exophiala*	0.55 a	0.68 a	0.36 a	0.31 a
Basidiomycota	*Agaricales*	*Conocybe*	0.22 a	1.07 a	0.45 a	0.13 a
Basidiomycota	*Wallemiales*	*Wallemia*	0.31 a	0.75 a	0.31 a	0.05 b
Ascomycota	*Pezizales*	*Cephaliophora*	0.17 a	0.53 b	0.22 ab	0.42 b
Ascomycota	*Pleosporales*	*Pyrenochaetopsis*	0.71 a	0.25 a	0.10 a	0.07 a
Basidiomycota	*Cystofilobasidiales*	*Itersonilia*	0.10 a	0.10 ab	0.50 a	0.02 b

Table S3-5. Permutational multivariate analysis of variance (PERMANOVA) results for prokaryotic and fungal communities in soil aggregates. The dataset was separated to determine the effect of aggregate size in explaining microbial community composition nested in each sieving treatment. DS, dry sieving; WS, wet sieving.

Community	Sieving Treatment	Variable	Df	SumSqs	R^2	F	P-value
Prokaryotes	Field + DS	Aggregate Size	3	0.543	0.195	2.261	***0.001**
	Field + WS	Aggregate Size	3	0.342	0.141	1.478	***0.004**
	Dry + DS	Aggregate Size	3	0.511	0.185	2.114	***0.001**
	Dry + WS	Aggregate Size	3	0.296	0.131	1.403	***0.008**
Fungi	Field + DS	Aggregate Size	3	1.380	0.287	3.762	***0.001**
	Field + WS	Aggregate Size	3	0.460	0.151	1.655	***0.006**
	Dry + DS	Aggregate Size	3	1.251	0.331	4.620	***0.001**
	Dry + WS	Aggregate Size	3	0.187	0.143	1.284	0.071

Figure S3-1. Homogeneity of within-group dispersions based on Bray-Curtis distances of prokaryotic (a) and fungal (b) communities by treatment across all samples. Different letters indicate statistically significant ($P < 0.05$) differences between sieving treatments. DS, dry sieving; WS, wet sieving.

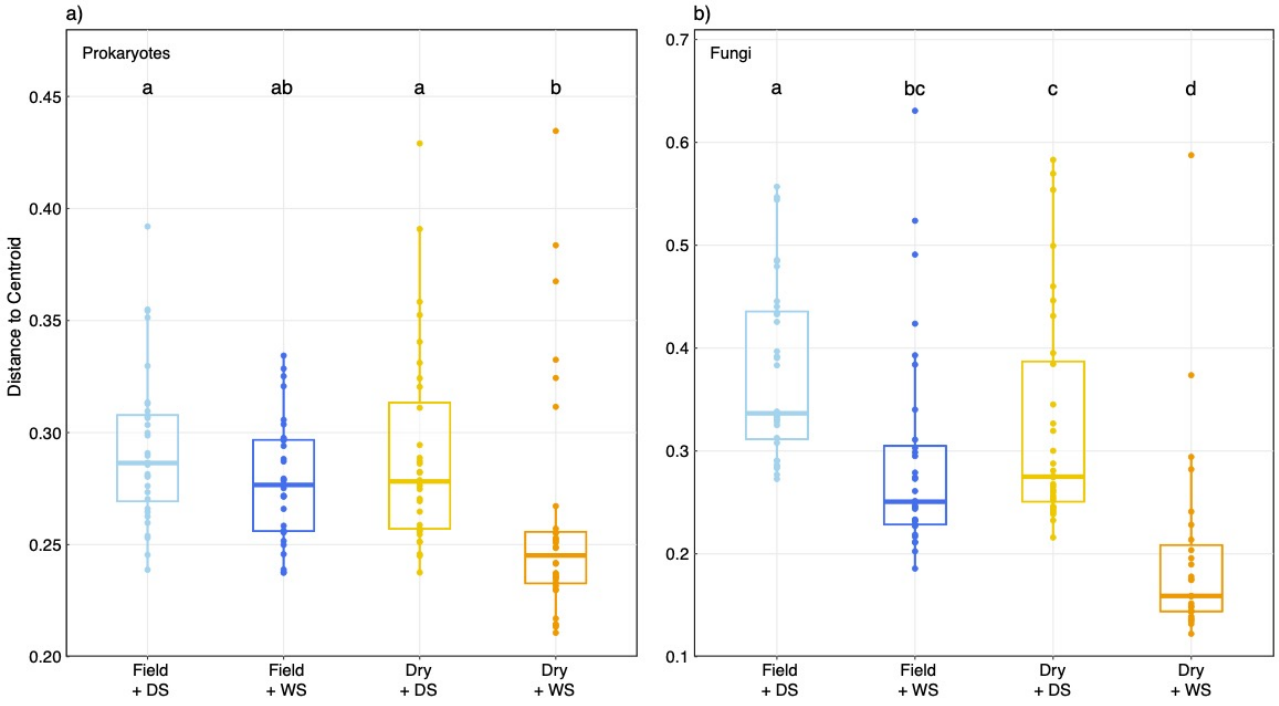


Figure S3-2. Alpha diversity based on Richness (a & c) and Shannon index (b & d) metrics of prokaryotic and fungal communities by treatment across all samples. Different letters represent significant differences ($P < 0.05$) between sieving treatments. DS, dry sieving; WS, wet sieving.

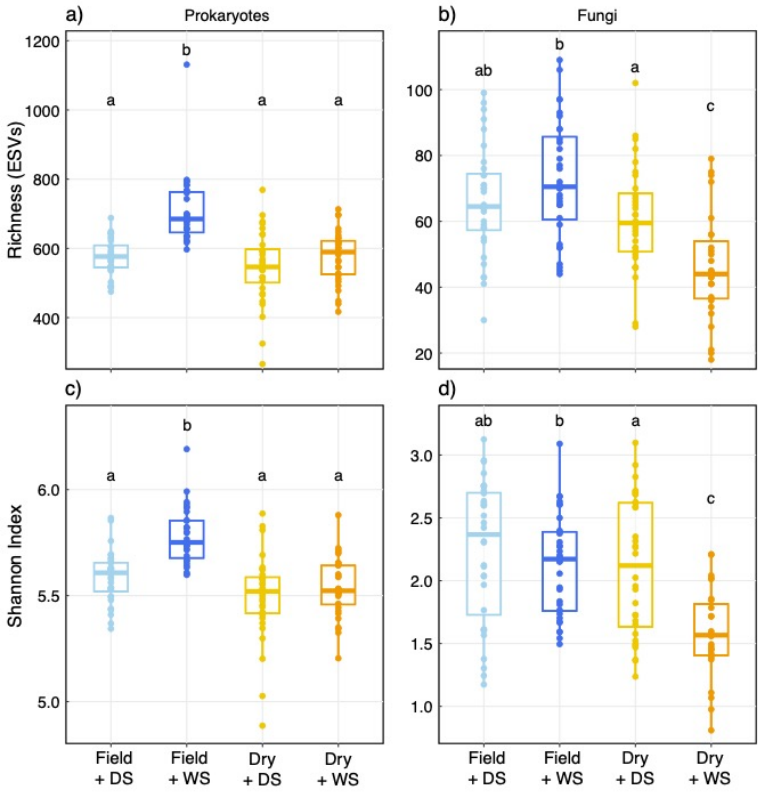


Figure S3-3. Bubble chart of the 10 most abundant prokaryotic phyla by sieving treatment based on the 16S rRNA marker gene. Each bubble is scaled by size according to relative abundance and colored by treatment. DS, dry sieving; WS, wet sieving.

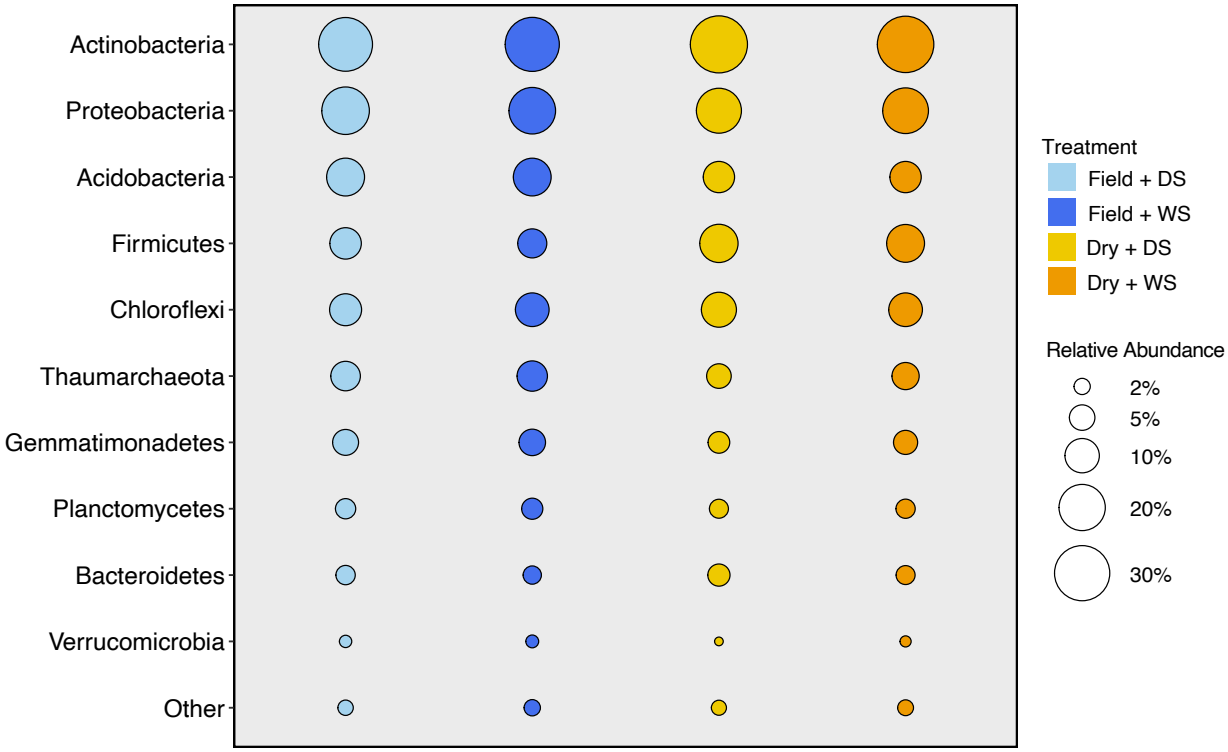


Figure S3-4. Bubble chart of the most abundant fungal phyla by sieving treatment based on the ITS region. Each bubble is scaled by size according to relative abundance and colored by treatment. DS, dry sieving; WS, wet sieving.

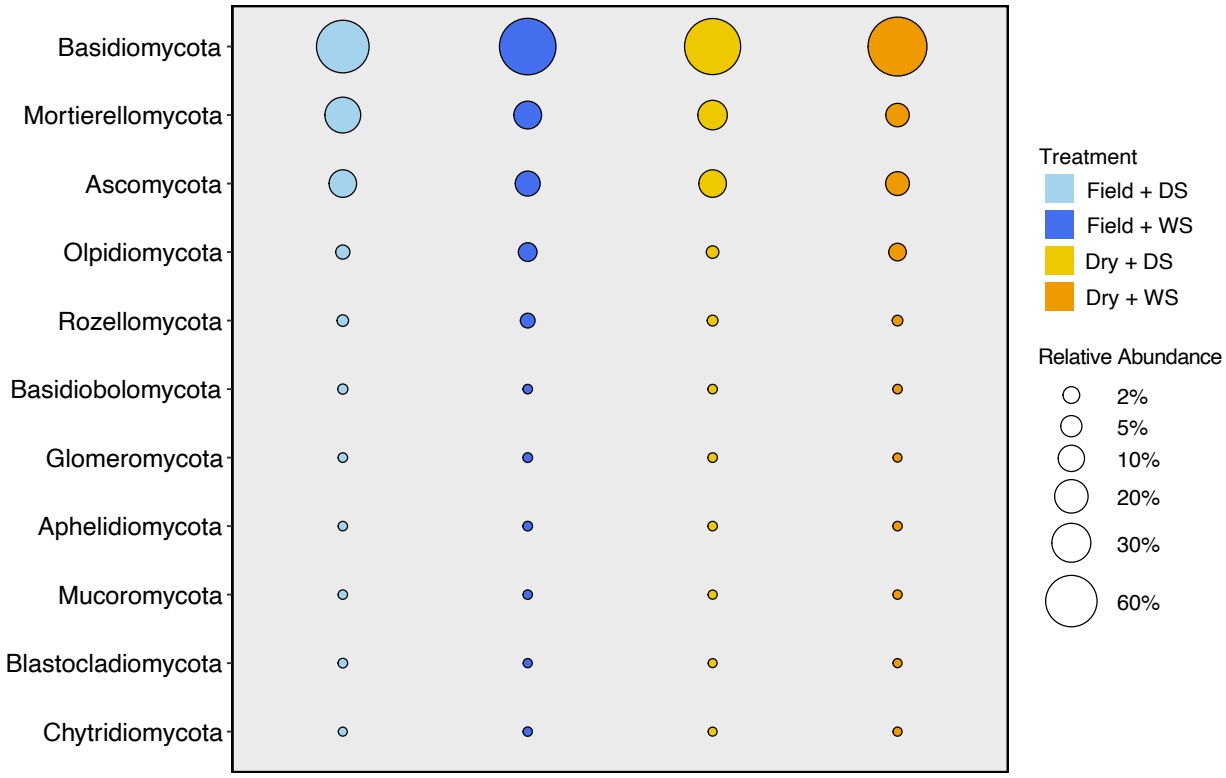


Figure S3-5. Community-weighted mean estimated traits for ribosomal RNA gene (*rrn*) copy number (a) and genome size (b) across all samples. Traits were estimated from 16S rRNA gene sequences. Different letters represent significant differences ($P < 0.05$) between sieving treatments. DS, dry sieving; WS, wet sieving.

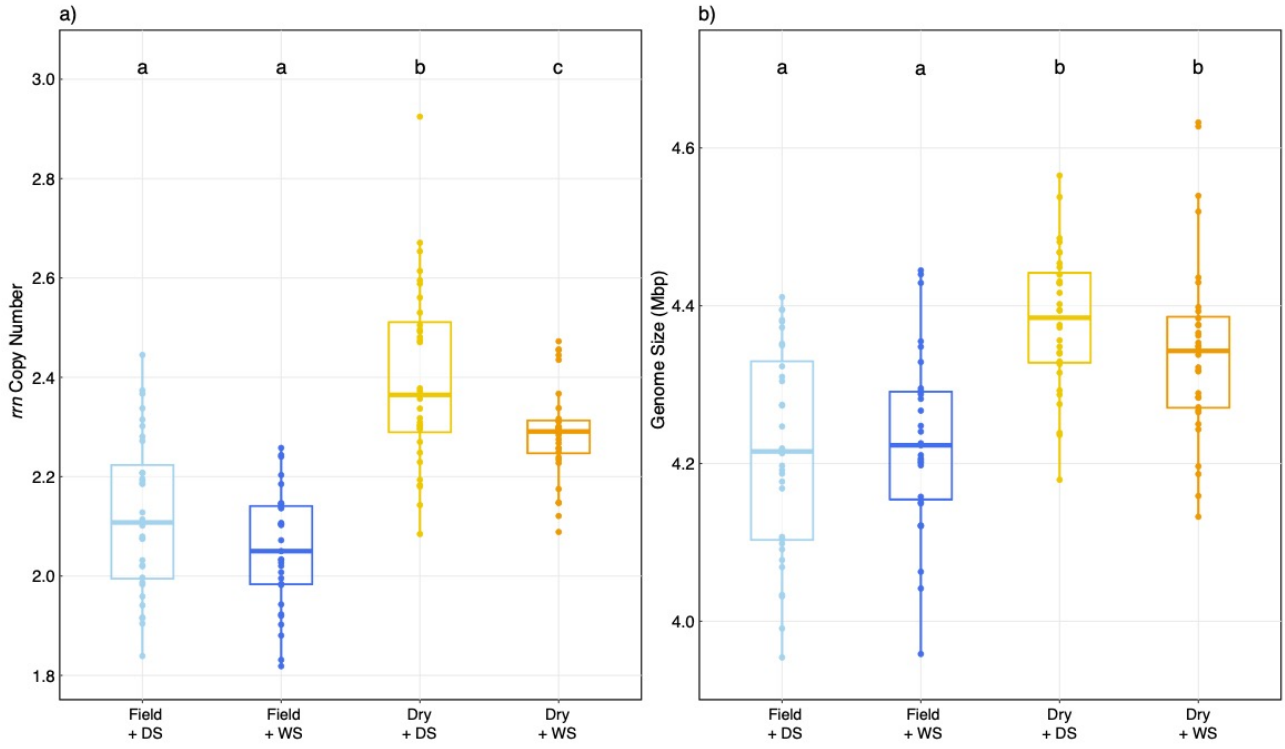


Figure S3-6. Aggregate total C (a) and N (b) concentrations by treatment across all samples. Samples are colored by fertilizer treatments. Asterisks indicate statistically significant ($P < 0.05$) differences between fertilizer treatments within each treatment group, and different letters indicate significant differences between sieving treatments. DS, dry sieving; WS, wet sieving.

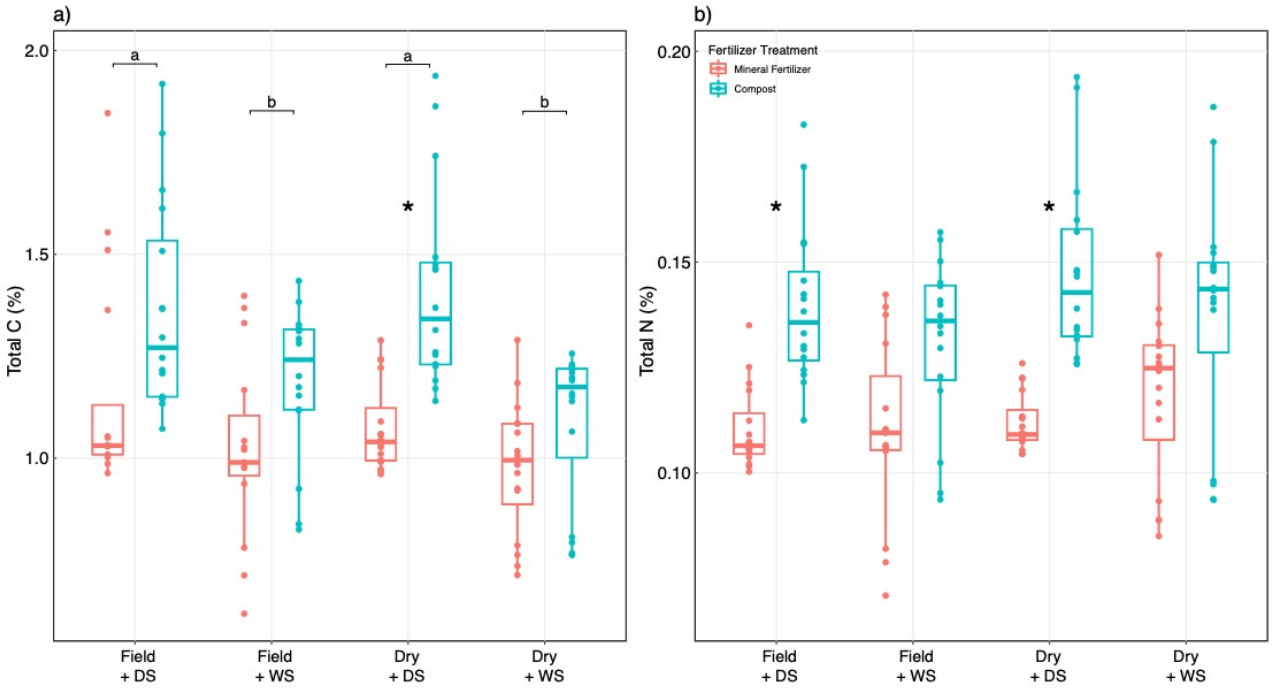
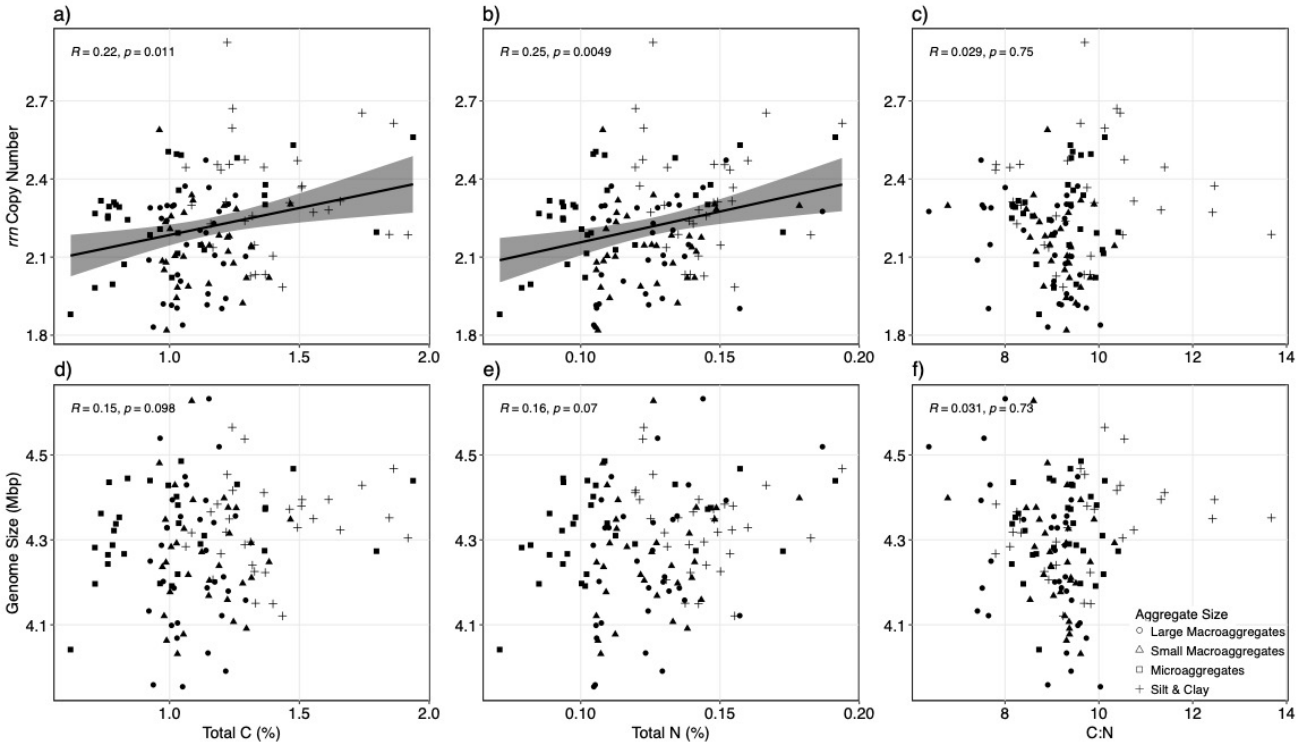


Figure S3-7. Spearman correlations of estimated traits for ribosomal RNA gene (*rrn*) copy number (a-c) and genome size (d-f) with aggregate total C, N, and C:N ratio. Trendlines were only shown for significant correlations ($P < 0.05$).



CHAPTER 3

SPATIAL STRATIFICATION OF MICROBIAL COMMUNITY TAXONOMY AND FUNCTION IN SOIL AGGREGATES OF DIFFERENT SIZE

List of Authors: Jonathan Y. Lin, Christian B. Erikson, Jane D. Fudyma, Daoyuan Wang, Erika H. Yao, Guadalupe Barajas, Jordan M. Sayre, Joanne B. Emerson, David A. Lipson, Cristina Lazcano, Kate M. Scow, Jorge L.M. Rodrigues

ABSTRACT

The soil matrix is a heterogeneous mixture composed of aggregates – three dimensional complexes composed of organic materials and mineral particles. Soil aggregates vary considerably in physical and chemical properties by size, making them unique habitats for distinct microbial communities and metabolic pathways. Yet, this spatial variability is often overlooked in studies that use homogenized soil cores. In this study, we investigated the taxonomic community, functional gene composition, and metabolic output of four size fractions of aggregates ranging from 8 mm to less than 53 μm collected from agricultural soils under two different management practices. The functional gene composition was significantly different by aggregate size, with higher abundances of genes for the degradation of plant-derived compounds in the macroaggregates and biomass recycling in the two smallest size fractions. These differences corroborated with significant differences in the composition of the metabolome but not specific enzyme activities. Both taxonomic profiling and reconstruction of genomes from metagenomes revealed a higher abundance of ammonia-oxidizing archaea in the macroaggregates, and analysis of their genomes revealed complementary metabolisms potentially enabling them to colonize different niches within the same habitat. Together, our results show that soil microbial communities and their functions are shaped by the spatial structure of the soil, which are likely driven by differences in resource availability between the macro- and microaggregates.

INTRODUCTION

Soils harbor complex assemblages of resident bacteria, archaea, and eukaryotes that together play fundamental roles in sustaining global biogeochemical cycles (Jansson and Hofmockel, 2018) and supporting plant (Jacoby et al., 2017) and human health (Samaddar et al., 2021). In the face of climate change and a growing population, the soil microbiome and its functions represent a critical biotic tool that can be leveraged to meet these challenges (Bender et al., 2016; Kallenbach et al., 2019). However, despite the use of modern high-throughput sequencing techniques and a plethora of studies that have characterized soil microbiomes under different land uses (Levine et al., 2011; Rodrigues et al., 2013; Arcand et al., 2017), rising temperatures (Dunbar et al., 2012; Hayden et al., 2012; Woodcroft et al., 2018; Wilson et al., 2021), and moisture conditions (de Vries et al., 2018; Schimel, 2018), we still lack a consistent fundamental understanding of the relationship between specific compositional states and diversity levels with microbial functions. The composition of the soil microbiome does not always predict specific microbial functions (Jansson and Hofmockel, 2018), and the contrasting interactions observed between diversity and ecosystem functions observed to date (Griffiths et al., 2001, 2004; Wertz et al., 2006; Philippot et al., 2013; Wagg et al., 2014; Maron et al., 2018) suggest that more studies are needed to decouple the factors confounding the relationship between microbial taxonomy and function in soils.

One major factor contributing to this variation is the physical structure of the soil, which is a heterogeneous matrix composed of minerals, organic matter, and pore spaces that provide a multitude of habitats for microbes (Wilpiszeski et al., 2019). The soil matrix contains unique environments at the micron to millimeter range that cannot be captured by sampling soil in bulk (Wilpiszeski et al., 2019) and therefore represent the length scales that are more ecologically relevant for soil microbes and their functional contribution to ecosystems (Vos et al., 2013; Cordero and Datta, 2016; Rillig et al., 2017; Wilpiszeski et al., 2019). All soils are primarily

composed of microaggregates – complexes defined as less than 250 μm in diameter and composed of minerals, carbonates, and other particles bound tightly together (Tisdall and Oades, 1982). In many soils, microaggregates are further enmeshed by organic material such as bacterial polysaccharides, fungal hyphae, and plant roots to form macroaggregates (> 250 μm) (Tisdall and Oades, 1982; Six et al., 2000). Aggregates have distinct physical and chemical properties from the bulk soil as well as from each other by size (Totsche et al., 2017), such as Carbon (C) content (Six et al., 2000; Davinic et al., 2012; Yu et al., 2015) and pore size affecting the diffusion of water and gasses through their networks (Sexstone et al., 1985), providing evidence that this spatial variability contains microhabitats for distinct microbial communities and metabolic pathways (Foster, 1988; Hattori, 1988; Vos et al., 2013).

Previous studies have shown that soil aggregates of different size fractions harbor distinct bacterial (Mummey et al., 2006; Davinic et al., 2012; Trivedi et al., 2017; Bach et al., 2018; Wang et al., 2022) and fungal communities (Bach et al., 2018; Upton et al., 2019; Lin et al., 2022) and have different rates of extracellular enzyme activity (Fansler et al., 2005; Bach and Hofmockel, 2014; Trivedi et al., 2015; Han et al., 2021). Yet, patterns in the abundance of specific taxa, diversity levels, and enzymatic activity among different size have insofar been difficult to generalize. Most of these studies have focused solely on the taxonomic composition or enzyme activity in aggregates, whereas the composition of functional genes and their connection with specific taxa have been less studied (Choi et al., 2018; Wu et al., 2021). Furthermore, no studies have attempted to link functional potential inferred from metagenomics with metabolic output in soil aggregates, a critical step towards understanding expressed functions of the microbes with resource availability in their immediate environment (Jansson and Hofmockel, 2018). Thus, in this study, agricultural soils from a long-term experimental plot were collected and sieved to obtain four fractions ranging in size from large macroaggregates (2000-8000 μm) to silt and clay (< 53 μm). The taxonomic and functional gene profiles of the prokaryotic community in soil

aggregates were characterized using shotgun metagenomics. To capture the metabolic output of the microbial community, soils were sampled again for analysis of metabolomes and enzyme activities for each aggregate size fraction. Our specific objectives were to 1) evaluate differences in the taxonomic and functional gene composition of the prokaryotic community between soil aggregates from different size fractions, 2) identify the major taxa associated with specific potential functions in aggregates by reconstructing metagenome-assembled genomes, and 3) compare the differences in metagenomics with differences in metabolite levels and enzyme rates in each aggregate size fraction.

MATERIALS AND METHODS

Long term field trial and soil sampling

Soils used in this study were collected from the Russell Ranch Sustainable Agricultural Research Facility, University of California, Davis (UC Davis), USA (38°32′47″N, 121°52′28″W). A field experiment was initiated in May 2012 to investigate the long-term impacts of different soil amendments on soil carbon storage, microbial communities, and crop production in a Mediterranean agricultural ecosystem. The cropping system consisted of a 2-year rotation of processing tomatoes (*Lycopersicon esculentum* Mill.) and corn (*Zea mays* L.) managed using practices and equipment similar to those used by local commercial growers. Two fertility management systems were tested: mineral fertilizer or poultry manure compost, for which the concentrations applied were scaled to have equivalent total N inputs. One application of biochar derived from walnut shells was added at the start of the experiment, resulting in four treatments: 1) mineral fertilizer without biochar; 2) mineral fertilizer with biochar, 3) compost without biochar; and 4) compost with biochar. Specific details on the management, amendments, and soil type have been described previously (Griffin et al., 2017; Wang et al., 2022). The experiment was arranged in a randomized block design with four replicate plots per treatment. In March 2018,

three soil subsamples were collected from a depth of 15 cm per plot using a soil knife, combined into a representative sample (4 for each of the field treatments), and transported to the laboratory for subsequent analysis.

Soil aggregate sieving

Freshly collected bulk soils were first passed through an 8 mm sieve by breaking clods by hand along natural planes of weakness. Aggregates were then isolated from bulk soil using the wet sieving method as described previously (Wang et al., 2022). Briefly, 50 g of the moist, 8 mm sieved soil was submerged in deionized water on top of a 2000 µm sieve for 5 min. The sieve was then moved up and down (~ 3 cm) for 2 min at a rate of 50 repetitions per min (Elliott, 1986). The soil and water passing through the sieve were then transferred by gently rinsing the material with deionized water onto the next smaller size sieve, and the same procedure was repeated. Three sieve sizes (2000 µm, 250 µm, and 53 µm) were used to generate four aggregate fractions: 1) Large macroaggregates (2000-8000 µm); 2) small macroaggregates (250-2000 µm); 3) microaggregates (53-250 µm); and 4) silt & clay (< 53 µm). For each fraction, subsamples of the aggregates that were retained on each sieve were collected in sterile 50 mL polypropylene tubes and immediately stored at -80°C until DNA extraction. A total of 16 samples were collected for each aggregate size fraction resulting in 64 samples.

Soil DNA extraction and metagenomic sequencing

DNA from each soil aggregate size fraction was extracted in duplicate using the Powerlyzer PowerSoil DNA Isolation kit (Qiagen, Germantown, MD, USA) using a vortex adaptor according to the manufacturer's instructions. The duplicate extractions of the same soil sample were pooled, and the DNA was quantified using the Qubit High-Sensitivity dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA) and inspected by gel electrophoresis on a 1% agarose gel. Libraries were prepared using an insert size range of 250-400 bp and sequenced using the

Illumina NovaSeq platform (S4 flow cell, paired-end 150 bp). Both library preparation and sequencing were performed at the UC Davis DNA Technologies core facility. A total of 1.926 Tbp of sequencing was obtained to reach an average depth of 30.09 Gbp per sample.

Microbial taxonomic profile

Trimmomatic (v.0.39) (Bolger et al., 2014) was used to remove sequencing adaptors and quality-trim the raw reads using a sliding window of 4 bp, minimum average quality of 30, and a minimum length of 50 bp. To obtain a community-wide taxonomic profile, Kraken2 (v.2.1.2) (Wood et al., 2019) was used to classify the trimmed reads against a pre-built database of bacteria, archaea, and viruses downloaded from the National Center for Biotechnology Information (NCBI) GenBank (Benson et al., 2013) in February 2022.

Functional gene profile

The quality trimmed reads were assembled into contigs using Megahit (v.1.0.6) (Li et al., 2015) with a minimum contig length of 1000 bp. Prodigal (v.2.6.3) (Hyatt et al., 2010) was used to identify open reading frames from the contigs and the protein-coding sequences were annotated using KOfamscan (v.1.3.0) (Aramaki et al., 2020), which assigns KEGG orthologs (KO) to query genes using Hidden Markov Model (HMM) profiles with predefined score thresholds from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2017). Only K number assignments with scores above the predetermined thresholds for individual KO terms from the KEGG database were counted as annotated genes and used for analysis. To quantify the functional genes, the raw reads were mapped back to the assembled contigs using Bowtie2 (v.2.4.5) (Langmead and Salzberg, 2012) and the mapping files were used to generate gene counts using HTSeq (v.1.99.2) (Anders et al., 2015).

Metagenome-assembled genomes

To infer taxonomy from the potential functions assigned at the community level, the assembled contigs were binned using MaxBin2 (v.2.2.7) (Wu et al., 2016), which recovers genomes from metagenomes based on tetranucleotide frequencies and contig coverages. MaxBin2 was run using the BAM files generated from Bowtie2 that were converted to per-contig coverage information with the "pileup.sh" script in BBMap (v.38.18) (Bushnell, n.d.). The resulting bins were checked for completeness and quality with CheckM (v.1.1.3) (Parks et al., 2015). Only bins with > 40% completion and < 10% contamination determined by CheckM were used for downstream analysis, hereafter referred to as metagenome-assembled genomes (MAGs). The taxonomy of each MAG was assigned using GTDB-Tk (v.2.0.0) (Chaumeil et al., 2020) against the Genome Taxonomy Database (GTDB) release 07-RS207. Phylogenetic trees based on multiple sequence alignments of single copy marker genes were also constructed in GTDB-Tk and visualized in Iroki (Moore et al., 2020). The MAGs were annotated using Prodigal and KOfamscan against the KEGG database as described above. CoverM (v.0.6.1) (https://github.com/wwood/CoverM) was used to determine the relative abundance of the MAGs classified as archaea after dereplicating them at an average nucleotide identity (ANI) of 99% (Varghese et al., 2015). Metabolic pathways were constructed from the K number assignments using KEGG mapper (Kanehisa and Sato, 2020).

Quantitative PCR

Quantitate PCR (qPCR) was used to quantify the abundance of ammonia-oxidizing archaea in soil aggregates. The primers crenamo A23f/crenamo A616f (Tourna et al., 2008), which target the *amoA* gene responsible for encoding the ammonia monooxygenase enzyme in archaea, were used. Amplifications were performed on a Bio-Rad CFX Connect System (Bio-Rad Laboratories, Hercules, CA, USA) in 20 µL reaction mixtures containing 10 µL SsoAdvanced Universal SYBER Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA), 0.5 µM each

primer, 5 ng template DNA, and 4 µL sterile ddH_2O. Standard curves with a detection range of 10^2-10^8 copies were generated with the pCR Blunt II-TOPO vector (Invitrogen, Carlsbad, CA, USA) containing PCR-amplified fragments for the target gene. The coefficient of determination (R^2) and amplification efficiency was 0.927 and 99.93%, respectively. Triplicate reactions were performed for each sample and a melting curve analysis was run to ensure specificity of the amplified products.

Untargeted metabolomics and enzyme assays

We performed untargeted metabolomics and enzyme assays to test whether differences in functional gene potential in aggregates translated into metabolic output differences. Subsamples of soils from the same field were collected in April 2020 and separated by wet sieving as above. Aliquots for each aggregate size fraction were either immediately used for enzyme assays described below or frozen at -80°C until submission to the UC Davis West Coast Metabolomics center for untargeted analysis of primary metabolites using a LECO Pegasus GC-TOF MS (St. Joseph, MI, USA) (Lai et al., 2018). The metabolites analyzed included carbohydrates and sugar phosphates, amino acids, hydroxyl acids, free fatty acids, purines, pyrimidines, aromatics, and exposome-derived chemicals.

Soil extracellular enzymes were assayed according to Bell et. al (Bell et al., 2013). Briefly, 2.75 g of soil from each aggregate size fraction was blended in 50 mM tris buffer (pH 7.2) to create a slurry, incubated with 4-methylumbelliferone (MUB) or 7-amino-4-methylcoumarin (MUC)-linked substrates for 3 h, and read on a microplate fluorimeter (Agilent, Santa Clara, CA, USA). The enzymes measured included β-D-cellobiosidase, β-xylosidase, α-glucosidase, β-glucosidase, N-acetyl-β-glucosaminidase, leucine aminopeptidase, and phosphatase. Enzyme activities are reported per gram aggregate size fraction. A total of 8 samples for each aggregate size fraction were assayed for enzymes and metabolites resulting in 32 samples.

Statistical analysis

All statistical analyses were performed in R (v.3.6.2). For all analyses, a value of $P < 0.05$ was considered statistically significant. No differences in the taxonomic and functional gene profile were detected from the addition of biochar to the mineral fertilizer and manure compost treatments, which was consistent with our previous report on its limited effect on soil C and bacterial community composition (Wang et al., 2022). Therefore, the replicates for biochar addition for each fertilizer treatment (mineral fertilizer vs. manure compost) were combined for downstream analyses. Bray-Curtis distance matrices were calculated for the taxonomic, functional gene, and metabolite community profiles for Non-Metric Multidimensional Scaling (NMDS) analysis to visualize overall differences in composition between aggregate size fractions. Differences were tested by permutational multivariate analysis of variance (PERMANOVA) running 999 permutations with the 'adonis' function in the 'vegan' package (Oksanen et al., 2019) using aggregate size, fertilizer treatment, and their interaction as predictor variables. The 'edgeR' package (Robinson et al., 2010) was used to estimate dispersions, transform reads to $\log_2$-counts per million (CPM), and perform statistical tests to identify differentially abundant taxa and functional genes between the large and small macroaggregates with the microaggregates and silt & clay. Differentially abundant taxa and functional genes were visualized using heatmaps produced with the 'pheatmap' package (Kolde, 2019). The 'clusterProfiler' package (Yu et al., 2012) was used to identify enriched KEGG pathways based on differentially abundant KO terms. Differential abundance analysis of the metabolomics data was performed using the 'limma' package (Ritchie et al., 2015) after normalizing samples by the sum of all peak heights for all identified metabolites. Metabolite data are presented as $\log_{10}$-transformed peak heights. All other univariate data were first tested for assumptions of normality and homogeneity of variance before comparison using a one-way Analysis of Variance (ANOVA) test and Tukey Honestly Significant

Difference post-hoc test to identify significant differences between aggregate sizes. Data that failed to meet assumptions for ANOVA were compared by using a non-parametric Kruskal-Wallis test followed by a Dunn's post-hoc test.

Data availability

The sequencing data were deposited to the NCBI sequence read archive (SRA) under BioProject PRJNA847587.

RESULTS

Soil aggregate taxonomic diversity and composition

After quality-filtering, the total number of metagenomic reads per sample ranged from 27,974,166 to 106,652,794 with an average of 84,019,069 (**Table S4-1**). To characterize the community-wide taxonomic composition of soil aggregates, we used Kraken2 (Wood et al., 2019) to profile the microbial communities. Overall, the microbial community composition at the species level was significantly different by aggregate size (**Table 4-1a**). A small but significant difference in community composition was also found across all samples by fertilizer management, but not from the interaction between aggregate size and fertilizer treatment (**Table 4-1a**). These differences are most apparent with the separation of microbial communities by aggregate size observed in the NMDS ordination (**Figure 4-1a**). Among the different aggregate size fractions, the silt & clay had a lower richness compared to the small macroaggregates (**Figure 4-2a**). However, alpha diversity was higher in the microaggregates and silt & clay than in the large and small macroaggregates based on the Shannon index (**Figure 4-2b**), and these differences corresponded with a higher Evenness of the microbial community in the microaggregates and silt & clay than in the two macroaggregate fractions (**Figure 4-2c**). Across

all samples, the taxonomic alpha diversity was higher in the manure compost than mineral fertilizer treatment (**Figure S4-1a-c**).

Differences in the abundance of specific taxonomic groups among the different aggregate sizes followed phylum- and order-specific patterns. First, the microaggregates and silt & clay harbored higher abundances of several genera within the phyla *Actinobacteria* and *Firmicutes* compared to the large and small macroaggregates (**Figure 4-3**). By contrast, 8 genera within the *Proteobacteria* and 3 genera within the Archaeal phylum *Thaumarcheota* were enriched in the large and small macroaggregates compared to the microaggregates and silt & clay (**Figure 4-3**). The genera *Pontibacter*, *Rufibacter*, and *Rhodocytophaga* from the *Bacteroidetes* phylum *Cytophagales* were found in higher abundance in the microaggregates and silt & clay, whereas *Fluviicola*, *Chryseobacterium*, and *Flavobacterium* within the *Flavobacteriales* and *Pedobacter* and *Sphingobacterium* in the *Sphingobacteriales* were enriched in the large and small macroaggregates (**Figure 4-3**). Across all samples, the manure compost treatment had a higher abundance of *Pseudomonas*, *Chryseobacterium*, *Rhodococcus*, *Sphingobacterium*, and *Pedobacter*, while the mineral fertilizer treatment had a higher abundance of *Rhodanobacter*, *Pantoea*, and *Nitrosospira* (**Table S4-2a**). Together, our results show that soil aggregates of different size fractions harbor microbial communities that are distinct in composition and diversity, with the differences in microbial groups following broad taxonomic levels.

Functional gene profile and associated taxa in soil aggregates

We assembled our metagenomic reads, identified genes from the contigs, and annotated them using the KEGG database (Kanehisa et al., 2017) to obtain a community-wide functional gene profile in aggregates. The microbial functional gene composition was different by aggregate size, fertilizer management, and the interaction between both variables (**Table 4-1b**). Aggregate

size had the largest effect in the PERMANOVA model, explaining 10.5% of the variation in functional gene composition (**Table 4-1b**) for which a strong effect of separation of samples by aggregate size was observed in the NMDS plot (**Figure 4-1b**). No differences in functional gene alpha diversity by aggregate size were detected (**Figure 4-2d-f**). Across all size fractions, the manure compost treatment had higher functional gene richness and Shannon index but not evenness compared to the mineral fertilizer treatment (**Figure S4-1d-f**).

We screened our 144 differentially abundant genes (**Table S4-3**) for metabolic functions involved with biogeochemical C and N cycling, microbial substrate assimilation, biosynthesis, and other cellular processes. The results were plotted on a heatmap, and metagenome assembled genomes (MAGs) were recovered from the assemblies and annotated to link specific genes to taxonomic groups. The microaggregates and silt & clay were enriched in genes for nitrate reduction (*narG*/*Z*, *narH*/*Y*), mannitol transport (*mtlA*), synthesis of microbial polymers (AS) and co-polymers (*tagF*, *tarS*), and degradation of fatty acids (ACOX1,3, *alkM*), glycans (*rpfB*), and sulfoquinovose (*yihS*) (**Figure 4-4a**). On the other hand, the large and small macroaggregates contained a higher abundance of genes for waste N recycling, including allantoinase (*hpxB*) and urease (*ureC*); and degradation of potential plant-derived compounds, including starch (*susC*, *susD*), xyloglucan (CEL74A), xylan (*faeB*), sialate (SIAE), furfural (*hmfF*), gentisate (*gdo*), and other polycyclic aromatic hydrocarbons (*pht3*, *pht5*) (**Figure 4-4a**). Genes encoding for bacterial toxins (*tccC*, parE1_3_4) and biofilm formation (*exoP*, *yegE*) were also enriched in the large and small macroaggregates (**Figure 4-4a**). The manure compost treatment was significantly enriched in genes for fatty acid degradation, amino acid metabolism, and transporters of various compounds compared to the mineral fertilizer treatment

(**Table S4-2b**). Based on the differentially abundant KO terms, the KEGG pathways for beta-lactam resistance, arginine biosynthesis, and two-component systems were enriched in the

large and small macroaggregates, whereas pathways for arabinogalactan biosynthesis and biosynthesis of unsaturated fatty acids were enriched the microaggregates and silt & clay (**Figure S4-2**).

Annotation of the MAGs revealed that the functional genes in aggregates varied in their representation among diverse taxa. First, the gene encoding for the ability to degrade the heterocyclic compound furfural (*hmfF*) was represented by all 9 phyla of the MAGs (**Figure 4-4b**). Genes involved in microbial breakdown pathways for the aromatic compounds gentisate (*gdo*) and phthalate (*pht5*) were also detected in 8 and 7 out of the 9 phyla, respectively (**Figure 4-4b**). The genes for components of a mannitol phosphotransferase (*mtlA*, 7 out of 9 phyla) and urease (*ureC*, 8 out of 9 phyla) were also present in many of the MAGs (**Figure 4-4b**). Second, by contrast, the genes encoding for subunits of nitrate reductase enzyme were limited to the *Acidobacteriota* (*narG*/*Z*, *narH*/*Y*), *Chloroflexota* (*narH*/*Y*), *Firmicutes* (*narH*/*Y*), and *Thermoproteota* (*narH*/*Y*), whereas the gene encoding for the insecticidal toxin complex (*tccC*) was found only in the *Acidobacteriota*, *Firmicutes*, and *Desulfobacterota* (**Figure 4-4b**). Finally, all genes that were enriched in the large and small macroaggregates were represented in a higher number of phyla compared to the genes that were enriched in the microaggregates and silt & clay (**Figure S4-3**).

Soil aggregate metabolomes and enzyme activities

To determine whether differences in functional potential observed by metagenomics translated into metabolic output differences, we performed metabolomics on each aggregate size fraction. Overall, the metabolomics profile was different by aggregate size and fertilizer treatment, but not from their interaction, with the greatest effect of separation in the metabolites observed by aggregate size (**Figure 4-5a & b**). Out of 173 detectable metabolites, 96 were identified (55.5%). The top 8 identifiable metabolites that were differentially abundant based on

log-fold change are shown in **Figure 4-5c**. The remaining 6 metabolites with statistical significance are shown in **Figure S4-4**. Notable compounds that were different by aggregate size included lactic acid, galactinol, glutamic acid, glycolic acid, and adenosine (**Figure 4-5c** & **S4-3**), which were enriched in the macroaggregates; and N-acetyl aspartate diethylester and putrescine (**Figure 4-5c & S4-3**), which were enriched the microaggregates and silt & clay. The manure compost treatment was higher in isomaltose and beta-sitosterol, whereas the mineral fertilizer treatment was higher in palmitic acid across all samples (**Table S4-2c**).

No differences in overall enzyme activities by aggregate size or fertilizer treatment were detected (**Figure S4-5a & b**). No significant differences in the activity of specific enzymes were detected between any of the aggregate size fractions except for phosphatase, which had higher activity in the silt & clay compared to the small macroaggregates (**Figure S4-5c**).

Abundance and potential function of archaea MAGs

We retrieved a total of 253 MAGs (with > 40 % completeness and < 10% contamination) from our aggregate samples. Of this total, 147 belonged to the Bacteria and 108 were from the Archaea (**Figure S4-6**). The bacterial MAGs were represented by various lineages within the phyla *Actinobacteriota* (n = 44), *Acidobacteriota* (n = 28), *Chloroflexota* (n = 19), *Firmicutes* (n =18), *Proteobacteria* (n = 16), *Desulfobacterota* (n = 8), *Gemmatimonadota* (n = 7), and *Nitrospirota* (n = 5), whereas all the archaeal MAGs were classified as belonging to the order *Nitrososphaerales* within the phylum *Thermoproteota* (formerly *Thaumarchaeota*, (Rinke et al., 2021)) (n = 108, **Figure S4-6 & Table S4-4**). Recently, it was reported that macroaggregates contained higher abundances of ammonia-oxidizing archaea (AOA) (Han et al., 2020) and nitrification rates (Li et al., 2020). Thus, we further analyzed our archaea MAGs to characterize their phylogeny, evaluate their ecological distribution, and determine their potential functions in aggregates. Of the 108 archaeal MAGs, 46 were nearly complete (> 90%) and of high quality

(**Table S4-4**) (Bowers et al., 2017), which were used for analysis. Phylogenetic analysis revealed that the archaea MAGs represented two distinct genera within the family *Nitrososphaera* (**Figure S4-7**). The first group, represented by 8 MAGs, was classified as a species within the genus *Nitrososphaera* while the other 38 MAGs formed a cluster within the as-yet uncultivated genus TH1177 (**Figure S4-7**).

Using CoverM (https://github.com/wwood/CoverM) to dereplicate the MAGs and estimate their abundance, we found that both lineages were higher in relative abundance in the macroaggregates. A46_maxbin.003, the representative MAG for *Nitrososphaeraceae* TH1177, was more abundant in the large macroaggregates than in the microaggregates and silt & clay (**Figure S4-8a**), whereas A49_maxbin.001 representing the *Nitrososphaera* species was more abundant in both the large and small macroaggregates compared to the two smallest aggregate size fractions (**Figure S4-8b**). These findings were confirmed by qPCR targeting the *amoA* gene in archaea, where AOA copy numbers were higher in the large and small macroaggregates compared to the silt & clay (**Figure S4-8c**), and significantly correlated with the relative abundances of both archaea MAGs (**Figure S4-8d** & **e**). Pathway reconstruction revealed differences in the genomic composition between the two MAGs (**Figure 4-6**). A46_maxbin.003 possessed a gene for one subunit of ammonia monooxygenase (*amoB*), a gene for an ammonium transporter (*amt*), a complete gene cluster for urease (*ureBCDEFG*) and urea transporter (*utp*), zinc transporter (*znuABC*), both high and low affinity phosphate transporters (*pstSACB* & *PiT*) and genes for CO_2 fixation (**Figure 4-6a**). On the other hand, A49_maxbin.001 contained a complete gene cluster for ammonia monooxygenase (*amoCAB*), nitrite reductase (*nirK*), an incomplete zinc transporter (*znuAC*), and only the high affinity phosphate transporter (*pstSACB*) (**Figure 4-6b**). Furthermore, compared to A46_maxbin.003, A49_maxbin.001 possessed genes for several ABC-type transporters of multiple saccharides and polyols (*malK, msmX, msmK,*

smoK, aglK, msiK) as well as a higher number of genes involved in the CO_2-fixing hydroxypropionate/hydroxybutyrate (HP/HB) cycle (**Figure 4-6b**).

DISCUSSION

The different physicochemical properties within macro- and microaggregates make them unique habitats for distinct communities of microorganisms within the soil environment (Hattori, 1988; Rillig et al., 2017). This was confirmed in our study, where we found significant differences in the microbial taxonomic and functional gene composition and metabolites by aggregate size. Our results depicted the different aggregate size fractions as resource-rich or resource-poor microhabitats. While microaggregates and silt & clay appeared to be metabolically limiting for microbes, reflecting a low availability of C, enrichment of functional genes in both large and small macroaggregates reflected a higher concentration and diversity of nutrients for more microbial competition or niche differentiation to occur.

Macroaggregates have more Thaumarchaeota while microaggregates harbor more Actinobacteria

We found that the microbial community composition and diversity in aggregates was different by aggregate size. Notably, microaggregates and silt & clay had higher alpha diversity (Shannon) compared to large and small macroaggregates. While microaggregates and silt & clay harbored higher abundances of members within the phyla *Actinobacteria* and *Firmicutes*, large and small macroaggregates had higher abundances of *Thaumarchaeota* and *Proteobacteria*. These findings corroborate our previous study, using 16S rRNA gene sequencing to characterize the same samples, where we found a higher proportion of *Actinobacteria* orders *Micrococcales*, *Streptomycetales*, and *Propionibacteriales* in microaggregates and silt & clay and a higher proportion of the *Thaumarchaeota* order *Nitrososphaerales* in large and small macroaggregates (Wang et al., 2022).

While molecular characterizations of microbial communities in aggregates based on amplicon sequencing or clone libraries have been conducted in a number of studies (Wilpiszeski et al., 2019), only a few have found microbial groups inhabiting the different size fractions in patterns that were similar to this study. For instance, Mummey et. al (Mummey and Stahl, 2004; Mummey et al., 2006) found that the *Actinobacteria* were abundant in microaggregates including their interiors. This was consistent with a recent report by Bach et. al (Bach et al., 2018), which found higher relative abundances of several *Actinobacteria* orders as well as higher alpha diversity in the microaggregates (Bach et al., 2018; Upton et al., 2019). Similarly, higher abundances of α-*Proteobacteria* (Trivedi et al., 2015) and ammonia-oxidizing archaea (Han et al., 2020) were found in macroaggregates using qPCR. However, contrasting patterns were reported by Davinic et. al (Davinic et al., 2012), where most *Actinobacteria* groups (except *Rubrobacteriales*) were detected at higher abundances in the macroaggregates; and Trivedi et. al (Trivedi et al., 2017), where the macroaggregates harbored more *Actinobacteria* and higher alpha diversity compared to the microaggregates. This discordance between aggregate size with microbial groups and diversity levels suggests that while aggregates of different size provide ecological niches for distinct microbial communities, consistent patterns of specific taxa that preferentially colonize macro- versus microaggregates are not yet evident (Totsche et al., 2017). We surmise that these differences may be due to two factors. First, different methods are frequently used across studies to isolate aggregates from the bulk soil (Blaud et al., 2017). Soil aggregates can be collected using dry sieving, which involves shaking bulk soil on top of a stack of sieves; or wet sieving, which submerges the bulk soil in water followed by repeated vertical strokes on top of a sieve while still immersed (Elliott, 1986). As these sieving methods can alter microbial extracellular enzyme activity (Bach and Hofmockel, 2014) and bacterial and fungal community composition in aggregates (Lin et al., 2022), the variation in preferences of specific taxa in aggregates reported across the literature may reflect differences in aggregate

fractionation methods that were used (B. Wang et al., 2018). Second, differences in the distribution of specific taxa, compositional states, or species diversity among aggregate size fractions may be due to local adaptation of resident microbes to regional variations in soil texture, management, or the order of species establishment by different microbes (Gómez et al., 2016).

Differences in functional genes and associated taxa suggest that resource availability and functional redundancy vary by aggregate size

Macroaggregates are composed of microaggregates bound together by various organic compounds, including plant roots, fungal hyphae, and bacterial exopolysaccharides (Gupta and Germida, 2015; Costa et al., 2018). As macroaggregates form, this pool of organic matter can become occluded within their interiors (Six et al., 2000), which is protected from degradation by excluded microbes (i.e. on aggregate surfaces) but can serve as a nutrient source for the microbial communities within (Rillig et al., 2017). Our analysis revealed that the functional gene composition of the microbial community within aggregates was significantly different by aggregate size. Many of the genes that were enriched in the large and small macroaggregates were associated with the degradation of plant-derived compounds. For instance, we found a significant enrichment of genes for the breakdown of xylan (*faeB*), xyloglucan (CEL74A), furfural (*hmfF*), and gentisate (*gdo*). Feruloyl esterase, encoded by *faeB*, cleaves ester linkages in xylan to release aromatic acids (Benoit et al., 2008), whereas xyloglucan exo-beta-1, 4-glucanase (CEL74A) hydrolyses xyloglucan to release oligosaccharides (Grishutin et al., 2004). Furfural is formed naturally by the dehydration of xylose, a sugar that is abundant in the hemicellulose fraction of lignocellulosic biomass (Mathew et al., 2018), while gentisate (2, 5-dihydroxybenzoate) is a phenolic acid that is widely distributed as a secondary plant product (Abedi et al., 2020). In an earlier study, we found in the same samples that the small macroaggregates had higher total C compared to the microaggregates and silt & clay (Wang et

al., 2022), indicating that the higher abundance of these genes in macroaggregates in our study likely reflects a higher nutrient availability for microbes in these size fractions. This was corroborated by our metabolite data, that showed a higher abundance of compounds involved in microbial energy metabolism in macroaggregates than other size fractions. Other studies have shown that macroaggregates contain higher total soil organic carbon (SOC) than microaggregates and silt & clay (Davinic et al., 2012), and that macroaggregates preferentially accumulate phenolic, carboxyl, and methoxyl/N-alkyl C compounds from agricultural inputs during their formation from microaggregates (Yu et al., 2015). This suggests that compared to the microaggregates, incorporation of plant residues into macroaggregates make them resource-rich habitats for microbes (Choi et al., 2018).

Higher concentrations of microbially-available nutrients in the larger aggregates may coincide with elevated competition or facilitation between microbes in macroaggregates. We detected not only genes for breakdown of plant-derived compounds, but also for bacterial toxin formation, allantoinase, urease, and biofilm formation. Toxin production is a strategy used by some lineages of bacteria to defend against predation by insects or soil microfauna (Waterfield et al., 2001). A higher abundance of the *ureC* gene in macroaggregates has been reported previously (L. Wang et al., 2018), and together with the gene for allantoinase (*hpxB*), is an indicator that some microbial groups can benefit by recycling waste N or other co-metabolites produced by other microbes. Bacteria can form biofilms to stabilize conditions in their immediate environment and increase the efficiency of resource acquisition (Cai et al., 2019). Bacteria and fungi preferentially occupy pores measuring between 1 μm and 1 mm in size (Chenu and Cosentino, 2011), and biofilms can form in the large pores prevalent within macroaggregates to take advantage of moisture, oxygen, and nutrients diffusing from the soil matrix (Chenu et al., 2001). This finding contrasts with our detection of genes for dissimilatory nitrate reduction (*narG*/*Z*, *narH*/*Y*) that were enriched in microaggregates and silt & clay, both of which contain

smaller pores that can limit gas diffusion into their interiors, particularly when saturated with water (Totsche et al., 2017). This indicated that more prokaryotes in the smallest aggregate size fractions have the capability to use alternate terminal electron acceptors under anoxic conditions.

Other genes that increased in microaggregates and silt & clay were ones related to the synthesis and degradation of fatty acids and other polymers of microbial origin. This implies that compared to the macroaggregates, the microorganisms residing in the microaggregate and silt & clay fractions are more limited in nutrient availability and access. Microaggregates often contain little to no plant debris (Oades and Waters, 1991) and any organic matter present is typically associated with phyllosilicates, metal oxides, and other minerals that restrict microbes from access (Totsche et al., 2017). Therefore, degradation of microbial biomass, including microbial solutes, extracellular polymeric substances (EPS), and cell wall remnants may be the only sources of nutrients available for the resident microbes in these size fractions. The genes that we identified lend support to this notion. For example, we found significant enrichment of a gene for a mannitol phosphotransferase system (*mtlA*) in microaggregates and silt & clay. Mannitol is a widely distributed polyol that is regularly produced by fungi as a major storage compound and by bacteria as a solute in response to osmolarity stress (Song and Vieille, 2009). The increased abundance of *mtlA*, together with the other genes for the degradation of microbial glycans and aliphatic hydrocarbons, suggests that the microbes within the microaggregates and silt & clay can assimilate microbial products for survival. Recent studies have shown that most of the stable organic matter formed in soil is driven by microbial biomass (Kallenbach et al., 2016; Kästner et al., 2021), and that mineral-associated organic matter (MAOM) in microaggregates is predominantly microbial in origin (Plaza et al., 2013), providing evidence that microbial biomass in the microaggregates and silt & clay can be used by other microbes as a nutrient source.

However, whether resident microbes “entombed” within microaggregates (Liang et al., 2017) are active in recycling biomass has not yet been determined.

Genes enriched in large and small macroaggregates were represented in a wider range of phyla from our MAGs compared to genes enriched in microaggregates and silt & clay. This pattern was significant when quantified (**Fig. S2**), which suggests that genes that are abundant in macroaggregates may be functionally redundant amongst diverse soil phyla. Interestingly, we observed that the genes involved in the degradation pathways of heterocyclic and aromatic compounds were distributed among many of the phyla (i.e. *hmfF* & *gdo*), indicating that these potential functions may be a common feature of many soil prokaryotes. The extent to which the capacity to degrade organic matter is phylogenetically conserved in soil microbes is not yet known, but emerging evidence points to a degree of functional redundancy (Martiny et al., 2013). Recently, Barnett et. al (Barnett et al., 2021) found that the C assimilation dynamics for substrates of varying chemical complexity were poorly phylogenetically conserved amongst diverse soil bacteria. While the taxa that assimilated C from vanillin, cellulose, and palmitic acid showed the strongest phylogenetic clustering, this effect was not significant and was highly variable with time (Barnett et al., 2021). They proposed that in addition to possessing specific catabolic pathways, ecological traits governing growth rate, motility, and access are important for determining microbial soil C assimilation (Barnett et al., 2021). Thus, our finding of C degradation genes that were broadly distributed among diverse taxa in the macroaggregates supports the paradigm that the stability of soil organic matter is limited by microbial access rather than chemical recalcitrance (Schmidt et al., 2011; Dungait et al., 2012).

Soil aggregates differed by size in metabolites but not enzyme activity

In addition to differences in functional gene composition, we confirmed by GC-TOF MS that the metabolic output of aggregates also significantly differed by aggregate size. Notably, we

found that glutamic acid, lactic acid, galactinol, glycolic acid, and adenosine were generally more enriched in the macroaggregates than in the two smallest aggregate size fractions. These metabolites reflect a higher abundance of resources in the macroaggregates to support microbial growth and energy metabolism.

Interestingly, we saw no differences in enzyme activity by aggregate size for most of the enzymes that we assayed, indicating that the differences in functional potential in the different aggregate sizes did not correspond with measured activity. Overall enzyme rates were low, which may indicate that the substrates used by the enzymes (i.e., cellulose) were not present or at very low concentrations at the time of sampling. The field was bare and had been fallow for more than 5 months when samples were collected (April 2020), which was one month later than our first sampling for the metagenomes (March 2018). This time point (April 2020) may have been a period when much of the available plant residues were decomposed and when microbial enzyme activity in aggregates is low prior to the start of the season (Bach and Hofmockel, 2016). Supporting this, we found that total organic matter in the bulk soils was significantly higher in manure compost than in the mineral fertilizer plots in March 2020, but no differences between the two treatments were detected in April 2020 (unpublished data). Furthermore, we found that the substrates and products of the enzymes assayed were not significantly different by aggregate size or fertilizer treatment (xylose, sucrose, glucose, phosphate, leucine) in our metabolomics dataset or not identified.

Ammonia oxidizing archaea vary in genomic content and occupy different niches in macroaggregates

Our analysis of the archaea MAGs revealed that they represented two distinct genera within the *Nitrososphaeraceae*, a family formerly belonging to the *Thaumarcheota* that was recently reclassified as a group within the new phylum *Thermoproteota* (Rinke et al., 2021). The

Nitrososphaeraceae are a group of ammonia-oxidizing archaea (AOA) common in agricultural soils managed using synthetic ammonium fertilizer and where they play an important role in nitrification (Zhalnina et al., 2013; Huang et al., 2021). Our quantification of these MAGs revealed that they were more abundant in the large and small macroaggregates than in the microaggregates and silt & clay. This finding was consistent with our Kraken2-based results showing higher abundances of *Thaumarcheota* in the macroaggregates and was confirmed by qPCR, demonstrating that soil AOA may have a distinct preference for occupying niches within macroaggregates. The canonical genes for CO_2 fixation, ammonia monooxygenase, and nitrite reductase were present in the MAGs, suggesting that the large pores in macroaggregates may allow for diffusion of these nutrients from the soil matrix to create an optimal habitat. Our results corroborate a few previous studies that detected higher abundances of AOA in macroaggregates compared to microaggregates using qPCR (Li et al., 2019; Han et al., 2020) but contrast with two studies that found high variation by location, where some soils had higher abundances in the microaggregates (Nahidan et al., 2017) or no differences by aggregate size (Chen et al., 2016). Notably, a recent study by Wu et. al (Wu et al., 2021) used a similar coverage-based method to estimate the abundance of AOA MAGs in aggregates from an agricultural soil. While two of their *Nitrososphaeraceae* MAGs were more abundant in macroaggregates, another was more abundant in the microaggregates, indicating potential ecotypic variation (Wu et al., 2021).

Interestingly, two AOA MAGs appeared to vary in functional potential. Despite being nearly complete (99.03%), the MAG of A46_maxbin.003 (genus TH1177) did not contain a complete operon for ammonia monooxygenase compared to A49_maxbin.001 (genus *Nitrososphaera*), which has a complete *amoCAB* gene cluster in addition to *nirK*. Instead, A46_maxbin.003 possesses a gene cluster for urease and a urea transporter, which A49_maxbin.001 lacks. A49_maxbin.001 also contains several genes encoding for ABC-type transporters of various polysaccharides, oligosaccharides, and polyols which are absent in the

genome of A46_maxbin.003. These differences between the two MAGs suggest a potential for the two lineages to use different nutrient sources to support complementary metabolisms. In the absence of ammonia, some AOA (including *Nitrososphaera*) can grow on urea as their sole energy source (Tourna et al., 2011; Qin et al., 2014), whereas others can couple organic C assimilation with CO_2 fixation for an energetic advantage (Hommes et al., 2003; Qin et al., 2014; Stieglmeier et al., 2014), showing in principle that these two lineages can occupy different niches within the same macroaggregate habitat (Prosser and Nicol, 2012).

CONCLUSION

Our study enabled us to obtain novel insights on soil aggregates as diverse microbial habitats within the soil matrix. We found that the microbial taxonomic and functional gene composition was significantly different by aggregate size, and our analysis of the functional genes and metabolites characterized macroaggregates as environments rich in resources originating from plant residues. On the other hand, the microaggregates and silt & clay represented resource-limited habitats with metabolic pathways largely restricted to the recycling of microbial biomass and anaerobic respiration. We found that many functional genes in the macroaggregates were redundant amongst diverse phyla but demonstrated that some groups, such as the AOA may vary in genomic content and potentially colonize different niches within the same macroaggregate habitats. Taken together, our results provide evidence that well-aggregated soils offer a larger diversity of physical habitats and diversified C sources for microorganisms to survive and carry out a broad range of metabolic activities (Chenu and Cosentino, 2011). Therefore, soils with more macroaggregates are in principle expected to harbor a higher diversity of microbes and microbially-driven ecosystem benefits. However, aggregates are rarely considered in surveys of soil microbial diversity (Bach et al., 2018), indicating that a substantial proportion of microbial taxa and functional genes are overlooked

when studying bulk soils alone. Our study demonstrates that a closer examination of microbes and their spatial stratification in aggregates would strengthen linkages between microbial community diversity and function. This integrated knowledge is needed to develop novel, biologically based solutions to improve soil management and increase the sustainability of our agricultural systems.

ACKNOWLEDGEMENTS

We thank the staff of the Russell Ranch Sustainable Agriculture Facility at UC Davis for their support in the field study. This research was supported in part by the USDA National Institute of Food and Agriculture (NIFA) through Hatch Formula Funding CA-D-LAW-2616-H, a UC Davis Henry A. Jastro Award (to JYL), and a USDA NIFA AFRI Predoctoral Fellowship (2020-67034-31937 to JYL).

REFERENCES

Abedi, F., Razavi, B.M., Hosseinzadeh, H., 2020. A review on gentisic acid as a plant derived phenolic acid and metabolite of aspirin: Comprehensive pharmacology, toxicology, and some pharmaceutical aspects. Phytotherapy Research 34, 729–741.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169.

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., Ogata, H., 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36, 2251–2252.

Arcand, M.M., Levy-Booth, D.J., Helgason, B.L., 2017. Resource legacies of organic and conventional management differentiate soil microbial carbon use. Frontiers in Microbiology 8.

Bach, E.M., Hofmockel, K.S., 2016. A time for every season: Soil aggregate turnover stimulates decomposition and reduces carbon loss in grasslands managed for bioenergy. GCB Bioenergy 8, 588–599.

Bach, E.M., Hofmockel, K.S., 2014. Soil aggregate isolation method affects measures of intra-aggregate extracellular enzyme activity. Soil Biology and Biochemistry 69, 54–62.

Bach, E.M., Williams, R.J., Hargreaves, S.K., Yang, F., Hofmockel, K.S., 2018. Greatest soil microbial diversity found in micro-habitats. Soil Biology and Biochemistry 118, 217–226.

Barnett, S.E., Youngblut, N.D., Koechli, C.N., Buckley, D.H., 2021. Multisubstrate DNA stable isotope probing reveals guild structure of bacteria that mediate soil carbon cycling. Proceedings of the National Academy of Sciences 118.

Bell, C.W., Fricks, B.E., Rocca, J.D., Steinweg, J.M., McMahon, S.K., Wallenstein, M.D., 2013. High-throughput fluorometric measurement of potential soil extracellular enzyme activities. JoVE (Journal of Visualized Experiments) e50961.

Bender, S.F., Wagg, C., van der Heijden, M.G.A., 2016. An underground revolution: Biodiversity and soil ecological engineering for agricultural sustainability. Trends in Ecology & Evolution 31, 440–452.

Benoit, I., Danchin, E.G.J., Bleichrodt, R.-J., de Vries, R.P., 2008. Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. Biotechnology Letters 30, 387–396.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. Nucleic Acids Research 41, D36–D42.

Blaud, A., Menon, M., van der Zaan, B., Lair, G.J., Banwart, S.A., 2017. Effects of dry and wet sieving of soil on identification and interpretation of microbial community composition, in: Sparks, S.A.B. and D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 119–142.

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

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P.,

Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Murat Eren, A., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology 35, 725–731.

Bushnell, B., n.d. BBMap [WWW Document]. SourceForge. URL https://sourceforge.net/projects/bbmap/

Cai, P., Sun, X., Wu, Y., Gao, C., Mortimer, M., Holden, P.A., Redmile-Gordon, M., Huang, Q., 2019. Soil biofilms: Microbial interactions, challenges, and advanced techniques for *ex-situ* characterization. Soil Ecology Letters 1, 85–93.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2020. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927.

Chen, Z., Guo, Y., Du, Z., Wu, W., Meng, F., 2016. Change in the abundance and community composition of ammonia-oxidizing bacteria and archaea at soil aggregate level as native pasture converted to cropland in a semiarid alpine steppe of central Asia. Journal of Soils and Sediments 16, 243–254.

Chenu, C., Cosentino, D., 2011. Microbial regulation of soil structural dynamics., in: Ritz, K., Young, I. (Eds.), The Architecture and Biology of Soils: Life in Inner Space. CABI, Wallingford, pp. 37–70.

Chenu, C., Hassink, J., Bloem, J., 2001. Short-term changes in the spatial distribution of microorganisms in soil aggregates as affected by glucose addition. Biology and Fertility of Soils 34, 349–356.

Choi, J., Bach, E., Lee, J., Flater, J., Dooley, S., Howe, A., Hofmockel, K.S., 2018. Spatial structuring of cellulase gene abundance and activity in soil. Frontiers in Environmental Science 6.

Cordero, O.X., Datta, M.S., 2016. Microbial interactions and community assembly at microscales. Current Opinion in Microbiology, Environmental microbiology * Special Section: Megaviromes 31, 227–234.

Costa, O.Y.A., Raaijmakers, J.M., Kuramae, E.E., 2018. Microbial extracellular polymeric substances: Ecological function and impact on soil aggregation. Frontiers in Microbiology 9.

Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak, J.C., Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy reveal distinct aggregate stratification of soil bacterial communities and organic matter composition. Soil Biology and Biochemistry 46, 63–72.

de Vries, F.T., Griffiths, R.I., Bailey, M., Craig, H., Girlanda, M., Gweon, H.S., Hallin, S., Kaisermann, A., Keith, A.M., Kretzschmar, M., Lemanceau, P., Lumini, E., Mason, K.E., Oliver, A., Ostle, N., Prosser, J.I., Thion, C., Thomson, B., Bardgett, R.D., 2018. Soil bacterial networks are less stable under drought than fungal networks. Nature Communications 9, 3033.

Dunbar, J., Eichorst, S.A., Gallegos-Graves, L.V., Silva, S., Xie, G., Hengartner, N.W., Evans, R.D., Hungate, B.A., Jackson, R.B., Megonigal, J.P., Schadt, C.W., Vilgalys, R., Zak, D.R., Kuske, C.R., 2012. Common bacterial responses in six ecosystems exposed to 10 years of elevated atmospheric carbon dioxide. Environmental Microbiology 14, 1145–1158.

Dungait, J.A.J., Hopkins, D.W., Gregory, A.S., Whitmore, A.P., 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. Global Change Biology 18, 1781–1796.

Elliott, E.T., 1986. Aggregate structure and carbon, nitrogen, and phosphorus in native and cultivated soils. Soil Science Society of America Journal 50, 627–633.

Fansler, S.J., Smith, J.L., Bolton, H., Bailey, V.L., 2005. Distribution of two C cycle enzymes in soil aggregates of a prairie chronosequence. Biology and Fertility of Soils 42, 17–23.

Foster, R.C., 1988. Microenvironments of soil microorganisms. Biology and Fertility of Soils 6, 189–203.

Gómez, P., Paterson, S., De Meester, L., Liu, X., Lenzi, L., Sharma, M.D., McElroy, K., Buckling, A., 2016. Local adaptation of a bacterium is as important as its presence in structuring a natural microbial community. Nature Communications 7, 12453.

Griffin, D.E., Wang, D., Parikh, S.J., Scow, K.M., 2017. Short-lived effects of walnut shell biochar on soils and crop yields in a long-term field experiment. Agriculture, Ecosystems & Environment 236, 21–29.

Griffiths, B.S., Kuan, H.L., Ritz, K., Glover, L.A., McCaig, A.E., Fenwick, C., 2004. The relationship between microbial community structure and functional stability, tested experimentally in an upland pasture soil. Microbial Ecology 47, 104–113.

Griffiths, B.S., Ritz, K., Wheatley, R., Kuan, H.L., Boag, B., Christensen, S., Ekelund, F., Sørensen, S.J., Muller, S., Bloem, J., 2001. An examination of the biodiversity–ecosystem function relationship in arable soil microbial communities. Soil Biology and Biochemistry 33, 1713–1722.

Grishutin, S.G., Gusakov, A.V., Markov, A.V., Ustinov, B.B., Semenova, M.V., Sinitsyn, A.P., 2004. Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. Biochimica et Biophysica Acta - General Subjects 1674, 268–281.

Gupta, V.V.S.R., Germida, J.J., 2015. Soil aggregation: Influence on microbial biomass and implications for biological processes. Soil Biology and Biochemistry 80, A3–A9.

Han, S., Delgado-Baquerizo, M., Luo, X., Liu, Y., Van Nostrand, J.D., Chen, W., Zhou, J., Huang, Q., 2021. Soil aggregate size-dependent relationships between microbial functional diversity and multifunctionality. Soil Biology and Biochemistry 154, 108143.

Han, S., Luo, X., Tan, S., Wang, J., Chen, W., Huang, Q., 2020. Soil aggregates impact nitrifying microorganisms in a vertisol under diverse fertilization regimes. European Journal of Soil Science 71, 536–547.

Hattori, T., 1988. Soil aggregates as microhabitats of microorganisms. Reports of the Institute for Agricultural Research - Tohoku University (Japan).

Hayden, H.L., Mele, P.M., Bougoure, D.S., Allan, C.Y., Norng, S., Piceno, Y.M., Brodie, E.L., DeSantis, T.Z., Andersen, G.L., Williams, A.L., Hovenden, M.J., 2012. Changes in the microbial community structure of bacteria, archaea and fungi in response to elevated CO_2

and warming in an Australian native grassland soil. Environmental Microbiology 14, 3081–3096.

Hommes, N.G., Sayavedra-Soto, L.A., Arp, D.J., 2003. Chemolithoorganotrophic growth of Nitrosomonas europaea on fructose. Journal of Bacteriology 185, 6809–6814.

Huang, L., Chakrabarti, S., Cooper, J., Perez, A., John, S.M., Daroub, S.H., Martens-Habbena, W., 2021. Ammonia-oxidizing archaea are integral to nitrogen cycling in a highly fertile agricultural soil. ISME Communications 1, 1–12.

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

Jacoby, R., Peukert, M., Succurro, A., Koprivova, A., Kopriva, S., 2017. The role of soil microorganisms in plant mineral nutrition—current knowledge and future directions. Frontiers in Plant Science 8.

Jansson, J.K., Hofmockel, K.S., 2018. The soil microbiome — from metagenomics to metaphenomics. Current Opinion in Microbiology 43, 162–168.

Kallenbach, C.M., Frey, S.D., Grandy, A.S., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. Nature Communications 7, 13630.

Kallenbach, C.M., Wallenstein, M.D., Schipanksi, M.E., Grandy, A.S., 2019. Managing agroecosystems for soil microbial carbon use efficiency: Ecological unknowns, potential outcomes, and a path forward. Frontiers in Microbiology 10.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K., 2017. KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research 45, D353–D361.

Kanehisa, M., Sato, Y., 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Science 29, 28–35.

Kästner, M., Miltner, A., Thiele-Bruhn, S., Liang, C., 2021. Microbial necromass in soils—linking microbes to soil processes and carbon turnover. Frontiers in Environmental Science 9.

Kolde, R., 2019. pheatmap: Pretty heatmaps.

Lai, Z., Tsugawa, H., Wohlgemuth, G., Mehta, S., Mueller, M., Zheng, Y., Ogiwara, A., Meissen, J., Showalter, M., Takeuchi, K., Kind, T., Beal, P., Arita, M., Fiehn, O., 2018. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. Nature Methods 15, 53–56.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.

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

Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W., 2015. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

Li, J., Yang, C., Liu, X., Ji, H., Shao, X., 2020. Soil aggregate size influences the impact of inorganic nitrogen deposition on soil nitrification in an alpine meadow of the Qinghai–Tibet Plateau. PeerJ 8, e8230.

Li, P.-P., Han, Y.-L., He, J.-Z., Zhang, S.-Q., Zhang, L.-M., 2019. Soil aggregate size and long-term fertilization effects on the function and community of ammonia oxidizers. Geoderma 338, 107–117.

Liang, C., Schimel, J.P., Jastrow, J.D., 2017. The importance of anabolism in microbial control over soil carbon storage. Nature Microbiology 2, 17105.

Lin, J.Y., Wang, D., Law, K.C., Scow, K.M., Rodrigues, J.L.M., 2022. Differential responses of prokaryotic and fungal communities in soil microenvironments to drying and wetting as affected by soil aggregate isolation method. In Preparation.

Maron, P.-A., Sarr, A., Kaisermann, A., Lévêque, J., Mathieu, O., Guigue, J., Karimi, B., Bernard, L., Dequiedt, S., Terrat, S., Chabbi, A., Ranjard, L., 2018. High microbial diversity promotes soil ecosystem functioning. Applied and Environmental Microbiology AEM.02738-17.

Martiny, A.C., Treseder, K., Pusch, G., 2013. Phylogenetic conservatism of functional traits in microorganisms. The ISME Journal 7, 830–838.

Mathew, A.K., Abraham, A., Mallapureddy, K.K., Sukumaran, R.K., 2018. Lignocellulosic biorefinery wastes, or resources? in: Bhaskar, T., Pandey, A., Mohan, S.V., Lee, D.-J., Khanal, S.K. (Eds.), Waste Biorefinery. Elsevier, pp. 267–297.

Moore, R.M., Harrison, A.O., McAllister, S.M., Polson, S.W., Wommack, K.E., 2020. Iroki: Automatic customization and visualization of phylogenetic trees. PeerJ 8, e8584. doi:10.7717/peerj.8584

Mummey, D., Holben, W., Six, J., Stahl, P., 2006. Spatial stratification of soil bacterial populations in aggregates of diverse soils. Microbial Ecology 51, 404–411.

Mummey, D.L., Stahl, P.D., 2004. Analysis of soil whole- and inner-microaggregate bacterial communities. Microbial Ecology 48, 41–50.

Nahidan, S., Nourbakhsh, F., Henneberger, R., Lazzaro, A., Zeyer, J., 2017. Aggregate size distribution of ammonia-oxidizing bacteria and archaea at different landscape positions. Geomicrobiology Journal 34, 895–902.

Oades, J.M., Waters, A.G., 1991. Aggregate hierarchy in soils. Soil Research 29, 815–828.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: Community ecology package.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25, 1043–1055.

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M., Sarr, A., Maron, P.-A., 2013. Loss in microbial diversity affects nitrogen cycling in soil. The ISME Journal 7, 1609–1619.

Plaza, C., Courtier-Murias, D., Fernández, J.M., Polo, A., Simpson, A.J., 2013. Physical, chemical, and biochemical mechanisms of soil organic matter stabilization under

conservation tillage systems: A central role for microbes and microbial by-products in C sequestration. Soil Biology and Biochemistry 57, 124–134.

Prosser, J.I., Nicol, G.W., 2012. Archaeal and bacterial ammonia-oxidisers in soil: The quest for niche specialisation and differentiation. Trends in Microbiology 20, 523–531.

Qin, W., Amin, S.A., Martens-Habbena, W., Walker, C.B., Urakawa, H., Devol, A.H., Ingalls, A.E., Moffett, J.W., Armbrust, E.V., Stahl, D.A., 2014. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. Proceedings of the National Academy of Sciences 111, 12504–12509.

Rillig, M.C., Muller, L.A., Lehmann, A., 2017. Soil aggregates as massively concurrent evolutionary incubators. The ISME Journal 11, 1943–1948.

Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.-A., Davín, A.A., Waite, D.W., Whitman, W.B., Parks, D.H., Hugenholtz, P., 2021. A standardized archaeal taxonomy for the Genome Taxonomy Database. Nature Microbiology 6, 946–959.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43, e47.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rodrigues, J.L.M., Pellizari, V.H., Mueller, R., Baek, K., Jesus, E. da C., Paula, F.S., Mirza, B., Hamaoui, G.S., Tsai, S.M., Feigl, B., Tiedje, J.M., Bohannan, B.J.M., Nüsslein, K., 2013. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proceedings of the National Academy of Sciences 110, 988–993.

Samaddar, S., Karp, D.S., Schmidt, R., Devarajan, N., McGarvey, J.A., Pires, A.F.A., Scow, K., 2021. Role of soil in the regulation of human and plant pathogens: Soils' contributions to people. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20200179.

Schimel, J.P., 2018. Life in dry soils: Effects of drought on soil microbial communities and processes. Annual Review of Ecology, Evolution, and Systematics 49, 409–432.

Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49–56.

Sexstone, A.J., Revsbech, N.P., Parkin, T.B., Tiedje, J.M., 1985. Direct measurement of oxygen profiles and denitrification rates in soil aggregates. Soil Science Society of America Journal 49, 645–651.

Six, J., Elliott, E.T., Paustian, K., 2000. Soil macroaggregate turnover and microaggregate formation: A mechanism for C sequestration under no-tillage agriculture. Soil Biology and Biochemistry 32, 2099–2103.

Song, S.H., Vieille, C., 2009. Recent advances in the biological production of mannitol. Applied Microbiology and Biotechnology 84, 55–62.

Stieglmeier, M., Klingl, A., Alves, R.J.E., Rittmann, S.K.-M.R., Melcher, M., Leisch, N., Schleper, C. 2014, 2014. *Nitrososphaera viennensis* gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum *Thaumarchaeota*. International Journal of Systematic and Evolutionary Microbiology 64, 2738–2752.

Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. Journal of Soil Science 33, 141–163.

Totsche, K.U., Amelung, W., Gerzabek, M.H., Guggenberger, G., Klumpp, E., Knief, C., Lehndorff, E., Mikutta, R., Peth, S., Prechtel, A., Ray, N., Kögel-Knabner, I., 2017. Microaggregates in soils. Journal of Plant Nutrition and Soil Science 181, 104–146.

Tourna, M., Freitag, T.E., Nicol, G.W., Prosser, J.I., 2008. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environmental Microbiology 10, 1357–1364.

Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Urich, T., Engel, M., Schloter, M., Wagner, M., Richter, A., Schleper, C., 2011. *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. Proceedings of the National Academy of Sciences 108, 8420–8425.

Trivedi, P., Delgado-Baquerizo, M., Jeffries, T.C., Trivedi, C., Anderson, I.C., Lai, K., McNee, M., Flower, K., Pal Singh, B., Minkey, D., Singh, B.K., 2017. Soil aggregation and associated microbial communities modify the impact of agricultural management on carbon content. Environmental Microbiology 19, 3070–3086.

Trivedi, P., Rochester, I.J., Trivedi, C., Van Nostrand, J.D., Zhou, J., Karunaratne, S., Anderson, I.C., Singh, B.K., 2015. Soil aggregate size mediates the impacts of cropping regimes on soil carbon and microbial communities. Soil Biology and Biochemistry 91, 169–181.

Upton, R.N., Bach, E.M., Hofmockel, K.S., 2019. Spatio-temporal microbial community dynamics within soil aggregates. Soil Biology and Biochemistry.

Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K., Kyrpides, N.C., Pati, A., 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Research 43, 6761–6771.

Vos, M., Wolf, A.B., Jennings, S.J., Kowalchuk, G.A., 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiology Reviews 37, 936–954.

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

Wang, B., Brewer, P.E., Shugart, H.H., Lerdau, M.T., Allison, S.D., 2018. Soil aggregates as biogeochemical reactors and implications for soil-atmosphere exchange of greenhouse gases-a concept. Global Change Biology.

Wang, D., Lin, J.Y., Sayre, J.M., Schmidt, R., Fonte, S.J., Rodrigues, J.L.M., Scow, K.M., 2022. Compost amendment and cover cropping maintains soil structure and carbon storage by increasing available carbon and microbial biomass in agricultural soil – a six-year field study. Under Review.

Wang, L., Luo, X., Liao, H., Chen, W., Wei, D., Cai, P., Huang, Q., 2018. Ureolytic microbial community is modulated by fertilization regimes and particle-size fractions in a Black soil of Northeastern China. Soil Biology and Biochemistry 116, 171–178.

Waterfield, N.R., Bowen, D.J., Fetherston, J.D., Perry, R.D., ffrench-Constant, R.H., 2001. The tc genes of *Photorhabdus*: A growing family. Trends in Microbiology 9, 185–191.

Wertz, S., Degrange, V., Prosser, J.I., Poly, F., Commeaux, C., Freitag, T., Guillaumaud, N., Roux, X.L., 2006. Maintenance of soil functioning following erosion of microbial diversity. Environmental Microbiology 8, 2162–2169.

Wilpiszeski, R.L., Aufrecht, J.A., Retterer, S.T., Sullivan, M.B., Graham, D.E., Pierce, E.M., Zablocki, O.D., Palumbo, A.V., Elias, D.A., 2019. Soil aggregate microbial communities: Towards understanding microbiome interactions at biologically relevant scales. Applied and Environmental Microbiology 85, e00324-19.

Wilson, R.M., Tfaily, M.M., Kolton, M., Johnston, E.R., Petro, C., Zalman, C.A., Hanson, P.J., Heyman, H.M., Kyle, J.E., Hoyt, D.W., Eder, E.K., Purvine, S.O., Kolka, R.K., Sebestyen, S.D., Griffiths, N.A., Schadt, C.W., Keller, J.K., Bridgham, S.D., Chanton, J.P., Kostka, J.E., 2021. Soil metabolome response to whole-ecosystem warming at the Spruce and Peatland Responses under Changing Environments experiment. Proceedings of the National Academy of Sciences 118, e2004192118.

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. Genome Biology 20, 257.

Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A.F., Hoelzle, R.D., Lamberton, T.O., McCalley, C.K., Hodgkins, S.B., Wilson, R.M., Purvine, S.O., Nicora, C.D., Li, C., Frolking, S., Chanton, J.P., Crill, P.M., Saleska, S.R., Rich, V.I., Tyson, G.W., 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560, 49–54.

Wu, X., Peng, J., Liu, P., Bei, Q., Rensing, C., Li, Y., Yuan, H., Liesack, W., Zhang, F., Cui, Z., 2021. Metagenomic insights into nitrogen and phosphorus cycling at the soil aggregate scale driven by organic material amendments. Science of The Total Environment 785, 147329.

Wu, Y.-W., Simmons, B.A., Singer, S.W., 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y., 2012. clusterProfiler: An R package for comparing biological themes among gene clusters. Omics: A Journal of Integrative Biology 16, 284–287.

Yu, H., Ding, W., Chen, Z., Zhang, H., Luo, J., Bolan, N., 2015. Accumulation of organic C components in soil and aggregates. Scientific Reports 5, 13804.

Zhalnina, K., de Quadros, P., Gano, K., Davis-Richardson, A., Fagen, J., Brown, C., Giongo, A., Drew, J., Sayavedra-Soto, L., Arp, D., Camargo, F., Daroub, S., Clark, I., McGrath, S., Hirsch, P., Triplett, E., 2013. Ca. *Nitrososphaera* and *Bradyrhizobium* are inversely correlated and related to agricultural practices in long-term field experiments. Frontiers in Microbiology 4.

TABLES AND FIGURES

Table 4-1. Permutational multivariate analysis of variance (PERMANOVA) results for microbial taxonomy (a) and function (b) in soil aggregates.

a)

Taxonomy					
Variable	**Df**	**SumSqs**	**R^2**	**F**	**P-value**
Aggregate Size	3	0.373	0.065	1.346	***0.001**
Treatment	1	0.120	0.021	1.289	***0.001**
Aggregate Size:Treatment	3	0.289	0.050	1.041	0.069
Residuals	54	4.986	0.865		
Total	61	5.767	1.000		

b)

Function (KO)					
Variable	**Df**	**SumSqs**	**R^2**	**F**	**P-value**
Aggregate Size	3	0.394	0.105	2.362	***0.001**
Treatment	1	0.117	0.031	2.100	***0.002**
Aggregate Size:Treatment	3	0.228	0.061	1.363	***0.011**
Residuals	54	3.004	0.802		
Total	61	3.743	1.000		

Figure 4-1. Non-metric multidimensional scaling (NMDS) plots based on Bray-Curtis distances of microbial taxonomy (a) and function (b) in soil aggregates. Samples are colored by aggregate size and shapes correspond with different fertilizer treatments. Ordination stress values are 0.222 and 0.091 for taxonomy and function, respectively.

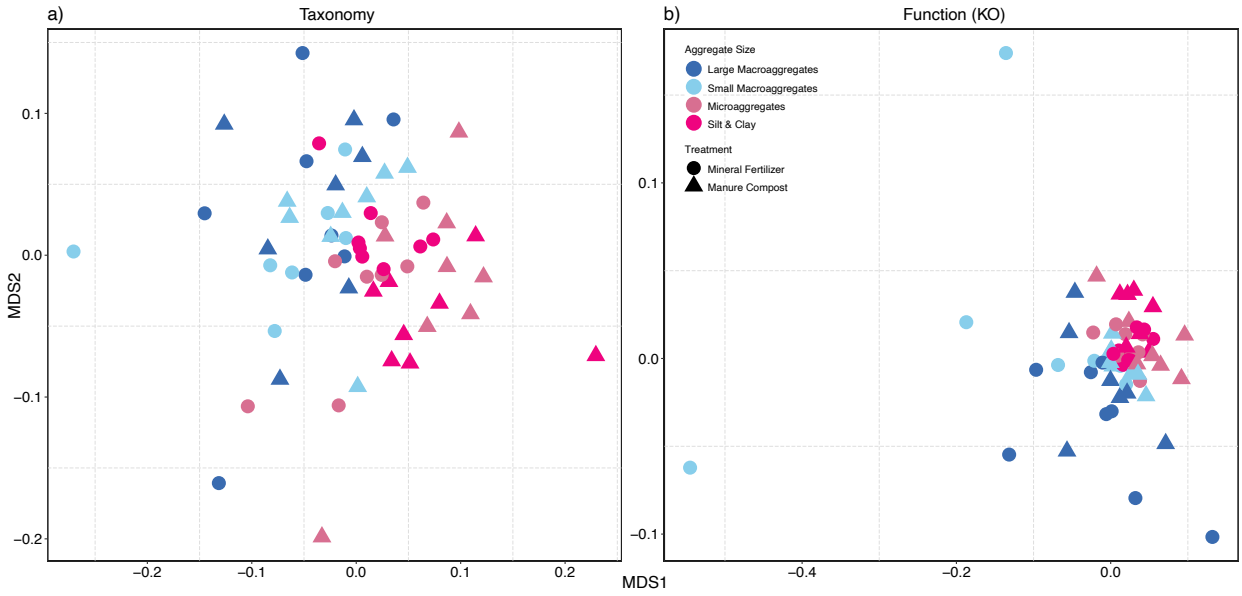


Figure 4-2. Alpha diversity of microbial taxonomy (a-e) and function (d-f) based on Richness, Shannon, and Evenness indexes in soil aggregates. Samples are colored by aggregate size and different letters represent significant differences ($P < 0.05$) between aggregate size fractions.

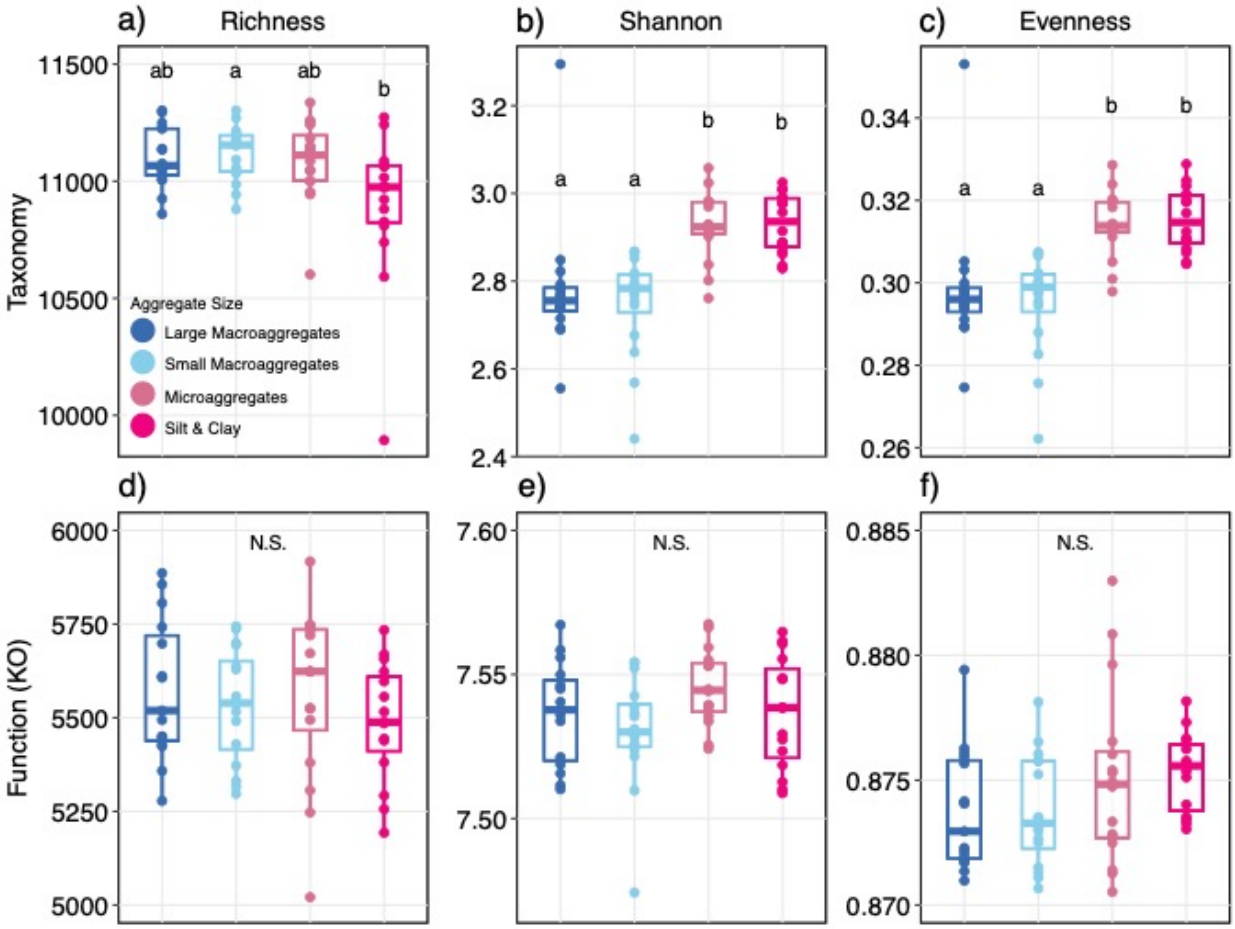


Figure 4-3. Heatmap showing taxa at the genus level that were differentially abundant ($P < 0.05$) between the large and small macroaggregates and microaggregates and silt & clay. Corresponding taxonomic ranks for order and phylum are given for each genus.

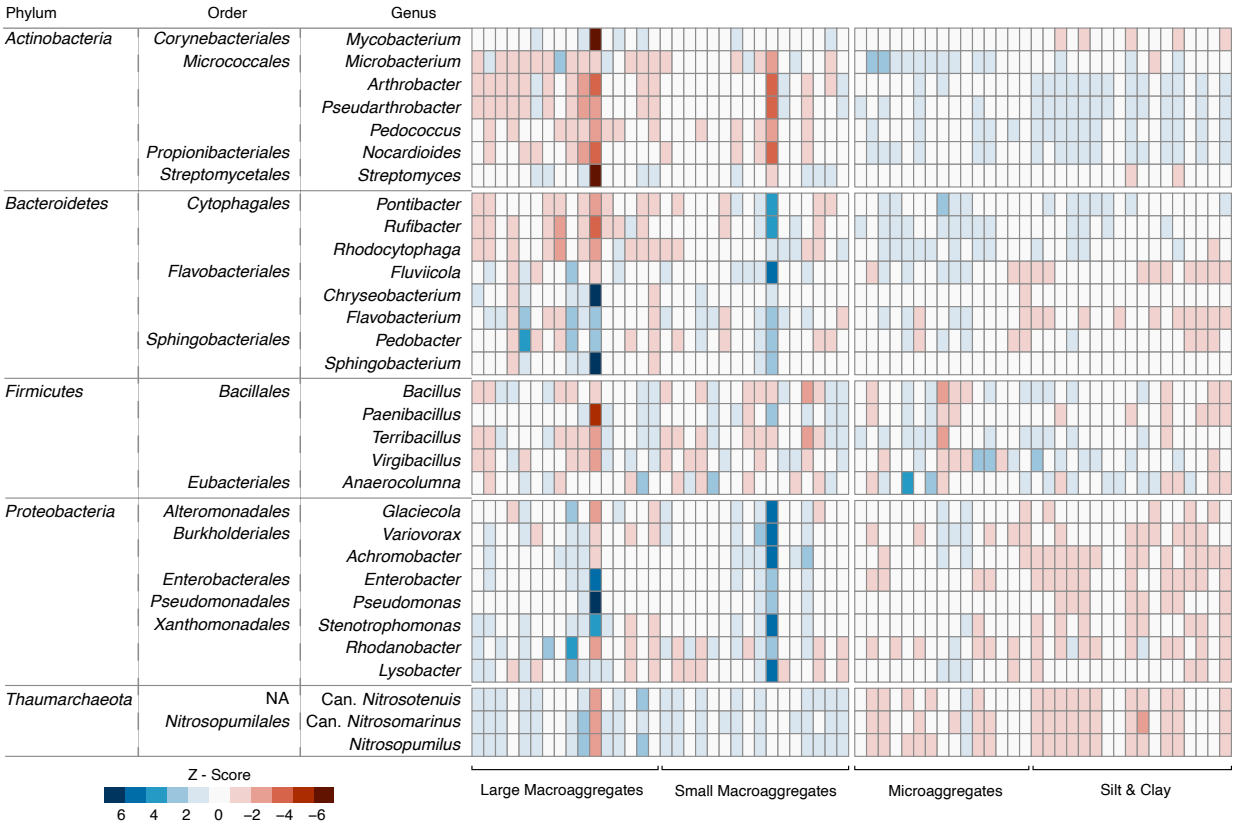


Figure 4-4. Heatmaps showing the a) functional genes in soil aggregates that were differentially abundant ($P < 0.05$) between the large and small macroaggregates and microaggregates and silt & clay and the b) percentage of metagenome assembled genomes (MAGs) within each phylum that contain each gene. The numbers in parentheses indicate the number of MAGs retrieved per phylum. Relevant KEGG pathways, modules, or enzyme names are given with the K numbers for each gene.

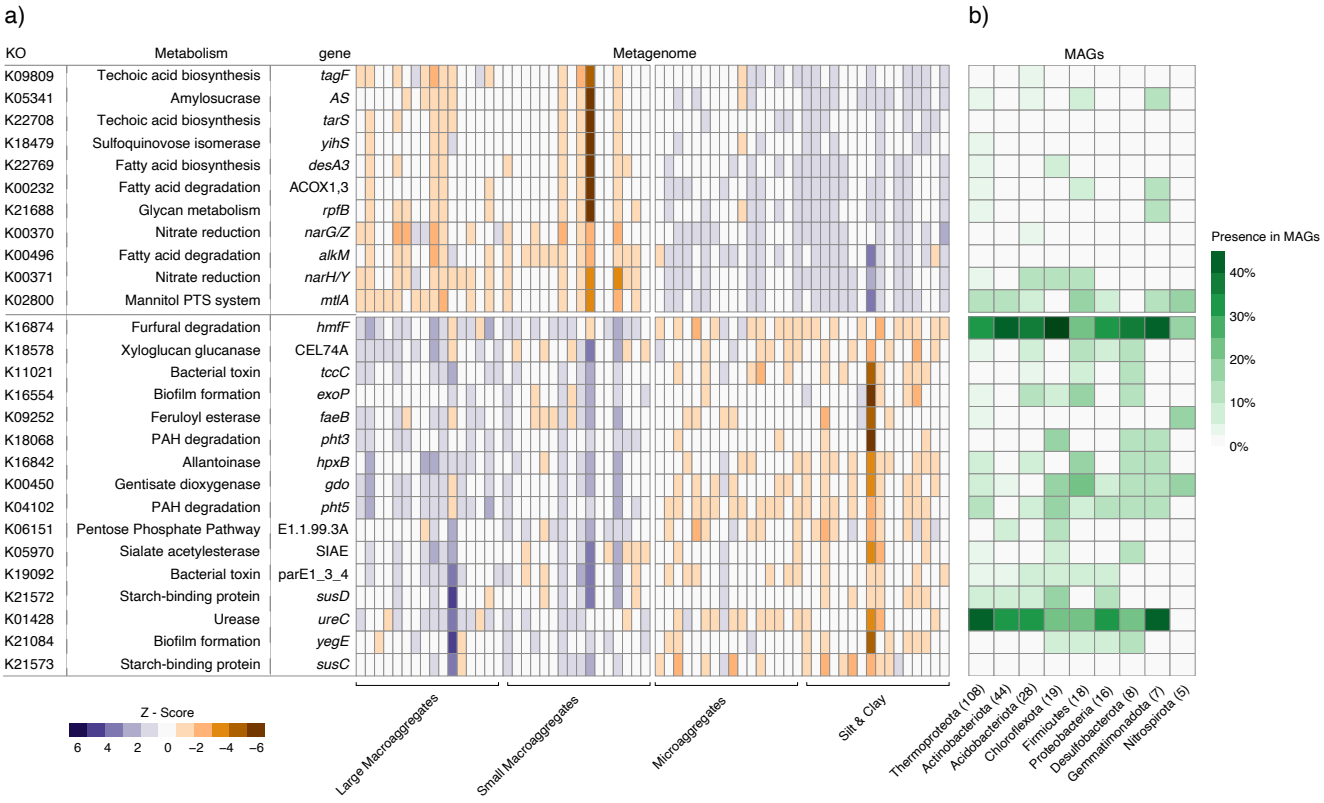
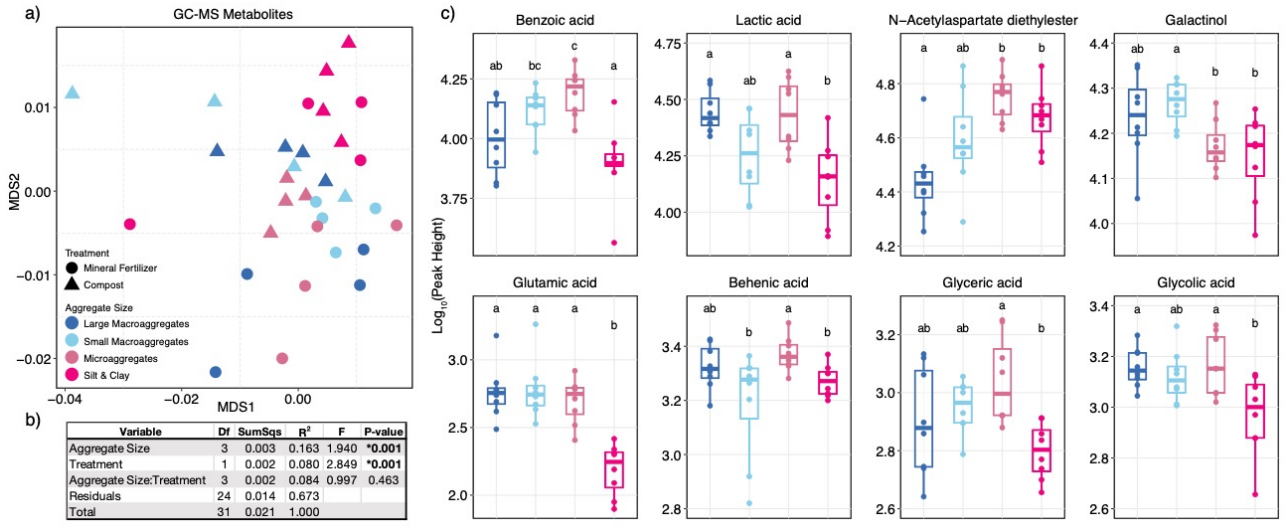
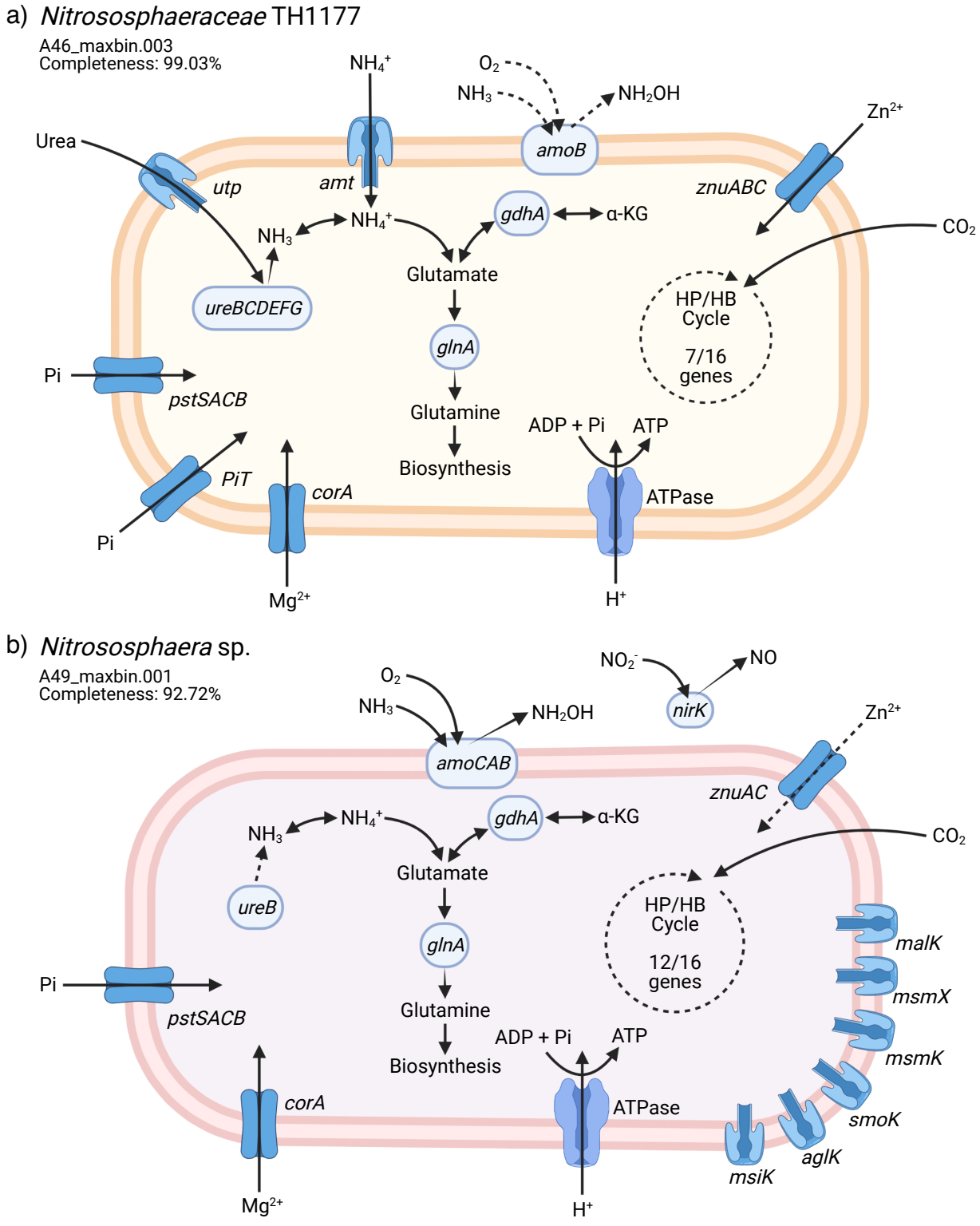


Figure 4-5. (a) Non-metric multidimensional scaling (NMDS) plot based on the Bray-Curtis distance of the metabolome of soil aggregates. Metabolites were identified by GC-MS and the ordination stress value was 0.171. (b) Permutational multivariate analysis of variance (PERMANOVA) results for the metabolome and (c) the top 8 identified metabolites that were differentially abundant based on log fold change. The remaining metabolites that were significantly different are shown in Fig. S3. Samples in the NMDS are colored by aggregate size and shapes correspond with different fertilizer treatments. Different letters in the boxplots represent significant differences ($P < 0.05$) between aggregate size fractions.



Variable	Df	SumSqs	R^2	F	P-value
Aggregate Size	3	0.003	0.163	1.940	***0.001**
Treatment	1	0.002	0.080	2.849	***0.001**
Aggregate Size:Treatment	3	0.002	0.084	0.997	0.463
Residuals	24	0.014	0.673		
Total	31	0.021	1.000		

Figure 4-6. Cell metabolism diagrams constructed from the MAGs of (a) A46_maxbin.003, which was classified as a species within the genus *Nitrososphaeraceae* TH1177, and (b) A49_maxbin.001, which was classified as a species within the genus *Nitrososphaera*. Dashed lines represent putative pathways due to incomplete operons or missing genes.



CHAPTER 3

SUPPLEMENTARY MATERIAL

Table S4-1. Metagenome assembly metrics of all samples used in this study.

Sample	Reads passing filter	Contigs (#)	Total length (bp)	Max. contig size (bp)	Avg. contig Size (bp)	N50
A1	93,068,016	408,657	720,777,762	66,484	1,764	1,726
A2	96,483,071	429,998	761,324,419	92,951	1,771	1,733
A3	88,264,717	371,426	639,173,091	69,751	1,721	1,669
A4	84,358,174	371,595	640,454,282	52,254	1,724	1,673
A5	83,701,364	364,059	619,293,274	52,212	1,701	1,650
A6	98,452,482	450,066	781,859,709	61,727	1,737	1,689
A7	81,309,855	339,679	578,017,829	59,468	1,702	1,646
A8	79,654,292	328,613	561,746,196	63,886	1,709	1,658
A9	82,490,151	342,004	602,316,569	73,303	1,761	1,712
A10	85,728,088	389,354	677,773,968	67,438	1,741	1,695
A11	76,999,523	293,301	497,961,869	65,610	1,698	1,637
A12	56,200,127	198,390	335,033,874	46,902	1,689	1,628
A13	80,699,188	354,840	613,951,980	60,341	1,730	1,679
A14	74,746,139	299,182	508,993,070	52,867	1,701	1,644
A15	92,664,020	368,254	625,650,265	62,283	1,699	1,642
A16	95,258,212	386,748	665,995,077	64,893	1,722	1,676
A17	89,830,850	356,082	610,969,904	50,049	1,716	1,668
A18	85,068,952	353,399	614,153,011	60,623	1,738	1,687
A19	91,955,110	346,781	589,662,339	72,638	1,700	1,645
A20	73,364,464	271,393	466,922,249	63,290	1,720	1,662
A21	73,273,468	301,636	514,483,438	62,127	1,706	1,649
A22	88,927,291	352,477	603,482,913	62,187	1,712	1,661
A23	78,653,266	285,858	484,651,061	66,059	1,695	1,633
A24	75,633,534	280,247	471,487,654	54,938	1,682	1,627
A25	97,469,181	449,521	798,580,696	64,333	1,777	1,742
A26	106,652,794	473,287	832,986,618	65,924	1,760	1,725
A27	83,809,942	333,195	567,738,222	61,651	1,704	1,645
A28	89,565,789	392,346	682,537,564	59,702	1,740	1,694
A29	82,254,539	341,799	584,275,980	65,925	1,709	1,659
A30	86,247,448	348,824	599,827,028	76,909	1,720	1,666
A31	77,747,030	265,642	450,021,323	92,267	1,694	1,632
A32	27,974,166	55,224	87,981,554	24,348	1,593	1,517
A33	75,093,158	244,992	411,408,434	63,396	1,679	1,613
A34	75,249,687	273,284	467,743,067	57,961	1,712	1,662
A35	75,378,768	252,217	424,007,364	59,301	1,681	1,618
A36	78,787,908	305,065	519,856,584	60,999	1,704	1,647
A37	85,835,752	399,565	700,883,231	86,352	1,754	1,713
A38	80,340,447	303,709	539,868,142	49,513	1,778	1,735
A39	54,799,017	184,998	308,331,957	53,885	1667	1,603
A40	71,419,450	291,586	499,900,323	59,178	1,714	1,656
A41	94,497,999	369,615	669,227,326	467,732	1,811	1,749
A42	93,724,606	400,107	691,174,139	74,112	1,727	1,682
A43	101,610,750	427,418	733,380,470	61,348	1,716	1,667
A44	87,602,634	369,022	634,789,340	64,933	1,720	1,671
A45	81,136,375	336,844	572,431,809	42,492	1,699	1,639
A46	79,270,921	315,016	539,344,879	67,001	1,712	1,658
A47	94,551,402	363,028	614,098,008	57,473	1,692	1,631
A48	76,807,798	275,816	463,433,179	61,023	1,680	1,614
A49	98,133,324	441,361	769,726,346	98,747	1,744	1,701
A50	88,492,880	372,559	641,393,435	63,551	1,722	1,685
A51	77,823,368	313,947	537,272,147	67,668	1,711	1,653
A52	65,127,954	267,208	453,898,238	57,822	1,699	1,641
A53	86,116,076	382,423	665,386,970	71,752	1,740	1,693
A54	92,272,665	415,047	723,651,756	53,935	1,744	1,697
A55	90,523,851	384,360	665,591,172	70,479	1,732	1,682
A56	78,539,170	308,870	526,429,786	53,688	1,704	1,644
A57	102,326,661	488,231	856,109,143	69,260	1,753	1,723
A58	96,226,982	447,141	787,761,723	60,487	1,762	1,723
A59	93,460,768	418,521	729,027,907	70,750	1,742	1,696
A60	77,579,732	327,854	562,489,066	61,012	1,716	1,665
A61	82,711,448	384,852	673,624,483	65,432	1,750	1,704
A62	97,496,814	428,067	748,300,399	83,007	1,748	1,710
A63	89,604,825	364,353	626,160,196	63,655	1,719	1,665
A64	96,172,004	397,690	691,187,243	74,390	1,738	1,689

Table S4-2. Differential abundance analysis ($P < 0.05$) of microbial taxonomy at the genus level (a), KEGG genes (b), and metabolites (c) by fertilizer management. A positive log_2-fold change value indicates enrichment under mineral fertilizer, while a negative log_2-fold change value indicates enrichment under manure compost.

a)

Taxonomy						
Genus	**Log_2FC**	**Log_2CPM**	**LR**	***P*-value**	**FDR**	**Classification**
Pseudomonas	-7.05	8.53	100.39	1.25E-23	3.74E-20	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae
Chryseobacterium	-9.15	7.99	141.72	1.12E-32	6.71E-29	Bacteria; Bacteroidetes; Flavobacteriia; Flavobacteriales; Weeksellaceae
Rhodococcus	-1.68	4.62	30.77	2.90E-08	3.47E-06	Bacteria; Actinobacteria; Actinomycetia; Corynebacteriales; Nocardiaceae
Sphingobacterium	-6.58	6.17	91.36	1.20E-21	2.39E-18	Bacteria; Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae
Pedobacter	-1.11	3.78	12.23	4.69E-04	1.05E-02	Bacteria; Bacteroidetes; Sphingobacteriia; Sphingobacteriales; Sphingobacteriaceae
Rhodanobacter	1.66	7.27	66.96	2.77E-16	2.07E-13	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Rhodanobacteraceae
Pantoea	1.13	4.50	9.89	1.66E-03	2.54E-02	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales; Erwiniaceae
Nitrosospira	1.51	4.80	66.00	4.51E-16	3.00E-13	Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales; Nitrosomonadaceae

b)

Function (KO)						
KEGG ID	**Log_2FC**	**Log_2CPM**	**LR**	***P*-value**	**FDR**	**Gene Name**
K00496	-0.48	6.17	20.42	6.22E-06	3.79E-03	alkane 1-monooxygenase [EC:1.14.15.3]
K06188	-0.42	6.11	47.85	4.59E-12	1.12E-08	aquaporin Z
K06895	-0.47	5.99	14.83	1.18E-04	1.65E-02	L-exporter protein LysE/ArgO
K18481	-0.46	7.22	18.50	1.70E-05	8.27E-03	Mce-membrane protein
K19577	-0.65	6.13	14.86	1.16E-04	1.65E-02	MFS transporter, family, membrane protein
K21672	-0.49	6.34	36.43	1.58E-09	1.93E-06	2,4-dehydrogenase [EC:1.4.1.1.4.1.26]

c)

Metabolites						
KEGG ID	**Log_2FC**	**Avg. Expression**	***t*-statistic**	***P*-value**	**Adj. *P*-value**	**Compound**
C00252	-472.81	467.66	-4.19	0.000	0.013	isomaltose
C00249	20180.38	161777.13	3.58	0.001	0.049	palmitic acid
C01753	-1901.69	1851.28	-3.50	0.001	0.049	beta-sitosterol

Table S4-3. Differentially abundant KEGG genes ($P < 0.05$) between aggregate size fractions. A positive log_2-fold change value indicates enrichment in the microaggregates and silt & clay, while a negative log_2-fold change value indicates enrichment in the large and small macroaggregates.

KEGG ID	Log_2 FC	Log_2 CPM	LR	*P*-value	FDR	Gene Name
K00127	-0.49	6.04	24.57	7.15E-07	8.11E-06	formate subunit gamma
K00150	-0.41	6.30	22.44	2.17E-06	2.01E-05	glyceraldehyde-3-dehydrogenase (NAD(P)) [EC:1.2.1.59]
K00232	0.46	7.65	35.00	3.29E-09	9.66E-08	acyl-oxidase [EC:1.3.3.6]
K00370	0.61	6.42	32.57	1.15E-08	2.68E-07	nitrate reductase / oxidoreductase, subunit [EC:1.7.5.1.7.99.-]
K00371	0.51	6.73	55.67	8.57E-14	1.31E-11	nitrate reductase / oxidoreductase, subunit [EC:1.7.5.1.7.99.-]
K00450	-0.53	6.75	46.97	7.20E-12	6.64E-10	gentisate 1,2-dioxygenase [EC:1.13.11.4]
K00496	0.67	6.16	52.99	3.35E-13	4.30E-11	alkane 1-monooxygenase [EC:1.14.15.3]
K00505	-0.57	6.32	46.93	7.35E-12	6.64E-10	tyrosinase [EC:1.14.18.1]
K00728	0.44	8.51	44.87	2.10E-11	1.77E-09	dolichyl-phosphate-mannose-mannosyltransferase [EC:2.4.1.109]
K00758	0.44	6.23	24.97	5.81E-07	6.79E-06	thymidine phosphorylase [EC:2.4.2.4]
K00868	0.52	6.23	24.97	5.82E-07	6.79E-06	pyridoxine kinase [EC:2.7.1.35]
K01153	-0.47	6.43	42.79	6.09E-11	4.01E-09	type restriction enzyme, subunit [EC:3.1.21.3]
K01301	-0.46	5.93	6.68	9.77E-03	2.46E-02	N-acetylated-alpha-acidic dipeptidase [EC:3.4.17.21]
K01302	-0.41	7.83	11.90	5.62E-04	2.27E-03	carboxypeptidase Q [EC:3.4.17.-]
K01428	-0.44	6.84	25.01	5.70E-07	6.78E-06	urease alpha [EC:3.5.1.5]
K01467	-0.51	5.72	7.74	5.41E-03	1.52E-02	beta-class C [EC:3.5.2.6]
K01576	-0.46	6.15	43.80	3.63E-11	2.68E-09	benzoylformate decarboxylase [EC:4.1.1.7]
K01886	-0.55	6.48	21.55	3.45E-06	2.96E-05	glutaminyl-synthetase [EC:6.1.1.18]
K01958	0.42	6.65	13.10	2.95E-04	1.31E-03	pyruvate carboxylase [EC:6.4.1.1]

K01989	-0.48	8.58	60.63	6.87E-15	2.10E-12	putative tryptophan/transport substrate-protein
K02005	-0.40	8.42	14.00	1.83E-04	8.86E-04	HlyD secretion protein
K02014	-0.50	9.96	7.16	7.47E-03	1.97E-02	iron outermembrane protein
K02117	-0.44	6.94	18.16	2.03E-05	1.38E-04	V/A-H+/Na+-ATPase A [EC:7.1.2.7.2.2.1]
K02172	-0.50	8.10	16.15	5.86E-05	3.39E-04	bla protein blaR1
K02322	-0.52	7.86	27.73	1.40E-07	2.17E-06	DNA II subunit [EC:2.7.7.7]
K02448	-0.45	6.21	15.52	8.15E-05	4.49E-04	nitric reductase protein
K02453	-0.67	5.33	16.70	4.38E-05	2.64E-04	general pathway D
K02481	-0.55	7.52	24.04	9.44E-07	1.01E-05	two-system, family, regulator
K02517	-0.54	6.01	14.25	1.60E-04	7.94E-04	Kdo2-IVA lauroyltransferase/acyltransferase [EC:2.3.1.2.3.1.-]
K02558	-0.52	5.80	11.63	6.50E-04	2.58E-03	UDP-N-acetylmuramate: L-alanyl-gamma-D-glutamyl-meso-ligase [EC:6.3.2.45]
K02668	-0.41	6.18	12.07	5.13E-04	2.12E-03	two-system, family, histidine PilS [EC:2.7.13.3]
K02674	-0.50	6.30	15.72	7.35E-05	4.10E-04	type pilus protein PilY1
K02683	-0.43	6.31	24.40	7.84E-07	8.69E-06	DNA small subunit [EC:2.7.7.102]
K02800	0.56	7.17	56.03	7.13E-14	1.16E-11	mannitol system or component [EC:2.7.1.197]
K03042	-0.45	7.99	27.93	1.26E-07	1.96E-06	DNA-RNA subunit A" [EC:2.7.7.6]
K03296	-0.59	7.24	9.57	1.98E-03	6.61E-03	hydrophobic/exporter-1 (G- bacteria), family
K03299	0.44	7.79	35.86	2.12E-09	6.55E-08	gluconate:H+ symporter, family
K03367	-1.27	5.34	21.09	4.38E-06	3.68E-05	D-alanine--poly(phosphoribitol) subunit 1 [EC:6.1.1.13]
K03427	-0.54	6.59	78.11	9.73E-19	7.91E-16	type restriction M protein [EC:2.1.1.72]
K03566	-0.42	6.82	5.59	1.80E-02	4.03E-02	LysR transcriptional regulator, cleavage transcriptional activator
K03583	0.40	8.08	19.49	1.01E-05	7.51E-05	exodeoxyribonuclease gamma subunit [EC:3.1.11.5]

K03648	0.45	6.49	28.04	1.19E-07	1.90E-06	uracil-glycosylase [EC:3.2.2.27]
K03654	-0.66	6.09	11.04	8.90E-04	3.39E-03	ATP-DNA RecQ [EC:3.6.4.12]
K03694	-0.57	5.48	8.76	3.08E-03	9.56E-03	ATP-Clp ATP-subunit ClpA
K04034	-0.54	7.40	25.10	5.45E-07	6.58E-06	anaerobic magnesium-IX ester cyclase [EC:1.21.98.3]
K04070	-0.43	6.42	42.25	8.03E-11	4.90E-09	putative formate activating enzyme [EC:1.97.1.4]
K04102	-0.54	7.69	81.89	1.44E-19	1.75E-16	4,5-decarboxylase [EC:4.1.1.55]
K05341	0.50	7.08	25.67	4.05E-07	5.22E-06	amylosucrase [EC:2.4.1.4]
K05365	-0.67	5.43	13.41	2.50E-04	1.14E-03	penicillin-protein 1B [EC:2.4.1.3.4.16.4]
K05830	-0.41	6.30	13.52	2.37E-04	1.09E-03	LysW-gamma-L-lysine/LysW-L-aminotransferase [EC:2.6.1.2.6.1.-]
K05831	-0.52	5.98	10.52	1.18E-03	4.31E-03	DEFINITION [group protein]-lysine/hydrolase [EC:3.5.1.3.5.1.132]
K05970	-0.52	5.65	15.82	6.98E-05	3.94E-04	sialate O-acetylesterase [EC:3.1.1.53]
K06006	-0.45	6.76	17.35	3.12E-05	2.02E-04	periplasmic CpxP/Spy
K06151	-0.49	6.33	24.26	8.43E-07	9.17E-06	gluconate 2-alpha chain [EC:1.1.99.3]
K06160	-1.14	4.55	18.85	1.42E-05	1.00E-04	putative transport ATP-binding/protein
K06219	0.47	6.27	21.01	4.57E-06	3.78E-05	S-adenosylmethionine-methyltransferase
K06894	-1.05	7.13	19.71	9.01E-06	6.78E-05	alpha-2-macroglobulin
K06895	0.50	5.99	17.15	3.45E-05	2.19E-04	L-exporter protein LysE/ArgO
K06909	-0.73	6.26	64.58	9.27E-16	3.77E-13	phage large subunit
K06965	-0.40	6.30	15.50	8.24E-05	4.53E-04	protein pelota
K06989	-0.41	6.72	57.73	3.00E-14	6.65E-12	aspartate dehydrogenase [EC:1.4.1.21]
K07003	-0.47	6.83	12.41	4.26E-04	1.80E-03	uncharacterized protein
K07165	-0.66	7.19	13.74	2.09E-04	9.90E-04	transmembrane sensor

K07277	-0.41	8.01	13.08	2.98E-04	1.32E-03	outer protein porin family
K07303	-0.62	6.57	11.24	8.03E-04	3.12E-03	isoquinoline 1-subunit beta [EC:1.3.99.16]
K07689	-0.42	5.82	14.05	1.78E-04	8.67E-04	two-system, family, response UvrY
K07705	-0.54	6.36	19.13	1.22E-05	8.87E-05	two-system, family, regulator LytT
K07713	-0.51	7.65	21.02	4.53E-06	3.77E-05	two-system, family, regulator HydG
K07714	-0.51	8.63	20.92	4.80E-06	3.90E-05	two-system, family, regulator AtoC
K07768	0.48	6.55	22.48	2.12E-06	1.97E-05	two-system, family, histidine SenX3 [EC:2.7.13.3]
K07787	-1.08	5.58	16.56	4.72E-05	2.81E-04	copper/efflux protein
K07795	-0.44	10.50	24.72	6.63E-07	7.62E-06	putative transport protein
K08191	-0.70	5.49	22.35	2.27E-06	2.09E-05	MFS transporter, family, transporter
K08303	-0.64	5.56	15.90	6.67E-05	3.78E-04	U32 peptidase [EC:3.4.-.-]
K08372	0.47	7.50	18.97	1.33E-05	9.60E-05	putative protease PepD [EC:3.4.21.-]
K08676	-0.78	7.15	10.93	9.44E-04	3.56E-03	tricorn protease [EC:3.4.21.-]
K09136	-0.41	7.39	46.13	1.11E-11	9.63E-10	ribosomal S12 accessory factor
K09252	-0.80	5.11	15.11	1.01E-04	5.39E-04	feruloyl esterase [EC:3.1.1.73]
K09781	0.45	6.71	35.89	2.09E-09	6.55E-08	uncharacterized protein
K09809	0.42	7.28	12.82	3.43E-04	1.48E-03	CDP-glycerophosphotransferase [EC:2.7.8.12]
K09861	0.53	6.53	33.26	8.08E-09	2.05E-07	uncharacterized protein
K10007	0.47	6.32	22.52	2.08E-06	1.95E-05	glutamate system protein
K10532	-0.54	6.15	21.57	3.42E-06	2.94E-05	heparan-alpha-N-acetyltransferase [EC:2.3.1.78]
K10843	0.49	7.31	36.98	1.20E-09	4.23E-08	DNA repair ERCC-3 [EC:3.6.4.12]
K11021	-1.32	5.80	34.53	4.20E-09	1.18E-07	insecticidal complex TccC

K11085	-0.77	5.76	17.74	2.53E-05	1.67E-04	ATP-cassette, B, MsbA [EC:7.5.2.6]
K11089	-0.51	5.90	22.03	2.68E-06	2.42E-05	60 SS-A/ribonucleoprotein
K11414	0.52	6.20	29.96	4.42E-08	7.98E-07	NAD+-protein sirtuin 4 [EC:2.3.1.286]
K11891	-0.62	6.29	14.33	1.53E-04	7.71E-04	type secretion protein ImpL
K11893	-0.48	5.57	12.93	3.23E-04	1.41E-03	type secretion protein ImpJ
K11895	-0.45	5.31	11.61	6.54E-04	2.59E-03	type secretion protein ImpH
K11904	-0.61	7.40	14.71	1.26E-04	6.50E-04	type secretion secreted VgrG
K12444	-1.43	6.80	29.76	4.88E-08	8.68E-07	phthiocerol/synthesis type-polyketide E [EC:2.3.1.292]
K12503	0.41	6.34	22.57	2.03E-06	1.91E-05	short-Z-diphosphate synthase [EC:2.5.1.68]
K13017	-0.43	6.44	18.13	2.06E-05	1.40E-04	UDP-2-acetamido-2-deoxy-ribo-aminotransferase [EC:2.6.1.98]
K13288	0.41	6.90	38.17	6.49E-10	2.68E-08	oligoribonuclease [EC:3.1.-.-]
K13572	0.44	6.79	25.54	4.34E-07	5.52E-06	proteasome factor B
K13573	0.40	6.55	22.74	1.86E-06	1.79E-05	proteasome factor C
K13992	-0.64	6.27	19.40	1.06E-05	7.77E-05	photosynthetic center c subunit
K14055	-0.50	6.21	13.61	2.25E-04	1.04E-03	universal protein E
K14645	-0.48	8.12	41.70	1.07E-10	6.04E-09	serine protease [EC:3.4.21.-]
K14954	0.45	6.28	42.18	8.33E-11	4.95E-09	lipoprotein LprG
K14986	-0.47	7.64	17.93	2.29E-05	1.55E-04	two-system, family, kinase FixL [EC:2.7.13.3]
K15016	-0.44	6.52	14.80	1.19E-04	6.23E-04	enoyl-hydratase / 3-hydroxyacyl-dehydrogenase [EC:4.2.1.1.1.1.35]
K15019	-0.51	6.19	36.22	1.76E-09	5.71E-08	3-hydroxypropionyl-A dehydratase [EC:4.2.1.116]
K15532	-0.46	6.08	12.08	5.09E-04	2.11E-03	unsaturated hydrolase [EC:3.2.1.172]
K15836	-0.56	7.09	13.15	2.88E-04	1.28E-03	formate transcriptional activator

K16090	-0.56	5.86	9.46	2.10E-03	6.91E-03	catecholate receptor
K16163	0.50	6.41	35.87	2.11E-09	6.55E-08	maleylpyruvate isomerase [EC:5.2.1.4]
K16554	-0.65	6.29	11.36	7.50E-04	2.93E-03	polysaccharide transport protein
K16649	0.59	6.53	34.28	4.78E-09	1.29E-07	rhamnopyranosyl-N-acetylglucosaminyl-diphospho-beta-1,3/1,4-galactofuranosyltransferase [EC:2.4.1.287]
K16650	0.63	6.78	39.80	2.81E-10	1.37E-08	galactofuranosylgalactofuranosylrhamnosyl-N-acetylglucosaminyl-diphospho-beta-1,5/1,6-galactofuranosyltransferase [EC:2.4.1.288]
K16842	-0.40	6.70	38.03	6.96E-10	2.75E-08	allantoinase [EC:3.5.2.5]
K16874	-0.46	7.11	66.69	3.18E-16	1.55E-13	2,5-decarboxylase 1
K17713	-0.55	8.96	16.88	3.99E-05	2.48E-04	outer protein factor BamB
K17734	-0.49	7.13	10.80	1.01E-03	3.79E-03	serine AprX [EC:3.4.21.-]
K17758	-0.64	5.30	13.52	2.36E-04	1.09E-03	ADP-NAD(P)H-dehydratase [EC:4.2.1.136]
K17837	-0.55	6.11	14.39	1.49E-04	7.51E-04	metallo-beta-class B [EC:3.5.2.6]
K18068	-0.79	6.23	36.32	1.68E-09	5.52E-08	phthalate 4,5-dioxygenase [EC:1.14.12.7]
K18138	-0.52	6.40	6.44	1.11E-02	2.75E-02	multidrug pump
K18139	-0.56	6.73	9.15	2.48E-03	7.89E-03	outer protein, efflux system
K18455	0.40	6.71	24.81	6.34E-07	7.36E-06	mycothiol S-amidase [EC:3.5.1.115]
K18479	0.49	6.13	16.16	5.82E-05	3.38E-04	sulfoquinovose isomerase [EC:5.3.1.31]
K18481	0.59	7.22	35.73	2.26E-09	6.81E-08	Mce-membrane protein
K18578	-0.48	5.90	18.87	1.40E-05	9.95E-05	xyloglucan-exo-beta-1,4-glucanase [EC:3.2.1.155]
K18601	-0.51	6.78	21.66	3.26E-06	2.83E-05	aldehyde dehydrogenase [EC:1.2.1.-]
K18691	-0.42	5.63	8.87	2.90E-03	9.08E-03	membrane-lytic transglycosylase F [EC:4.2.2.-]
K18926	0.49	6.27	8.41	3.73E-03	1.12E-02	MFS transporter, family, resistance protein
K19092	-0.61	5.83	32.62	1.12E-08	2.68E-07	toxin ParE1/3/4

K19701	-0.47	6.90	12.06	5.14E-04	2.12E-03	aminopeptidase YwaD [EC:3.4.11.3.4.11.10]
K21084	-0.51	6.81	12.38	4.33E-04	1.83E-03	diguanylate cyclase [EC:2.7.7.65]
K21572	-0.81	8.13	20.34	6.49E-06	5.05E-05	starch-outer protein, SusD/family
K21573	-2.90	6.29	29.19	6.55E-08	1.12E-06	TonB-starch-outer protein SusC
K21688	0.54	6.74	32.85	9.98E-09	2.43E-07	resuscitation-factor RpfB
K22447	-0.42	7.84	36.79	1.32E-09	4.53E-08	archaeal chaperonin
K22476	0.43	6.70	43.43	4.39E-11	3.15E-09	N-synthase [EC:2.3.1.1]
K22486	-0.71	5.94	11.03	8.98E-04	3.42E-03	transcriptional HilA, transcriptional of SPI1
K22708	0.40	6.69	13.73	2.11E-04	9.93E-04	poly(ribitol-phosphate) beta-N-acetylglucosaminyltransferase [EC:2.4.1.355]
K22769	0.51	6.56	21.91	2.86E-06	2.55E-05	NADPH-stearoyl-9-desaturase [EC:1.14.19.-]
K22770	0.47	7.09	26.20	3.08E-07	4.26E-06	stearoyl-9-NADPH oxidoreductase
K23424	-0.71	6.06	22.03	2.69E-06	2.42E-05	protein O-mannosyl-transferase [EC:2.4.1.-]
K23842	0.47	6.52	31.77	1.73E-08	3.72E-07	NAD(P)dehydrogenase (quinone) [EC:1.6.5.2]
K23980	0.44	6.82	32.08	1.48E-08	3.28E-07	cysteinylglycine-S-dipeptidase [EC:3.4.13.23]
K24017	0.54	6.10	27.55	1.53E-07	2.33E-06	phosphoribosyl A [EC:5.3.1.5.3.1.24]

Table S4-4. Summary metrics and taxonomic classifications of metagenome assembled genomes (MAGs) used in this study.

MAG ID	Completion (%)	Contamination (%)	Heterogeneity (%)	Genome Size (Mbp)	GC Content (%)	GTDB Classification						
						Domain	Phylum	Class	Order	Family	Genus	Species
A1_maxbin.008	99.03	8.74	8.33	3.61	28.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A1_maxbin.002	69.42	5.55	41.67	1.57	50.1	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A1_maxbin.038	40.75	8.93	14.29	3.02	69.7	Bacteria	Actinobacteriota	UBA4738	UBA4738	HRBIN12	AC-51	
A1_maxbin.123	40.15	3.42	50.00	3.61	73.0	Bacteria	Actinobacteriota	Actinomycetia				
A2_maxbin.001	92.72	4.85	0.00	1.76	29.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A2_maxbin.017	64.07	2.95	25.00	2.02	55.9	Bacteria	Proteobacteria	Gammaproteobacteria	JACCYU01	JACCYU01		
A3_maxbin.012	99.03	3.88	0.00	2.54	58.1	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A3_maxbin.004	58.73	2.90	29.41	2.39	38.5	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A3_maxbin.151	49.12	3.51	50.00	2.31	70.5	Bacteria	Actinobacteriota	Actinomycetia	Actinomycetales	Microbacteriaceae	Agromyces	
A3_maxbin.117	49.10	7.83	33.33	3.49	72.9	Bacteria	Actinobacteriota	Actinomycetia				
A3_maxbin.005	40.76	8.02	10.71	1.83	56.2	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	OLB17	
A4_maxbin.005	99.03	4.85	0.00	3.00	28.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A4_maxbin.004	69.82	7.00	55.00	1.56	40.0	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A4_maxbin.047	63.35	4.85	22.22	5.03	33.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	JAFAQB01	
A4_maxbin.045	55.68	6.70	33.33	4.03	75.3	Bacteria	Actinobacteriota					
A4_maxbin.041	46.63	6.11	18.67	3.03	66.4	Bacteria	Actinobacteriota	Actinomycetia	Mycobacteriales	Mycobacteriaceae	Mycobacterium	
A4_maxbin.003	41.93	5.49	40.00	3.04	57.3	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	PSRF01	
A4_maxbin.094	40.27	5.56	8.33	4.55	72.6	Bacteria	Actinobacteriota	Actinomycetia				
A5_maxbin.007	55.17	5.17	0.00	2.57	70.6	Bacteria	Chloroflexota	Limnocylindria	Limnocylindrales	CSP1-4	SPCO01	
A5_maxbin.003	51.70	4.65	0.00	2.74	68.8	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JABFSM01	

A5_maxbin.012	43.10	3.45	100.00	1.78	69.6	Bacteria	Chloroflexota	Limnocylindria	Limnocylindrales	CSP1-4			
A6_maxbin.001	99.03	5.65	27.27	1.95	50.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera		
A6_maxbin.002	85.28	6.31	0.00	1.69	30.1	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177		
A6_maxbin.003	49.23	6.44	16.00	1.65	68.6	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738			
A6_maxbin.036	48.04	6.03	0.00	3.45	65.9	Bacteria	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Steroidobacteraceae			
A6_maxbin.022	46.14	8.19	0.00	2.38	68.1	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610			
A7_maxbin.004	99.03	2.91	0.00	2.86	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177		
A7_maxbin.001	60.36	0.49	100.00	1.07	50.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera		
A7_maxbin.012	40.57	8.62	13.33	2.40	58.0	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae			
A8_maxbin.003	93.20	2.91	0.00	2.96	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177		
A8_maxbin.057	73.38	7.77	18.18	4.79	33.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	JAFAQB01		
A8_maxbin.004	65.10	2.18	18.75	2.56	38.5	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium	
A8_maxbin.076	59.22	8.25	11.11	4.86	36.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH5893		
A8_maxbin.005	44.20	2.14	33.33	2.33	57.2	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae			
A9_maxbin.005	92.23	3.88	0.00	2.96	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177		
A9_maxbin.030	60.29	7.68	7.14	3.66	59.5	Bacteria	Acidobacteriota	Blastocatellia	RBC074	RBC074			
A10_maxbin.003	99.03	6.80	30.00	3.01	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177		
A10_maxbin.001	98.22	3.62	42.86	1.80	51.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera		
A10_maxbin.010	52.72	5.73	12.50	2.16	68.2	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610			
A11_maxbin.005	90.29	2.91	0.00	2.57	28.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177		
A11_maxbin.043	55.17	9.48	14.29	4.22	74.9	Bacteria	Actinobacteriota	Actinomycetia					
A11_maxbin.001	49.39	2.68	28.57	2.07	53.3	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	WHTF01		
A11_maxbin.037	47.30	7.32	4.35	4.60	68.0	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	UBA2999			

A11_maxbin.004	40.65	2.99	0.00	0.64	54.0	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A11_maxbin.010	40.56	1.13	0.00	1.14	40.3	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A11_maxbin.002	40.09	0.00	0.00	0.74	50.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A12_maxbin.002	87.38	3.88	0.00	2.68	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A12_maxbin.003	55.99	5.33	11.11	3.82	51.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A12_maxbin.047	46.55	3.45	33.33	3.67	74.8	Bacteria	Actinobacteriota					
A12_maxbin.020	40.65	7.89	0.00	3.78	67.5	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	UBA2999		
A13_maxbin.008	99.03	6.80	0.00	3.35	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A13_maxbin.080	40.52	9.48	44.44	3.67	75.5	Bacteria	Actinobacteriota	Actinomycetia				
A14_maxbin.004	99.03	3.88	0.00	2.96	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A14_maxbin.001	95.15	5.45	62.50	1.76	51.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A15_maxbin.002	99.03	4.85	0.00	3.09	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A15_maxbin.001	87.38	6.95	50.00	1.76	51.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A15_maxbin.065	60.34	8.62	28.57	5.17	74.7	Bacteria	Actinobacteriota	Actinomycetia				
A15_maxbin.005	55.78	1.70	23.08	1.89	38.6	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A15_maxbin.010	43.23	4.60	66.67	2.55	55.4	Bacteria	Chloroflexota	Anaerolineae	Anaerolineales	EnvOPS12	UBA12294	
A16_maxbin.010	94.17	4.85	0.00	3.09	28.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A16_maxbin.001	93.13	8.78	37.50	1.79	51.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A16_maxbin.008	86.97	8.70	18.52	3.77	38.5	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A16_maxbin.006	58.99	7.75	20.00	3.22	69.1	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JABFSM01	
A16_maxbin.011	46.06	6.81	7.69	2.59	57.3	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae		
A16_maxbin.040	42.47	1.74	25.00	2.26	65.5	Bacteria	Chloroflexota	Chloroflexia	54-19	JADMIH01		
A17_maxbin.008	99.03	3.88	0.00	3.01	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	

A17_maxbin.001	59.55	0.97	0.00	0.86	52.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A17_maxbin.011	53.09	4.76	10.00	2.69	57.5	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae	
A17_maxbin.019	43.27	7.31	12.50	1.04	39.1	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia
A18_maxbin.007	96.76	5.88	45.45	3.35	51.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A18_maxbin.001	48.29	1.99	0.00	1.58	68.5	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738	
A19_maxbin.004	99.03	7.77	16.67	3.17	28.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A20_maxbin.002	99.03	3.88	0.00	2.81	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A20_maxbin.057	54.62	3.45	0.00	3.71	75.0	Bacteria	Actinobacteriota	Actinomycetia			
A20_maxbin.001	51.80	2.59	20.00	0.82	53.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A20_maxbin.008	40.96	2.02	33.33	0.86	50.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A21_maxbin.003	99.03	5.83	14.29	2.94	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A21_maxbin.064	53.45	2.59	50.00	4.19	74.9	Bacteria	Actinobacteriota	Actinomycetia			
A21_maxbin.001	51.15	2.34	50.00	0.80	53.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A21_maxbin.002	43.90	0.97	0.00	0.96	50.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A22_maxbin.001	84.79	3.88	0.00	1.57	30.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A22_maxbin.019	44.58	7.59	0.00	1.38	68.2	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738	
A22_maxbin.036	41.22	6.41	11.11	2.74	63.1	Bacteria	Proteobacteria	Gammaproteobacteria			
A23_maxbin.009	99.03	2.91	0.00	3.26	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A23_maxbin.010	53.41	8.87	21.88	1.69	57.5	Bacteria	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira
A23_maxbin.061	49.76	9.48	12.50	4.11	74.9	Bacteria	Actinobacteriota	Actinomycetia			
A23_maxbin.032	42.92	8.40	20.00	2.37	62.9	Bacteria	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Steroidobacteraceae	
A24_maxbin.013	99.03	5.83	0.00	3.15	28.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177

A24_maxbin.002	88.04	2.99	25.00	1.86	52.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A24_maxbin.003	49.17	0.96	53.85	1.47	40.0	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A25_maxbin.030	41.15	5.98	0.00	2.66	64.3	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610		
A26_maxbin.039	46.32	6.03	20.00	3.72	70.8	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	Schlegelella	
A26_maxbin.015	45.59	1.52	37.50	1.89	56.0	Bacteria	Proteobacteria	Gammaproteobacteria	JACCYU01	JACCYU01		
A26_maxbin.003	44.01	3.60	0.00	3.12	57.2	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	DP-20	
A27_maxbin.002	79.94	2.91	0.00	1.38	30.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A27_maxbin.003	70.02	6.14	42.86	1.76	40.0	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A27_maxbin.074	62.92	6.84	28.57	4.45	74.9	Bacteria	Actinobacteriota	Actinomycetia				
A27_maxbin.001	54.56	5.95	35.71	1.05	50.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A28_maxbin.013	99.03	5.34	33.33	3.24	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A28_maxbin.005	63.30	9.26	0.00	2.72	61.2	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales			
A28_maxbin.002	43.84	8.50	22.73	1.38	56.4	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	OLB17	
A28_maxbin.040	43.75	7.99	73.33	2.00	69.0	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JACDDX01	
A28_maxbin.010	41.01	7.88	33.33	5.49	54.3	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	PSRF01	
A29_maxbin.006	99.03	4.85	0.00	3.11	28.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A29_maxbin.001	94.66	4.99	45.45	1.74	51.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A29_maxbin.002	63.93	6.25	33.33	3.18	53.4	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	WHTF01	
A29_maxbin.011	63.14	6.36	0.00	3.21	49.7	Bacteria	Chloroflexota	Anaerolineae	Anaerolineales	EnvOPS12	UBA12294	
A29_maxbin.004	48.29	2.51	0.00	1.45	68.5	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738		
A29_maxbin.010	44.33	8.56	0.00	2.74	69.2	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JABFSM01	
A29_maxbin.008	41.82	5.81	18.75	2.10	56.7	Bacteria	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira	
A30_maxbin.002	99.03	3.88	0.00	3.07	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	

A30_maxbin.064	48.65	1.21	0.00	4.29	68.3	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	UBA2999		
A30_maxbin.001	46.92	4.31	13.33	1.36	68.3	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738		
A31_maxbin.002	99.03	4.85	0.00	3.10	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A31_maxbin.004	49.67	9.26	50.00	3.94	50.0	Bacteria	Chloroflexota	Anaerolineae	Anaerolineales	EnvOPS12	UBA12294	
A31_maxbin.006	48.02	0.97	0.00	1.02	52.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A32_maxbin.008	57.51	2.43	25.00	1.53	35.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH5893	
A32_maxbin.002	44.47	7.88	4.76	2.47	54.1	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	WHTF01	
A33_maxbin.006	100.00	6.80	11.11	3.01	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A33_maxbin.100	43.10	7.76	20.00	3.39	68.5	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Rhodanobacter	sp018599185
A33_maxbin.005	42.87	7.39	0.00	2.81	55.3	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	WHTF01	
A34_maxbin.006	99.03	3.88	0.00	2.88	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A34_maxbin.001	66.67	3.24	40.00	0.97	52.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A34_maxbin.058	63.40	8.62	50.00	4.69	74.8	Bacteria	Actinobacteriota	Actinomycetia				
A34_maxbin.051	45.01	1.72	0.00	1.72	64.3	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	Microvirga	
A35_maxbin.005	93.20	3.88	0.00	2.81	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A35_maxbin.002	63.63	6.77	10.34	3.03	53.3	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	WHTF01	
A35_maxbin.001	58.03	3.24	25.00	0.90	53.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A36_maxbin.003	94.82	2.91	0.00	2.87	28.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A36_maxbin.004	67.70	6.04	17.39	2.55	38.5	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A36_maxbin.024	51.42	9.02	10.91	2.53	62.5	Bacteria	Proteobacteria	Gammaproteobacteria	Steroidobacterales	Steroidobacteraceae	PALSA-1196	
A36_maxbin.059	50.86	6.03	0.00	4.13	75.0	Bacteria	Actinobacteriota	Actinomycetia				
A37_maxbin.017	99.03	9.71	0.00	3.73	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	

A37_maxbin.002	92.31	8.44	25.00	3.85	68.8	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JABFSM01	
A37_maxbin.018	67.56	3.88	0.00	1.53	34.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TA-21	sp014523495
A37_maxbin.047	59.02	5.50	57.14	2.35	68.7	Bacteria	Chloroflexota	Limnocylindria	Limnocylindrales	CSP1-4		
A37_maxbin.044	42.44	7.77	9.09	2.27	40.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH5896	
A37_maxbin.055	42.23	3.40	20.00	1.87	35.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	JAFAQB01	
A38_maxbin.005	77.35	3.88	0.00	2.52	28.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A38_maxbin.053	49.34	4.44	20.00	3.42	63.4	Bacteria	Chloroflexota	Chloroflexia	54-19	JADMIH01	JADMIH01	
A38_maxbin.027	48.10	9.37	16.13	2.51	59.4	Bacteria	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira	
A38_maxbin.054	44.83	9.48	14.29	3.30	74.5	Bacteria	Actinobacteriota	UBA4738				
A38_maxbin.019	44.36	7.98	9.09	3.13	57.7	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales			
A38_maxbin.012	43.77	7.85	7.69	3.53	57.4	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	DP-20	
A38_maxbin.033	43.37	2.27	0.00	1.53	32.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A39_maxbin.004	92.23	2.91	0.00	2.79	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A39_maxbin.002	58.46	7.43	15.79	2.85	53.2	Bacteria	Desulfobacterota	Binatia	UBA9968	UBA9968	WHTF01	
A39_maxbin.001	51.78	1.94	0.00	0.85	53.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A39_maxbin.009	48.02	9.03	12.50	3.03	56.8	Bacteria	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira	
A39_maxbin.003	44.01	0.00	0.00	0.86	50.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A39_maxbin.071	42.73	7.01	14.29	1.18	64.0	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomicrobium	
A40_maxbin.001	89.97	8.33	10.00	1.73	51.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A40_maxbin.034	48.89	5.67	50.00	4.17	67.6	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales			
A41_maxbin.022	98.54	6.80	0.00	3.64	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A41_maxbin.010	95.98	8.10	1.47	4.31	67.0	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	maltophilia
A41_maxbin.042	65.95	7.95	60.00	2.68	68.5	Bacteria	Chloroflexota	Limnocylindria	Limnocylindrales	CSP1-4		

A41_maxbin.016	65.25	6.69	21.28	3.35	63.5	Bacteria	Actinobacteriota	Actinomycetia	Mycobacteriales	Mycobacteriaceae	Rhodococcus	erythropolis
A41_maxbin.018	64.24	6.47	18.60	3.37	43.2	Bacteria	Firmicutes	Bacilli	Bacillales	DSM-1321	Peribacillus	
A41_maxbin.040	52.59	5.17	33.33	2.43	70.7	Bacteria	Chloroflexota	Limnocylindria	Limnocylindrales	CSP1-4	SPCO01	
A41_maxbin.041	41.91	9.33	6.67	3.40	59.4	Bacteria	Acidobacteriota	Blastocatellia	RBC074			
A41_maxbin.125	40.61	6.55	31.25	3.76	72.1	Bacteria	Actinobacteriota	Actinomycetia				
A42_maxbin.003	100.00	8.25	0.00	3.47	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A42_maxbin.006	77.98	4.22	26.19	3.29	38.5	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A42_maxbin.008	46.02	4.85	0.00	2.19	68.4	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610		
A43_maxbin.010	99.03	2.91	0.00	3.43	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A43_maxbin.002	66.69	5.49	72.73	1.29	50.1	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A44_maxbin.064	44.77	3.43	0.00	4.08	66.4	Bacteria	Chloroflexota	Chloroflexia	54-19	JADMIH01		
A45_maxbin.007	99.03	5.83	11.11	3.28	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A45_maxbin.009	96.41	3.92	28.57	4.65	38.4	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A45_maxbin.001	65.53	0.97	0.00	1.01	52.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A45_maxbin.072	57.76	7.76	40.00	5.00	74.6	Bacteria	Actinobacteriota	Actinomycetia				
A45_maxbin.025	55.97	6.92	22.22	2.04	58.5	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	OLB17	
A45_maxbin.107	48.28	6.03	25.00	3.50	70.2	Bacteria	Chloroflexota	Limnocylindria	QHBO01	QHBO01	JACDBZ01	
A45_maxbin.039	45.93	8.66	39.39	2.08	62.0	Bacteria	Actinobacteriota	Actinomycetia	Propionibacteriales	Propionibacteriaceae		
A45_maxbin.078	43.04	4.85	0.00	4.04	32.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	JAFAQB01	
A45_maxbin.084	40.21	9.39	25.00	2.97	36.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH5893	
A46_maxbin.003	99.03	5.83	0.00	2.99	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A46_maxbin.071	42.30	9.48	16.67	3.30	69.8	Bacteria	Chloroflexota	Limnocylindria	QHBO01	QHBO01	JACDBZ01	
A47_maxbin.005	94.17	3.88	0.00	2.94	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A47_maxbin.008	78.79	6.91	46.15	1.91	39.8	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium

A47_maxbin.001	64.08	3.24	25.00	0.94	52.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A48_maxbin.006	86.23	3.88	0.00	2.79	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A48_maxbin.001	75.51	1.94	0.00	1.30	52.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A48_maxbin.066	59.83	7.12	0.00	4.69	74.4	Bacteria	Actinobacteriota					
A48_maxbin.092	46.55	4.31	33.33	4.03	70.2	Bacteria	Chloroflexota	Limnocylindria	QHBO01	QHBO01	JACDBZ01	
A48_maxbin.047	40.82	3.45	100.00	3.38	66.9	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	UBA2999		
A49_maxbin.001	92.72	1.94	0.00	1.80	51.3	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A49_maxbin.064	64.66	8.62	25.00	4.76	74.8	Bacteria	Actinobacteriota	Actinomycetia				
A49_maxbin.008	60.45	5.08	12.50	2.62	68.2	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610		
A49_maxbin.002	56.18	0.94	33.33	1.97	38.8	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A49_maxbin.015	50.05	2.81	21.43	1.76	55.8	Bacteria	Proteobacteria	Gammaproteobacteria	JACCYU01	JACCYU01		
A50_maxbin.010	99.03	5.83	0.00	3.15	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A50_maxbin.001	76.91	0.49	0.00	1.31	52.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A51_maxbin.004	99.03	4.85	0.00	2.91	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A51_maxbin.001	49.84	0.97	0.00	0.81	50.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A51_maxbin.002	47.44	0.00	0.00	0.78	53.1	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A52_maxbin.114	56.36	5.50	21.43	2.57	63.9	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	Microvirga	
A52_maxbin.001	46.34	0.00	0.00	0.68	51.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A52_maxbin.036	44.17	6.90	0.00	3.49	67.9	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	UBA2999		
A53_maxbin.010	99.03	6.87	12.50	3.35	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A53_maxbin.006	54.41	3.08	28.57	1.89	38.6	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A53_maxbin.008	50.62	8.12	0.00	2.49	68.2	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610		

A54_maxbin.003	50.06	4.39	19.05	1.61	68.4	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738	
A54_maxbin.020	46.57	3.03	7.14	1.72	56.4	Bacteria	Proteobacteria	Gammaproteobacteria	JACCYU01	JACCYU01	
A54_maxbin.010	41.71	6.21	50.00	2.79	55.1	Bacteria	Chloroflexota	Anaerolineae	Anaerolineales	EnvOPS12	UBA12294
A55_maxbin.002	99.03	5.34	14.29	3.28	28.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A55_maxbin.024	58.06	7.41	30.77	3.01	68.5	Bacteria	Chloroflexota	Limnocylindria	Limnocylindrales	CSP1-4	
A55_maxbin.007	50.48	3.86	26.32	2.57	57.0	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae	
A56_maxbin.002	93.20	2.91	0.00	2.90	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A56_maxbin.020	74.55	8.28	11.11	2.37	57.9	Bacteria	Nitrospirota	Nitrospiria	Nitrospirales	Nitrospiraceae	Nitrospira
A56_maxbin.006	56.02	7.87	29.41	2.94	69.0	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JABFSM01
A56_maxbin.001	48.07	1.94	0.00	0.86	53.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A56_maxbin.004	47.86	1.46	50.00	0.92	50.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A57_maxbin.007	71.19	8.45	33.33	3.66	57.2	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae	
A57_maxbin.091	68.22	9.22	80.00	1.22	56.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A57_maxbin.013	61.76	5.43	14.63	2.13	55.8	Bacteria	Proteobacteria	Gammaproteobacteria	JACCYU01	JACCYU01	
A57_maxbin.014	57.12	5.83	42.86	1.07	52.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A57_maxbin.056	54.76	8.97	17.39	3.13	65.4	Bacteria	Chloroflexota	Chloroflexia	54-19	JADMIH01	
A57_maxbin.024	45.08	3.45	0.00	2.09	61.1	Bacteria	Acidobacteriota	Vicinamibacteria	Vicinamibacterales	2-12-FULL-66-21	
A57_maxbin.044	41.69	9.48	42.86	2.33	75.3	Bacteria	Actinobacteriota	Acidimicrobiia	Acidimicrobiales		
A57_maxbin.016	41.06	4.53	0.00	0.77	50.2	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A58_maxbin.010	100.00	8.25	0.00	3.20	28.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177
A58_maxbin.004	63.80	8.83	9.09	2.84	68.3	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610	
A58_maxbin.011	48.66	9.63	14.29	1.74	70.4	Bacteria	Actinobacteriota	UBA4738	UBA4738	UBA4738	
A58_maxbin.049	43.83	6.90	100.00	2.69	66.8	Bacteria	Actinobacteriota	Actinomycetia	Actinomycetales	Micrococcaceae	Arthrobacter

A58_maxbin.033	43.69	8.01	70.00	0.95	45.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A59_maxbin.008	99.03	6.80	12.50	3.05	28.6	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A59_maxbin.001	93.04	4.85	16.67	1.76	51.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A59_maxbin.002	52.13	2.16	28.57	1.67	39.9	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A59_maxbin.021	45.66	5.49	12.50	3.39	68.9	Bacteria	Gemmatimonadota	Gemmatimonadetes	Gemmatimonadales	GWC2-71-9	JABFSM01	
A59_maxbin.010	45.58	6.84	9.52	2.32	57.4	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales			
A59_maxbin.031	40.13	9.71	72.73	0.55	44.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A60_maxbin.003	88.35	4.85	0.00	2.84	28.5	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A60_maxbin.001	53.48	3.88	25.00	0.88	53.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A61_maxbin.005	99.03	6.80	0.00	3.33	28.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A61_maxbin.001	94.82	8.78	67.50	1.87	51.4	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	
A61_maxbin.004	81.37	6.49	38.10	3.26	38.5	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Priestia	megaterium
A61_maxbin.002	61.54	7.93	2.86	3.03	68.2	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610		
A61_maxbin.010	50.11	6.55	13.04	2.96	57.0	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae		
A61_maxbin.028	48.27	3.29	0.00	2.64	65.4	Bacteria	Chloroflexota	Chloroflexia	54-19	JADMIH01		
A61_maxbin.034	47.42	3.97	18.18	3.24	60.5	Bacteria	Acidobacteriota	Blastocatellia	RBC074	RBC074		
A61_maxbin.019	43.65	4.78	66.67	2.13	55.4	Bacteria	Acidobacteriota	Blastocatellia	Pyrinomonadales	Pyrinomonadaceae	PSRF01	
A61_maxbin.047	42.04	7.76	22.22	2.83	67.2	Bacteria	Actinobacteriota	Actinomycetia	Actinomycetales	Micrococcaceae	Arthrobacter	
A62_maxbin.012	100.00	7.28	11.11	3.60	28.9	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	
A62_maxbin.009	64.18	5.67	18.75	3.43	57.0	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales	Koribacteraceae		
A62_maxbin.045	50.68	4.57	0.00	2.91	65.5	Bacteria	Chloroflexota	Chloroflexia	54-19	JADMIH01		
A62_maxbin.010	42.00	3.27	11.54	1.64	55.4	Bacteria	Proteobacteria	Gammaproteobacteria	JACCYU01	JACCYU01		
A63_maxbin.004	99.03	4.85	0.00	2.98	28.7	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	TH1177	

A63_maxbin.001	88.03	6.91	43.75	1.63	51.8	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera
A63_maxbin.059	54.31	3.45	0.00	4.74	74.8	Bacteria	Actinobacteriota				
A63_maxbin.012	46.59	9.31	0.00	2.68	57.4	Bacteria	Acidobacteriota	Acidobacteriae	Acidobacteriales		
A63_maxbin.013	43.34	6.90	0.00	1.96	68.2	Bacteria	Actinobacteriota	Acidimicrobiia	IMCC26256	PALSA-610	
A63_maxbin.057	40.61	6.80	27.27	3.12	35.0	Archaea	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	JAFAQB01

Figure S4-1. Alpha diversity of microbial taxonomy (a-c) and function (d-f) based on Richness, Shannon, and Evenness indexes by fertilizer treatment across all samples. Asterisks represent significant differences ($P < 0.05$) between fertilizer treatments, while N.S. represents no significance.

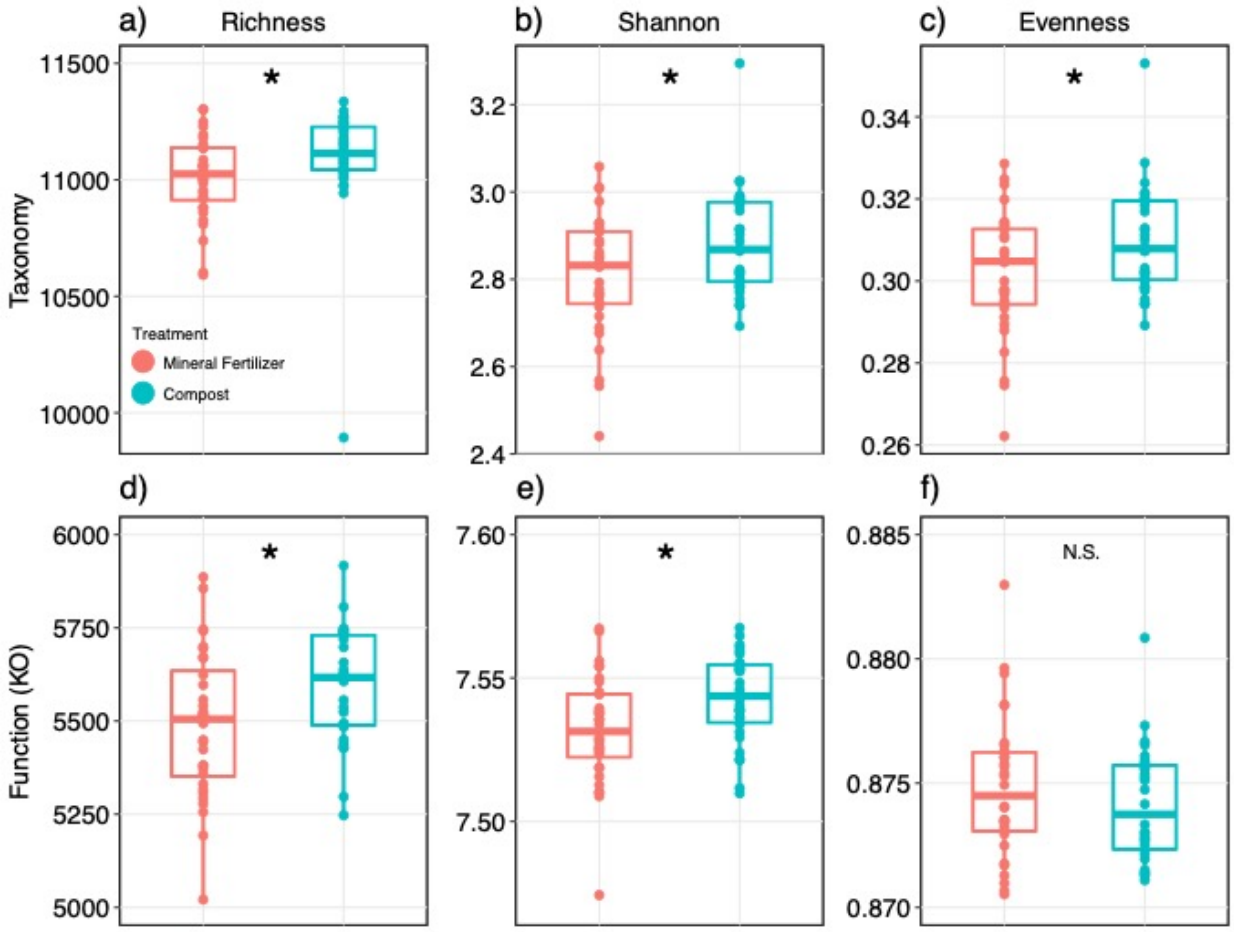


Figure S4-2. Pathway analysis showing enriched KEGG pathways based on differentially abundant KO terms. Representative K numbers for each pathway are shown on the x axis. A positive log_2-fold change value indicates enrichment in the microaggregates and silt & clay, while a negative log_2-fold change value indicates enrichment in the large and small macroaggregates.

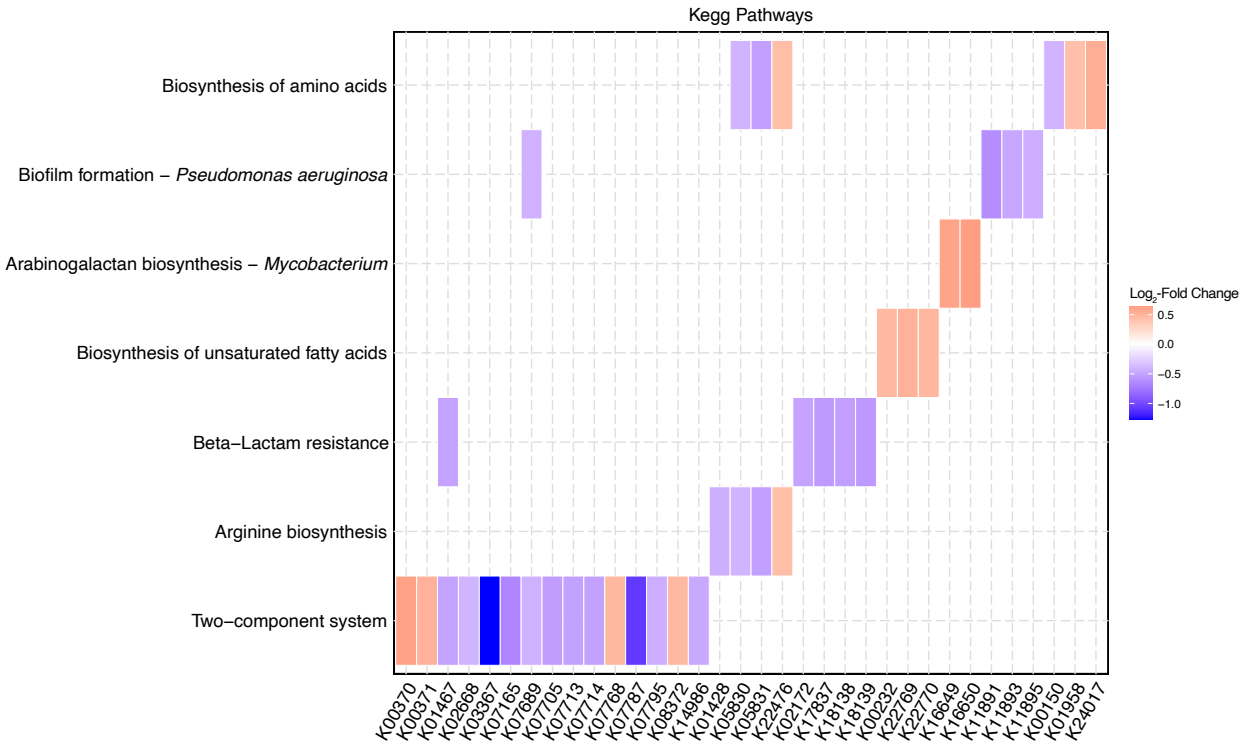


Figure S4-3. Taxonomic coverage of genes that were differentially abundant between the large and small macroaggregates and microaggregates and silt & clay. The number of phyla that possess each gene in their representative MAGs were plotted with the gene names in Fig. 4-4a shown. The asterisk indicates statistical significance ($P < 0.05$).

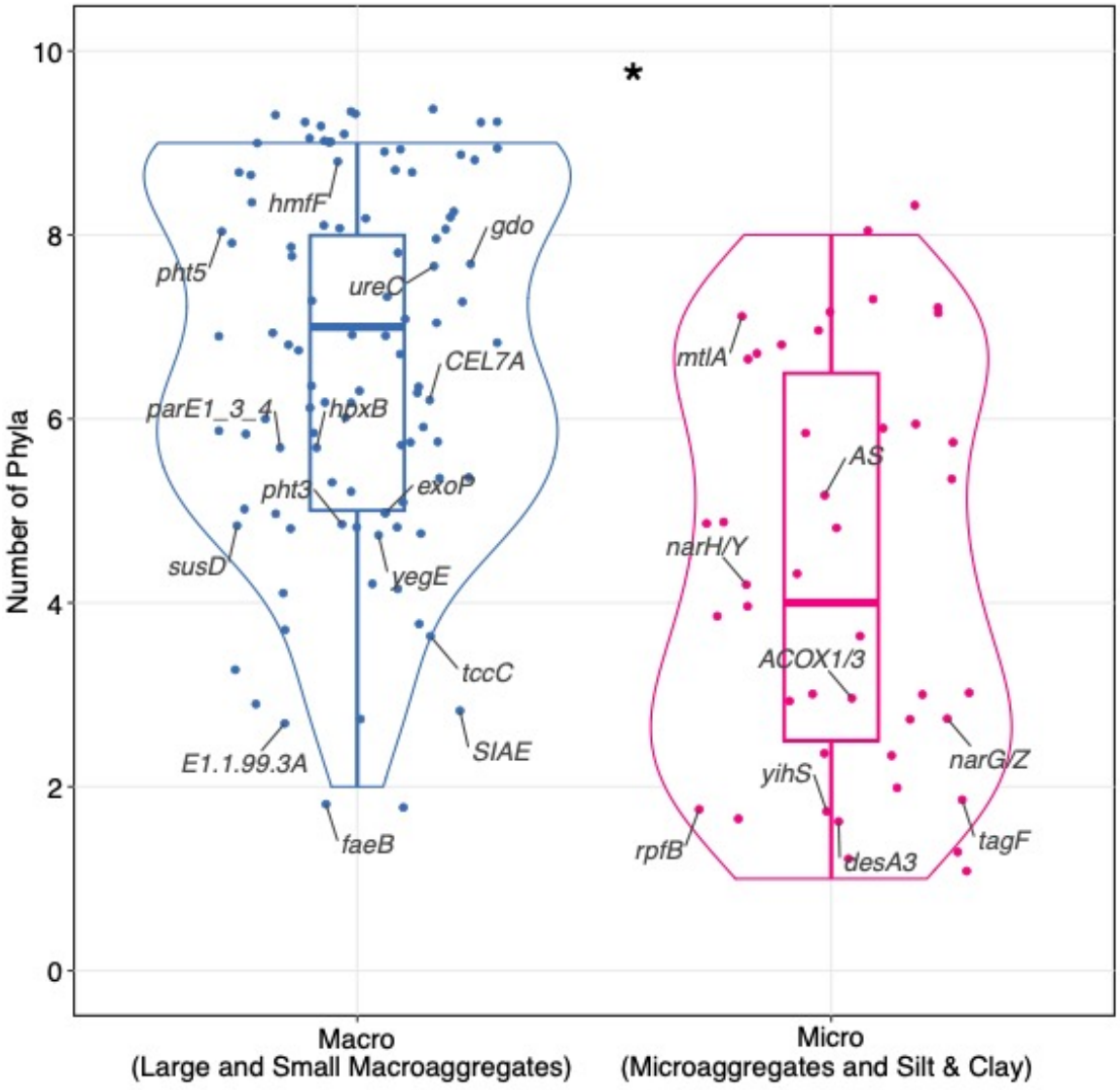


Figure S4-4. Differentially abundant metabolites by aggregate size. Different letters represent significant differences ($P < 0.05$) between aggregate size fractions.

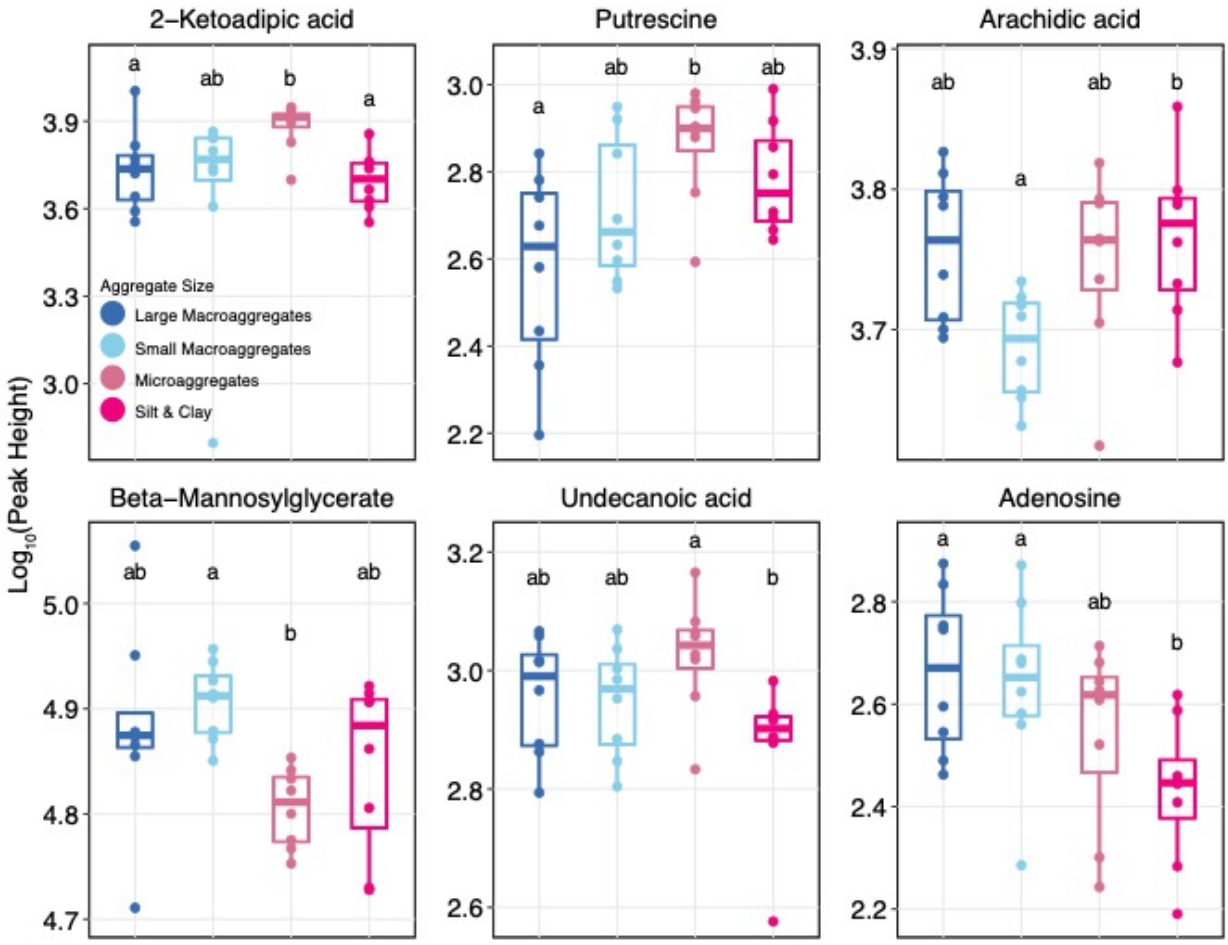
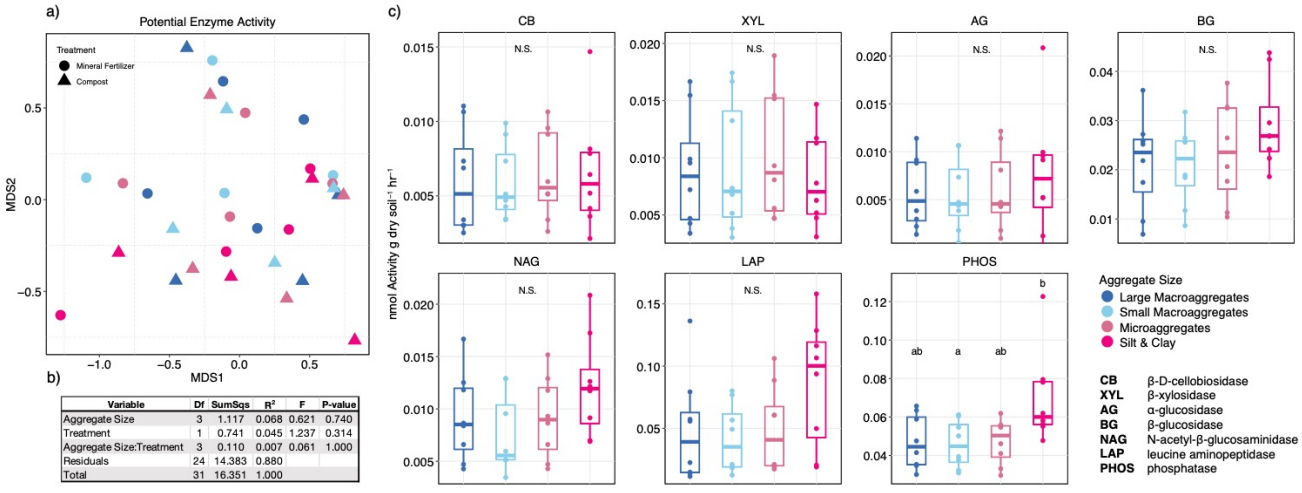


Figure S4-5. (a) Non-metric multidimensional scaling (NMDS) plot based on the Euclidean distance of enzyme activities of soil aggregates. The ordination stress value was 0.035. (b) Permutational multivariate analysis of variance (PERMANOVA) results for the enzymes and (c) the specific activities for each of the 7 enzymes assayed. Samples in the NMDS are colored by aggregate size and shapes correspond with different fertilizer treatments. Different letters in the boxplots represent significant differences ($P < 0.05$) between aggregate size fractions. Abbreviations for each enzyme are listed in the legend.



Variable	Df	SumSqs	R^2	F	P-value
Aggregate Size	3	1.117	0.068	0.621	0.740
Treatment	1	0.741	0.045	1.237	0.314
Aggregate Size:Treatment	3	0.110	0.007	0.061	1.000
Residuals	24	14.383	0.880		
Total	31	16.351	1.000		

Figure S4-6. Unrooted phylogenetic trees of bacterial (a) and archaeal (b) metagenome assembled genomes (MAGs) used in this study. The trees were constructed using multiple sequence alignments of 120 and 53 single-copy marker genes for bacteria and archaea, respectively. (c) Number of MAGs retrieved for each phylum by aggregate size.

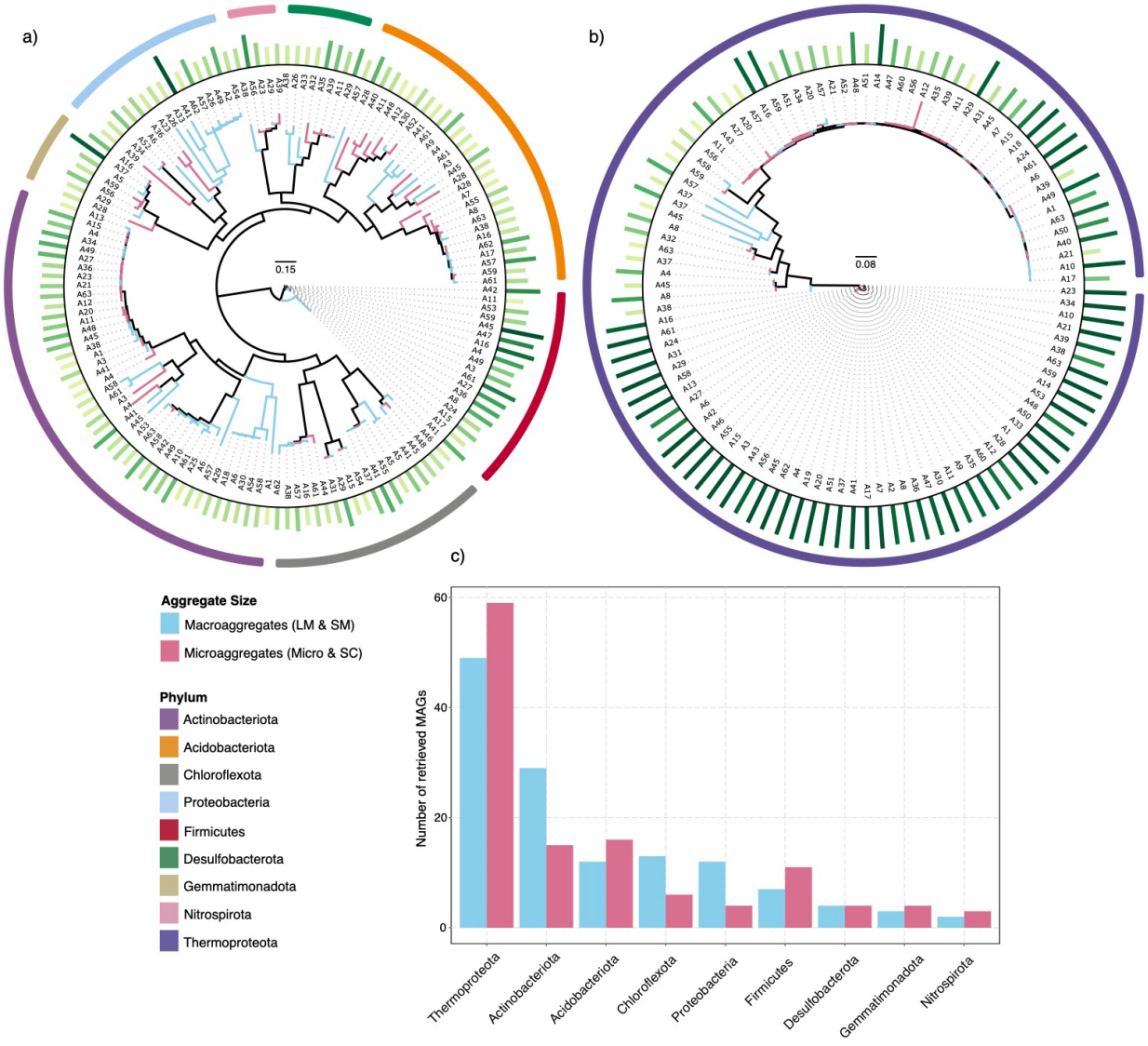


Figure S4-7. Phylogenetic tree of high-quality archaea MAGs used in this study. The tree, which was constructed using multiple sequence alignments of 53 single-copy marker genes, was trimmed to visualize MAG placement within the order *Nitrososphaerales*. Genome taxonomy database (GTDB) names are given at branch nodes and numbers indicate bootstrap support. Bar, 0.1 substitutions per position.

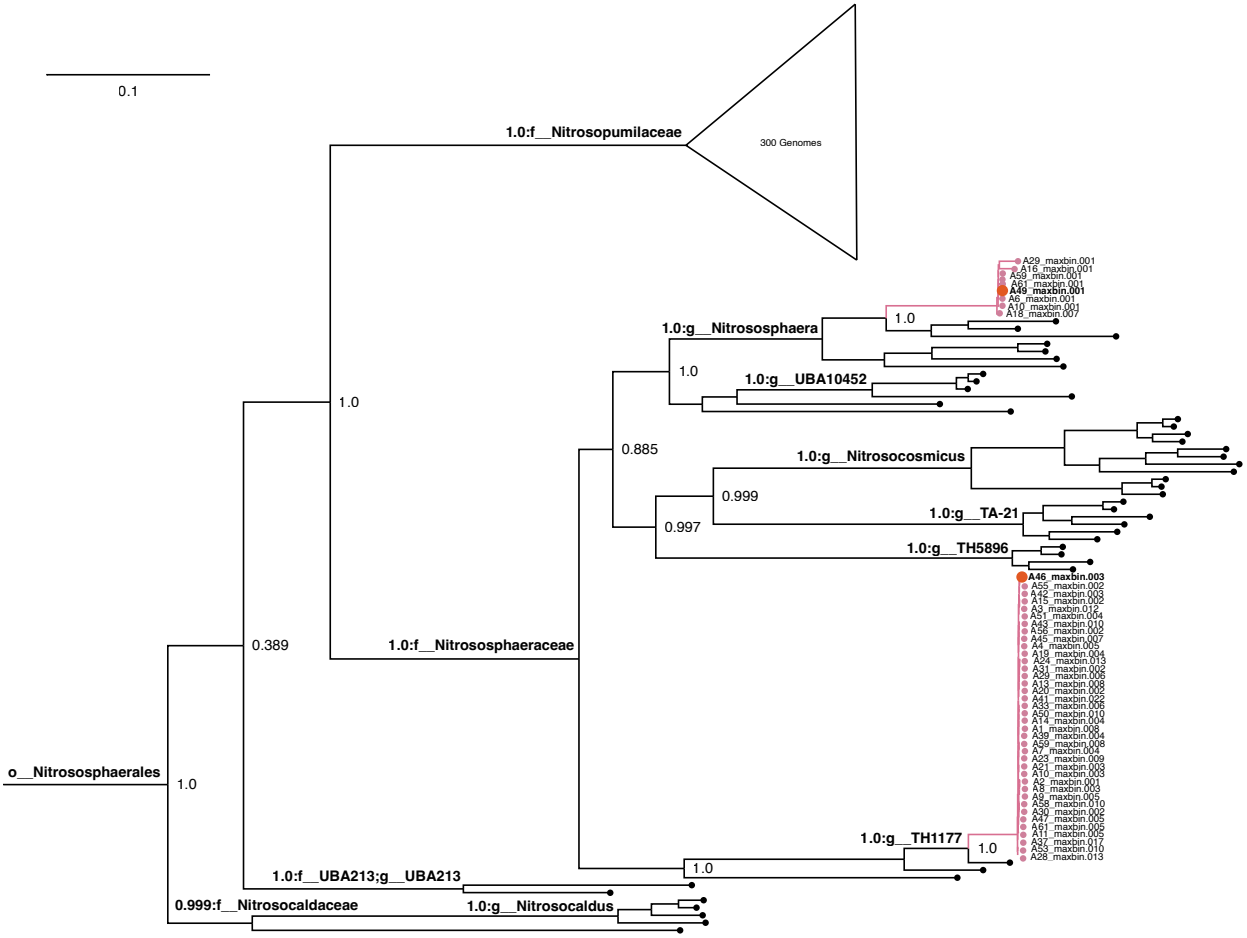
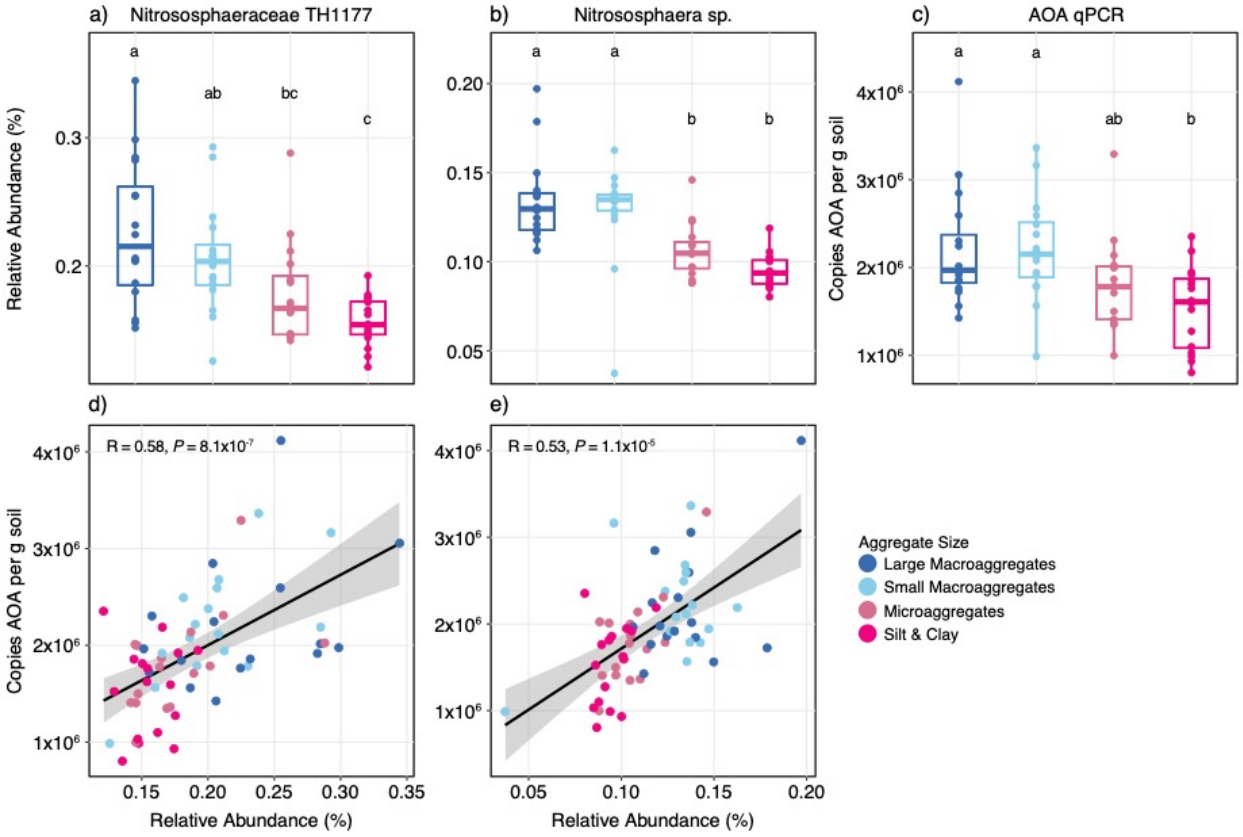


Figure S4-8. Relative abundances of the representative MAGs for (a) *Nitrososphaeraceae* TH1177 (A46_maxbin.003) and (b) *Nitrososphaera* sp. (A49_maxbin.001) in aggregates. (c) Quantification of ammonia-oxidizing archaea (AOA) in aggregates and spearman correlations of AOA copy numbers with each MAG (d & e). Different letters in the boxplots represent significant differences ($P < 0.05$) between aggregate size fractions.



APPENDIX A

GEMINISPHAERA

List of Authors: Jonathan Y. Lin and Jorge L.M. Rodrigues

Published in *Bergey's Manual of Systematics of Archaea and Bacteria*. 2020.
DOI: 10.1002/9781118960608.gbm01542

ABSTRACT

Cells are Gram-negative, non-motile, coccoid, and occur almost exclusively in pairs. Cells are facultatively anaerobic and microaerophilic with fermentative and respiratory metabolisms, respectively. Cells grow under air with the highest cell yields and shortest generation times at 2-8% O_2. Fermentation products from glucose are propionate, acetate, succinate, and CO_2. Sulfate and nitrate are not reduced with glucose as the electron donor. Growth occurs at temperatures between 15 and 35°C, pH values between 5.5 and 7.5, and a NaCl concentration of 1.5% (w/v). Capable of utilizing a number of monosaccharides, disaccharides, polysaccharides, and one or more components in yeast extract/casamino acids. Grows on a variety of media types. Capable of diazotrophic growth in a defined nitrogen-free medium.

Wertz et al., 2018b (effective publication: Wertz et al., 2018a)

Ge.mi.ni.sphae'ra. L. masc. adj. *geminus*, twin-born, twin; L. fem. n. *sphaera*, globe, sphere; N.L. fem. n. *Geminisphaera*, twin sphere.

DESCRIPTION

Cells are Gram-negative, non-motile, coccoid, and occur almost exclusively in pairs. Cells are facultatively anaerobic and microaerophilic with fermentative and respiratory

metabolisms. respectively. Cells grow under air with the highest cell yields and shortest generation times at 2-8% O_2. Fermentation products from glucose are propionate, acetate, succinate, and CO_2. Sulfate and nitrate are not reduced with glucose as the electron donor. Growth occurs at temperatures between 15 and 35°C, pH values between 5.5 and 7.5, and a NaCl concentration of 1.5% (w/v). Capable of utilizing a number of monosaccharides, disaccharides, polysaccharides, and one or more components in yeast extract/casamino acids. Grows on a variety of media types. Capable of diazotrophic growth in a defined nitrogen-free medium.

DNA G + C content (*mol %*): 60.8 (High-quality draft genome sequence).

Type species: *Geminisphaera colitermitum* Wertz et al., 2018b (effective publication: Wertz et al., 2018a)

Number of species with validated names: 1

Family classification: *Opitutaceae* (fbm00254)

FURTHER DESCRIPTIVE INFORMATION

Cell morphology and ultrastructure

The only currently described species of the genus *Geminisphaera*, *G. colitermitum* (type strain TAV2^T), was isolated from the hindgut of the Eastern subterranean termite, *Reticulitermes flavipes* (Kollar) (Stevenson et al., 2004). Cells are Gram-negative cocci, 0.5 to 0.6 µm in diameter, and occur almost exclusively in pairs (**Figure A-1**). Well-defined inner and outer membranes are visible under thin-section microscopy (Wertz et al., 2012). An extracellular polysaccharide capsule is not produced (Wertz et al., 2012). The cell morphology is consistent with cell morphologies observed in other genera within the family *Opitutaceae* – including *Opitutus*, '*Lacunisphaera*,' *Alterococcus*, *Cephaloticoccus*, *Ereboglobus*, and *Oleiharenicola* (Rast et al.,

2017; Shieh and Jean, 1998; Chin et al., 2001; Wertz et al., 2012; Rochman et al., 2018; Tegtmeier et al., 2018).

Colonial and cultural characteristics

On R2A agar media, visible colonies of *Geminisphaera colitermitum* appear within 2 weeks of incubation under normal (air) or hypoxic (2% (v/v) O_2 with 5% (v/v) CO_2) conditions (Wertz et al., 2012). Colonies are 2-4 mm in diameter, have an entire margin and a low convex, mucoid morphology, and are cream colored (Wertz et al., 2012).

Nutrition and growth conditions

Media that support the growth of *Geminisphaera colitermitum* currently include R2A agar or R2B broth (Wertz et al., 2012), fastidious anaerobe agar, and AM-5, a bicarbonate-buffered medium supplemented with yeast extract and casamino acids (0.1% each) (Tegtmeier et al., 2018). Diazotrophic growth also occurs in a defined, nitrogen-free media under a hypoxic or anoxic (80/20 N_2/CO_2) headspace (Wertz et al., 2012, Tegtmeier et al., 2018). Cells are facultatively anaerobic and microaerophilic with both fermentative and respiratory metabolisms, respectively (Tegtmeier et al., 2018). Cells prefer low oxygen conditions for growth with the shortest generation times and highest cell yields occurring under 2 to 8% O_2 in the headspace (Wertz et al., 2012, Tegtmeier et al., 2018). Under anoxic conditions, fermentation products produced from glucose (5 mM) include propionate, acetate, succinate, and CO_2 (Tegtmeier et al., 2018). Nitrate and sulfate are not reduced as terminal electron acceptors with glucose as the electron donor (Tegtmeier et al., 2018). Cells do not possess catalase, oxidase, or NADH/NADPH peroxidase activity (Wertz et al., 2012). Substrates utilized as energy sources include starch, D-cellobiose, D-maltose, D-glucose, D-galactose, and one or more components present in yeast extract (Wertz et al., 2012). Microcrystalline cellulose, methylcellulose, carboxymethylcellulose, xylan, D-fructose, D-mannose, D-trehalose, sucrose, D-ribose, D-xylose, L-arabinose, D-

mannitol, D-sorbitol, D-raffinose, D, L-lactate, sodium pyruvate, sodium fumarate, sodium acetate, allantoin, D-glucuronate, D-galacturonate, D-gluconic acid, xanthine, tannic acid, resourcinol, vanillic acid, sodium benzoate, and trimethyl benzoate are not used (Wertz et al., 2012). *Geminisphaera colitermitum* cells grow at a temperature range of 15 to 35°C, pH range of 5.5 to 7.5, and a NaCl concentration of 1.5% (w/v). No growth occurs at 4°C or 37°C and at pH ≤ 5 or ≥ 8 (Wertz et al., 2012).

Genome features

The current high-quality draft genome sequence of *Geminisphaera colitermitum* TAV2^T (whole genome shotgun GenBank accession no. ABEA00000000, version ABEA00000000.3; 5.67 Mb, 60.8 mol % GC content) contains 4,896 genes, 4,826 protein-coding sequences, and 1 rRNA operon (Rodrigues and Isanapong, 2014). The genome contains all genes necessary for glycolysis, the pentose phosphate pathway, and the tricarboxylic acid cycle. All of the genes necessary to express a cytochrome *cbb*$_3$ oxidase, but not a *bd* or *aa*$_3$-type terminal oxidase are present, suggesting that cytochrome *cbb*$_3$ oxidase is the sole oxidase capable of being expressed by *G. colitermitum* (Wertz et al., 2012). Genes for dinitrogen fixation, including complete operons for molybdenum-containing (*nifHDK*) and iron-only (*anfHDGK*) nitrogenase complexes are present in the genome (Wertz et al., 2012). Genes encoding a catalytic cellulase protein (glycohydrolase family 5), a β-1,4-xylanase (glycohydrolase family 16), and a starch phosphorylase (glycosyltransferase family 5) are also detected (Wertz et al., 2012).

Ecology

Along with several closely-related strains from the family *Opitutaceae* (Isanapong et al., 2012; Kotak et al., 2015, Kotak et al., 2020), *Geminisphaera colitermitum* TAV2^T was isolated from the hindgut of the wood-feeding Eastern subterranean termite, *Reticulitermes flavipes* (Kollar) (Stevenson et al., 2004). Members of the *Opitutaceae* are consistently found at

moderately low abundances in wood-feeding cockroaches (*Panesthia angustipennis*, 0.4%), lower termites (*Mastotermes darwiniensis*, 0.5%; *Reticulitermes flavipes*, 0.02%), and higher termites (*Alyscotermes trestus*, 0.4%; *Nasutitermes corniger*, 0.2%) determined by high-throughput amplicon sequencing (Köhler et al., 2012; Dietrich et al., 2014) or estimated by dilution-to-extinction PCR (Wertz et al., 2012). Despite these low abundances, members of the *Verrucomicrobia* appear to be autochthonous in the lower termite *R. flavipes*, suggesting that the niches occupied by these symbionts are important for the vitality of their host (Wertz et al., 2012). The capability for diazotrophic growth by *G. colitermitum* (Wertz et al., 2012, Tegtmeier et al., 2018) and the detection of increased nitrogenase, acetyl xylan esterase, and cytochrome bb_3 oxidase activities inferred by metatranscriptomics and proteomics under 2% O_2 compared to 20% O_2 substantiates the hypothesis that *G. colitermitum* cells contribute to *in situ* nitrogen fixation, hemicellulose degradation, and oxygen consumption in hypoxic microenvironments in the *R. flavipes* gut (Isanapong et al., 2013).

In addition to *G. colitermitum*, two other members of the *Opitutaceae* have been isolated from the intestinal tracts of insects, including two *Cephaloticoccus* species from *Cephalotes* ants and *Ereboglobus luteus* from *Shelfordella lateralis* cockroaches, respectively (Lin et al., 2016, Tegtmeier et al., 2018). The nitrogen acquisition capabilities of these symbionts are consistent with the nitrogen content of the diet of their respective hosts. The ability for diazotrophic growth by *G. colitermitum* suggests that these cells are among several important bacterial groups contributing to nitrogen fixation in wood-feeding termites (Brune, 2014), whereas the sugar-rich and nitrogen-poor diet of *Cephalotes* ants coincides with the presence of urea-degrading activity and potential amino acid provisioning by *Cephaloticoccus* symbionts (Lin et al., 2016; Hu et al., 2018). The absence of *nifH* genes and diazotrophic growth in *E. luteus* further confirms that nitrogen fixation is not an important process for omnivorous cockroaches (Breznak et al., 1973; Tegtmeier et al., 2018).

Maintenance Procedures

Geminisphaera colitermitum cells can be grown on solid R2A or liquid R2B media incubated under a normal atmosphere (air) or in a vinyl hypoxic chamber set to 2% (v/v) O_2 with 5% (v/v) CO_2 (Wertz et al., 2012, Isanapong et al., 2013). For liquid cultures, cells are grown in 50 mL Erlenmeyer flasks containing 15 mL R2B medium with shaking at 200 rpm at 25°C (Wertz et al., 2012). On solid media, cells can be maintained on plates for 2-3 weeks at 25°C before being transferred to fresh plates. Cells can also be grown under anoxic conditions in rubber-stoppered anaerobic culture tubes gassed with a headspace of 80/20 N_2/CO_2 in AM-5 medium at 30°C (Tegtmeier et al., 2018). Cells are stored by resuspending pelleted liquid cultures in R2B medium supplemented with glycerol (20% (v/v) final concentration) and freezing at -80°C.

Differentiation of the genus Geminisphaera from other genera

Table A-1 presents morphological, metabolic, and other physiological comparisons of all type species representing all genera currently within the *Opitutaceae*. To date, *G. colitermitum* TAV2^T is the only characterized isolate in the family *Opitutaceae* with nitrogen fixation genes and the capability for diazotrophic growth (**Table A-1**). Several closely-related *Opitutaceae* strains isolated from termite hindguts (Strains TAV1, 3, 4, and 5, **Figure A-2**) also possess genes for nitrogen fixation (Isanapong et al., 2012; Kotak et al., 2015; Kotak et al., 2020), but polyphasic taxonomic characterizations and proposal of formal names for these strains have not yet been performed. *G. colitermitum* can also be differentiated from other genera within the *Opitutaceae* on the basis of its *Diplococcus* cell shape, a characteristic currently shared only with *Oleiharenicola akalitolerans* (**Table A-1**). The sugar utilization profile of *G. colitermitum* is also distinct from other members of the *Opitutaceae* (**Table A-1**).

Taxonomic comments

The name for strain TAV2^T was originally proposed as '*Diplosphaera colitermitum*' (Wertz et al., 2012). This illegitimate genus name was emended to '*Didymococcus*' in 2017 (Wertz et al., 2017), before a subsequent correction to the name *Geminisphaera* in 2018 (Wertz et al., 2018a). Henceforth, *Geminisphaera colitermitum* is the validated name for strain TAV2^T in the official prokaryotic nomenclature (Wertz et al., 2018b).

Based on the 16S rRNA gene sequence, the closest cultivated neighbor of *G. colitermitum* with a validated name is *Oleiharenicola alkalitolerans* (92.6% nucleotide identity, **Figure A1**), followed by *Opitutus terrae* (92.4% nucleotide identity), *Cephaloticoccus primus*, and *Ereboglobus luteus* (91.5% nucleotide identity each).

List of species of the genus Geminisphaera

Geminisphaera colitermitum

Wertz et al., 2018b (effective publication: Wertz et al., 2018a)

(co.li.ter'mi.tum. L. neut. n. *colon*, colon, part of the large intestine; L. masc. n. *termes -itis*, wood-eating worm, termite; N.L. gen. pl. n. *colitermitum*, of the gut of termites).

Cells are coccoid (0.5 µm to 0.6 µm in diameter) and occur almost exclusively in pairs, with a Gram-negative cell wall morphology that includes an outer membrane. Cells are nonmotile. Facultatively anaerobic. Metabolism is both fermentative and respiratory. Fermentation products on 5 mM glucose are propionate, acetate, succinate and CO_2. Microaerophilic; grows under air; highest cell yields and shortest generation times are at 2-8% O_2. On solid R2A medium, colonies are 2 to 4 mm in diameter, have an entire margin and a low convex, mucoid morphology, and are cream colored. Cells do not possess catalase, oxidase, or NADH/NADPH peroxidase activity. Nitrogenase activity is inferred through growth on nitrogen-free medium. Growth occurs in liquid media between 15 and 35°C (optimum, 30°C), a pH range of 5.5 to 7.5 (optimum, 7.0), and a NaCl concentration of 1.5% (w/v). No growth occurs at 4°C

or 37°C and at pH ≤5 or ≥8. Substrates utilized as energy sources include starch, D-cellobiose, D-maltose, D-glucose, D-galactose, and one or more components present in yeast extract. Microcrystalline cellulose, methylcellulose, carboxymethylcellulose, xylan, D-fructose, D-mannose, D-trehalose, sucrose, D-ribose, D-xylose, L-arabinose, D-mannitol, D-sorbitol, D-raffinose, DL-lactate, sodium pyruvate, sodium fumarate, sodium acetate, allantoin, D-glucuronate, D-galacturonate, D-gluconic acid, xanthine, tannic acid, resourcinol, vanillic acid, sodium benzoate, and trimethylbenzoate are not utilized. The current high-quality draft genome sequence is 5.67 Mb in size (version ABEA00000000.3). The type strain, TAV2^{T}, was isolated from the hindgut of *Reticulitermes flavipes* (Kollar) termites collected from Dansville, MI, USA.

DNA G+C content (*mol %*): 60.8 (High-quality draft genome)

Type strain: TAV2^{T}, ATCC BAA-2264, DSM 25453, NRRL B-59605

EMBL/GenBank accession no. (16S rRNA gene): AY587232

EMBL/GenBank accession no. (genome sequence): ABEA00000000.3

REFERENCES

Breznak JA, Brill WJ, Mertins JW, & Coppel HC (1973) Nitrogen fixation in termites. *Nature* 244: 577-580.

Brune A (2014) Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol* 12: 168-180.

Chin KJ, Liesack W, & Janssen PH (2001) *Opitutus terrae* gen. nov., sp. nov., to accommodate novel strains of the division '*Verrucomicrobia*' isolated from rice paddy soil. *Int J Syst Evol Microbiol* 51: 1965-1968.

Dietrich C, Köhler T, & Brune A (2014) The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol* 80: 2261-2269.

Hu Y, Sanders JG, Łukasik P, D'Amelio CL, Millar JS, Vann DR et al. (2018) Herbivorous turtle ants obtain essential nutrients from a conserved nitrogen-recycling gut microbiome. *Nat Commun* 9: 964.

Isanapong J, Goodwin L, Bruce D, Chen A, Detter C, Han J et al. (2012) High-quality draft genome sequence of the *Opitutaceae* bacterium strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes*. *J Bacteriol* 194: 2744-2745.

Isanapong J, Sealy Hambright W, Willis AG, Boonmee A, Callister SJ, Burnum KE et al. (2013) Development of an ecophysiological model for *Diplosphaera colotermitum* TAV2, a termite hindgut Verrucomicrobium. *ISME J* 7: 1803-1813.

Köhler T, Dietrich C, Scheffrahn RH, & Brune A (2012) High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl Environ Microbiol* 78: 4691-4701.

Kotak M, Isanapong J, Goodwin L, Bruce D, Chen A, Han CS et al. (2015) Complete genome sequence of the *Opitutaceae* bacterium strain TAV5, a potential facultative methylotroph of the wood-feeding termite *Reticulitermes flavipes*. *Genome Announc* 3: e00060-15.

Kotak M, Lin JY, Isanapong J, & Rodrigues JLM (2020) Comparative genomics of strains TAV3 and TAV4 (*Verrucomicrobia*: *Opitutaceae*), isolated from a wood-feeding termite, and *in silico* analysis of their polysaccharide-degrading enzymes. Microbiol Resourc Announc 9 (2): e01192-19.

Lin JY, Russell JA, Sanders JG, & Wertz JT (2016) *Cephaloticoccus* gen. nov., a new genus of '*Verrucomicrobia*' containing two novel species isolated from *Cephalotes* ant guts. *Int J Syst Evol Microbiol* 66: 3034-3040.

Rast P, Glöckner I, Boedeker C, Jeske O, Wiegand S, Reinhardt R et al. (2017) Three novel species with peptidoglycan cell walls form the new genus *Lacunisphaera* gen. nov. in the family *Opitutaceae* of the Verrucomicrobial subdivision 4. *Front Microbiol* 8: 202.

Rochman FF, Kim JJ, Rijpstra WIC, Sinninghe Damsté JS, Schumann P, Verbeke T et al. (2018) *Oleiharenicola alkalitolerans* gen. nov., sp. nov., a new member of the phylum *Verrucomicrobia* isolated from an oilsands tailings pond. *Int J Syst Evol Microbiol* 68: 1078-1084.

Rodrigues JLM & Isanapong J (2014) The family *Opitutaceae*. In *The Prokaryotes*, 4th ed., Rosenberg E, Delong EF, Lory S, Stackenbrandt E and Thompson F (eds). Springer, Berlin; pp 751-756.

Shieh WY & Jean WD (1998) *Alterococcus agarolyticus*, gen. nov., sp. nov., a halophilic thermophilic bacterium capable of agar degradation. *Can J Microbiol* 44: 637-645.

Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, & Breznak JA (2004) New strategies for cultivation and detection of previously uncultured microbes. *Appl Environ Microbiol* 70: 4748-4755.

Tegtmeier D, Belitz A, Radek R, Heimerl T, & Brune A (2017) *Ereboglobus luteus* gen. nov. sp. nov. from cockroach guts, and new insights into the oxygen relationship of the genera *Opitutus* and *Didymococcus* (*Verrucomicrobia*: *Opitutaceae*). *Syst Appl Microbiol* 41: 101-112.

van Passel MWJ, Kant R, Palva A, Copeland A, Lucas S, Lapidus A et al. (2011) Genome sequence of the Verrucomicrobium *Opitutus terrae* PB90-1, an abundant inhabitant of rice paddy soil ecosystems. *J Bacteriol* 193: 2367-2368.

Wertz JT, Kim E, Breznak JA, Schmidt TM, & Rodrigues JLM (2012) Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. *Appl Environ Microbiol* 78: 1544-1555.

Wertz JT, Kim E, Breznak JA, Schmidt TM, & Rodrigues JLM (2017) Correction for Wertz et al., "Genomic and physiological characterization of the *Verrucomicrobia* isolate *Didymococcus colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes.". *Appl Environ Microbiol* 83: e00987-17.

Wertz JT, Kim E, Breznak JA, Schmidt TM, & Rodrigues JLM (2018a) Second correction for Wertz et al., "Genomic and physiological characterization of the *Verrucomicrobia* isolate *Geminisphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes.". *Appl Environ Microbiol* 84: e00952-18.

Wertz JT, Kim E, Breznak JA, Schmidt TM & Rodrigues JLM (2018b) In Validation list no. 183. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 68: 2707-2709.

TABLES AND FIGURES

Table A-1. Differential characteristics between *Geminisphaera colitermitum* TAV2^T and other cultivated type species from the family *Opitutaceae*

Characteristic	*G. colitermitum*	*C. primus*	*E. luteus*	*O. terrae*	*"L. limnophila"*	*A. agarolyticus*	*O. alkalitolerans*
Isolation source	Termite hindgut	Ant gut	Cockroach hindgut	Rice paddy soil	Freshwater lake	Coastal hot spring	Oilsands tailings pond
Cell shape	Diplococcus	Coccus	Coccus	Coccus	Coccus	Coccus	Coccus/Diplococcus
Cell diameter (µm)	0.5–0.6[a]	0.5–0.6	0.5–0.7	0.4–0.6	0.7–1.1	0.8–0.9	0.5–1.5
Gram stain	Negative	Negative	Negative	Negative	Negative	Negative	Negative
Colony color	Cream	Cream	Yellow	White	Cream	White opaque	White
Metabolism	Facultative anaerobe[b]	Obligate aerobe	Facultative anaerobe	Facultative anaerobe[b]	Aerobe	Facultative anaerobe	Obligate aerobe
Genome size (Mb)	5.67[c]	2.36	4.17	5.96[d]	4.20	ND	4.72
DNA G+C content (mol%)	60.8[a]	60.7	59.7	65.3[d]	66.5	65.5–67.0	66.1
Motility	–	–	+	+	+	+	–
Catalase	–	–	–	–	–	+	+
Oxidase	–	–	–	–	+	+	–
Nitrogen fixation/*nif* genes	+	–	–	–	–	ND	–
Diazotrophy	+	ND	–	–	ND	ND	ND
Temperature range (°C)	15–35	23–37	15–37	10–37	13–36	40–56	15–40
pH range	5.5–7.5	6.9–7.7	6.0–9.0	5.5–9.0	6.0–9.0	7.0–8.5	5.5–11.0
NaCl tolerance (%, w/v)	1.5	0.5–1.5	0–2.0	3.0	ND	1.0–3.0	3.0
Cellobiose	+	+	+	+	+	+	–
Maltose	+	+	+	+	+	ND	–
D-Lactose	ND	+	ND	+	+	+	–
Melibiose	ND	+	+	+	+	–	–
Sucrose	–	+	+	+	+	+	–
D-Fructose	–	+	+	+	+	ND	+
D-Glucose	+	+	+	+	+	+	+
Galactose	+	–	+	+	+	+	+
Starch	+	ND	–	+	ND	ND	–
D-Xylose	–	ND	–	–	ND	+	+

Data from Wertz et al. (2012) (*G. colitermitum*), Lin et al. (2016) (*C. primus*), Tegtmeier et al. (2018) (*E. luteus*), Chin et al. (2001) (*O. terrae*), Rast et al. (2017) ('*L. limnophila*'), and Rochman et al. (2018) (*O. alkalitolerans*). ND, not determined.

[a] Rodrigues and Isanapong (2014)
[b] Tegtmeier et al. (2018)
[c] Current high-quality draft genome sequence (version ABEA00000000.3)
[d] van Passel et al. (2011)

Figure A-1. Scanning electron micrograph of *Geminisphaera colitermitum* TAV2^T. Scale bar, 2 µm.

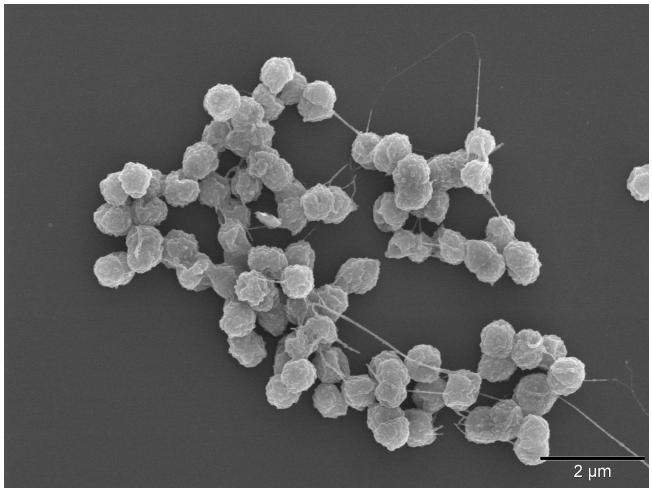
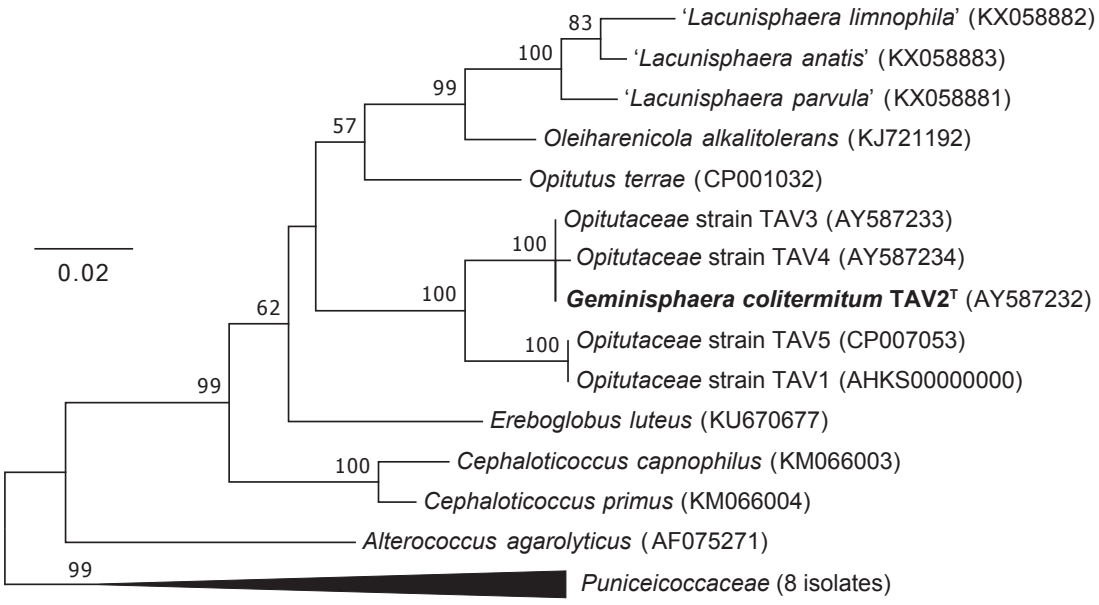


Figure A-2. Maximum likelihood-based 16S rRNA gene phylogeny of *Geminisphaera colitermitum* TAV2^T compared to other members from the family *Opitutaceae*. Members of the *Puniceicoccaceae*, a sister family within the *Opitutae*, were used as the outgroup. The phylogeny is based on 1,338 shared nucleotide positions and numbers at branch nodes indicate bootstrap support (1,000 replicates) above 50%. GenBank accession numbers are shown in parentheses. Bar, 0.02 substitutions per nucleotide.



APPENDIX B

TREPONEMA ENDOSYMBIONTS ARE THE DOMINANT BACTERIAL MEMBERS WITH UREOLYTIC POTENTIAL IN THE GUT OF THE WOOD-FEEDING TERMITE, *RETICULITERMES HESPERUS*

List of Authors: Jonathan Y. Lin, Laibin Huang, Sung J. Won, and Jorge L.M. Rodrigues

In revision at *Archives of Microbiology*

ABSTRACT

Termites are remarkable for their ability to digest wood as their main energy source, but the extremely low nitrogen (N) content of their diet presents a major challenge for N acquisition. Besides the N_2-fixing bacteria in the gut, the symbiotic groups that recycle N from waste products as a complementary N-provisioning mechanism in termites remains poorly understood. In this study, we used a combination of high-throughput amplicon sequencing, quantitative PCR, and cultivation to characterize the microbial community capable of degrading urea, a common waste product, into ammonia in the guts of two colonies of *Reticulitermes hesperus* termites. The abundance of the *ureC* gene, which encodes for the alpha subunit of the urease enzyme, ranged from 0.43 to 2.93% of the total prokaryotic community. Taxonomic analysis indicated that 27.6% of the *ureC* gene amplicons in the termite gut matched with a *Treponema* endosymbiont of gut protists previously found in several other termites, suggesting an important contribution to the nutrition of essential cellulolytic protists. This corroborated our cultivation efforts, where 92.2% the isolates recovered had ureolytic potential and matched the other taxa captured in our *ureC* gene sequences. Together, our results underscore a more important role for ureolysis by endosymbionts within protists than by free-swimming bacteria in the gut lumen of *R. hesperus*.

MAIN TEXT

Termites are social insects descended from wood-feeding cockroaches (Brune and Dietrich 2015) and have long been studied for their ability to digest lignocellulose (Brune 2014; Wertz and Béchade 2020). This ability is driven by essential contributions from deeply evolved, mutualistic symbionts found in their hindguts consisting of archaea and bacteria in the 'higher termites' (Family *Termitidae*) and a tripartite community of free-swimming archaea, bacteria, and cellulolytic protists with their ecto- and endosymbionts in the evolutionarily basal 'lower termites' (Brune 2014; Wertz and Béchade 2020). While biological nitrogen N_2 fixation (BNF) by bacterial symbionts is a prominent route of nitrogen (N) acquisition in termites to compensate for their N-limited diet (Brune 2014; Brune and Dietrich 2015; Wertz and Béchade 2020), BNF rates have been reported to be highly variable, suggesting that some species with low BNF rates such as the *Reticulitermes* must rely on different pathways to satisfy all host nutrient requirements (Breznak 2000). Termites release most nitrogenous waste as uric acid, and a previous study showed that gut bacteria in wood-feeding *Reticulitermes flavipes* termites are capable of recycling N in uric acid for re-absorption into host tissue (Potrikus and Breznak 1981). This process was confirmed in several bacterial strains isolated from termites that can ferment uric acid to produce ammonia (Potrikus and Breznak 1980; Thong-On et al. 2012). However, uric acid can be converted into urea by enzymes produced by the host or symbionts (Hansen et al. 2020); urea can also be excreted as a waste product by protist cells (Weatherby 1929). At this step, whether there are symbionts that can produce urease enzymes to catalyze the breakdown of urea to ammonia for re-assimilation in the termite gut remains unknown. This represents a knowledge gap in symbiont-mediated processes that may affect termite nutrition. To fill this gap, we characterized the taxonomic diversity and abundance of symbionts with ureolytic potential in the hindguts of two colonies of termites.

Termites were collected from two sites in the Central Valley in California, USA (**Figure B-1a**). Sequencing of the mitochondrial cytochrome oxidase II (*COII*) gene identified them as two genetically distinct colonies of *Reticulitermes hesperus* (**Figure B-1b**). DNA was extracted from hindguts dissected from workers and the *ureC* gene, which encodes for the alpha subunit of the urease enzyme, was sequenced using an Illumina MiSeq (Supplementary methods). The resulting sequences were classified using the GraftM pipeline against a *ureC* gene reference package (Boyd et al. 2018). Quantitative PCR (qPCR) was performed to determine the abundance of *ureC* genes in the termite gut, and bacteria were cultivated from hindgut homogenates and screened to confirm ureolytic potential. Across all samples, the composition of *ureC* gene at the phylum level consisted of the *Proteobacteria* (36.7%), *Spirochaetes* (27.6%), and *Firmicutes* (21.7%), with the remaining sequences not classified to any phyla (12.0%). The most abundant *ureC* sequences classified at the genus level were the *Treponema* (27.6%), followed by Candidatus *Accumulibacter* (7.3%), *Pseudomonas* (4.2%), *Bacillus* (3.0%), *Acinetobacter* (1.8%), and *Desulfovibrio* (1.1%) (**Figure B-1c**).

Notably, all *ureC* gene sequences annotated as *Treponema* were mapped to a single phylotype, Urec_98, which was the most abundant classified *ureC* sequence for both colonies. Phylogenetic analysis placed Urec_98 in a clade with a species previously identified as "*Candidatus* Treponema intracellularis," an endosymbiont of *Eucomonympha* protists in the termite gut (Ohkuma et al. 2015) (**Figure B-2**). Urec_98 shared a 96.61% similarity in the protein coding sequence with the *ureC* gene from these endosymbionts and was genetically divergent from the sequences of the free-living *Treponema bryantii* and *T. ruminis* (71.97% identity for both), indicating that Urec_98 is evolutionarily distant from other ureolytic *Treponema* species and likely an endosymbiont of protists in the termite gut. The abundance of *ureC* gene ranged from 131 to 2,171 copies ng DNA^{-1}, which was 0.43 to 2.93% of the total prokaryotic community when calculated as a proportion relative to the 16S rRNA gene copies (**Table B-1**). Out of a total

of 192 isolates retrieved from our cultivation procedures, 92.2% possessed ureolytic potential based on diagnostic PCR of the *ureC* gene (**Table SB-1**). Most of the identified strains were from the phyla *Proteobacteria* (148 isolates), *Bacteroidetes* (11 isolates) and *Firmicutes* (6 isolates) (**Table SB-1**). At the genus level, the strains were identified as *Citrobacter* (88 isolates), *Pseudomonas* (28 isolates), *Acinetobacter* (27 isolates), *Chryseobacterium* (11 isolates), and *Bacillus* (5 isolates), representing most of the diversity other than Urec_98 captured from our sequencing (Table S1). Taken together, our results demonstrate that the *ureC* gene is 1) present in the termite gut, are 2) prevalent in a majority of cultivated bacterial isolates, and 3) a significant proportion of the gene sequences are represented by endosymbionts.

Urec_98 was the single most abundant phylotype from our *ureC* gene dataset and is closely related to "*Candidatus* Treponema intracellularis," an endosymbiont of *Eucomonympha* protists previously found in wood-feeding *Hodotermopsis sjoestedti* termites (Ohkuma et al. 2015). The "*Candidatus* T. intracellularis" genome contains genes encoding for urease as well as a membrane-bound urea channel, indicating its ability to both transport and use urea excreted by its host (Ohkuma et al. 2015). "*Candidatus* T. intracellularis" falls within the termite *Treponema* cluster II (Ohkuma et al. 2015), a defined clade of *Treponem*a ectosymbionts attached to the cell surface of termite gut protists (Ohkuma et al. 1999; Noda et al. 2003). This clade, along with a group of free-swimming *Treponema* (cluster I) comprise an abundant and highly co-evolved community of *Spirochaetes* within the termite gut (Brune 2006). To date, the only other members of the *Treponema* for which genomes are available on NCBI that possess genes for urease and urea transporters are *T. bryantii* and *T. ruminis*, two *Spirochaetes* originally isolated from the bovine rumen (Stanton and Canale-Parola 1980; Newbrook et al. 2017), an environment where urea from the bloodstream is a major source of waste N in ruminant animals (Mobley and Hausinger 1989). By comparison, the genomes of *T. primitia* and *T. azonutricium* (Graber and Breznak 2005), two free-swimming species from *Treponema* cluster I isolated from the termite

gut (Leadbetter et al. 1999), do not contain genes encoding for any urease subunits or urea transporters. This suggests that unlike the rumen environment, the termite gut lumen likely does not have a significant flux of urea which may underscore a lack of selective pressure for free-swimming *Spirochaetes* to possess urease genes.

Besides the *Treponema* endosymbionts, several extracellular bacteria previously isolated from termite guts have been shown to have urease enzyme activity, such as *Comamonas odontotermitis* (Chou et al. 2007); or encode operons for urease and their transporters in their genomes including a *Citrobacter* strain (Fontes-Perez et al. 2015), *Sporomusa termitida* (Breznak et al. 1988), *Stenoxybacter acetivorans* (Wertz and Breznak 2007), and several *Verrucomicrobia* strains (Wertz et al. 2012; Isanapong et al. 2012; Kotak et al. 2015, 2020; Lin and Rodrigues 2020). Yet, the fact that we did not detect any of these taxa at proportions greater than 0.5% in our *ureC* gene dataset suggests that ureolysis in the gut lumen by free-swimming bacteria likely does not produce a significant quantity of recycled N for *Reticulitermes* termites. By contrast, our detection of *Treponema* endosymbionts at much higher proportions from our *ureC* sequences suggests a more important role for urea recycling inside protists. This is substantiated by a previous finding of another termite endosymbiont, "*Candidatus* Azobacteroides pseudotrichonymphae," a *Bacteroidales* strain that, like "*Candidatus* T. intracellularis," also possesses a gene cluster encoding a urease and urea transporter (Hongoh et al. 2008). This endosymbiont was found to be abundant in *Pseudotrichonympha* protists, a sister lineage to the *Eucomonympha* protists, which suggests that phylogenetically distant bacteria may have convergently established similar functional niches for N recycling within protist hosts (Ohkuma et al. 2015). These contributions are expected to enable the protists to grow efficiently and remain stable during nutrient fluctuations in the gut, thereby allowing the termite to maintain protists essential for cellulose degradation and host nutrition (Ohkuma et al. 2015).

In summary, our characterization of the *ureC* gene in termite guts showed a significant number of sequences matching to a *Treponema* endosymbiont of protists previously found in another termite, highlighting a more important role for ureolysis by endosymbionts within protists than by free-swimming bacteria in the gut lumen of *R. hesperus*. Thus, ureolytic endosymbionts are likely important for maintaining the stability of essential cellulolytic protists within the tripartite microbial community in the guts of the lower termites. Future work on ureolytic gut microbes is needed to determine the distribution and ecological relevance of *Treponema* endosymbionts in other termites.

Data Availability

The *ureC* gene sequences were deposited to the NCBI sequence read archive (SRA) under BioProject PRJNA660442 with the BioSample accession number SAMN15949813. The custom GraftM gene package used for taxonomic classification of *ureC* gene sequences is available at https://github.com/jonathanylin/Termite_gut_urease.

ACKNOWLEDGEMENTS

We thank the members of the Soil Ecogenomics lab, Eloi Parladé, Albert Barberán, and Michael Messina for assistance with sampling. We acknowledge Yang Ouyang for uploading *ureC* seed sequences to the Fungene Repository. This research was supported in part by a UC Davis Henry A. Jastro Research Award (to J.Y.L.) and a UC Davis Provost's Undergraduate Fellowship (to S.J.W.).

REFERENCES

Boyd JA, Woodcroft BJ, Tyson GW (2018) GraftM: A tool for scalable, phylogenetically informed classification of genes within metagenomes. Nucleic Acids Res 46:e59–e59.

Breznak JA (2000) Ecology of prokaryotic microbes in the guts of wood- and litter-feeding termites. In: Abe T, Bignell DE, Higashi M (eds) Termites: Evolution, Sociality, Symbioses, Ecology. Springer Netherlands, Dordrecht, pp 209–231

Breznak JA, Switzer JM, Seitz H-J (1988) *Sporomusa termitida* sp. nov., an H_2/CO2-utilizing acetogen isolated from termites. Arch Microbiol 150:282–288.

Brune A (2014) Symbiotic digestion of lignocellulose in termite guts. Nat Rev Micro 12:168–180.

Brune A (2006) Symbiotic associations between termites and prokaryotes. In: Dr MDP, Falkow S, Rosenberg E, et al. (eds) The Prokaryotes. Springer New York, pp 439–474

Brune A, Dietrich C (2015) The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution. Annual Rev Microbiol 69:145–166.

Chou J-H, Sheu S-Y, Lin K-Y, et al (2007) *Comamonas odontotermiti*s sp. nov., isolated from the gut of the termite *Odontotermes formosanus*. Int J Syst Evol Microbiol 57:887–891.

Fontes-Perez H, Olvera-García M, Chávez-Martínez A, et al (2015) Genome sequence of *Citrobacter* sp. CtB7.12, isolated from the gut of the desert subterranean termite *Heterotermes aureus*. Genome Announc 3.

Graber JR, Breznak JA (2005) Folate cross-feeding supports symbiotic homoacetogenic *Spirochetes*. Appl Environ Microbiol 71:1883–1889.

Hansen AK, Pers D, Russell JA (2020) Symbiotic solutions to nitrogen limitation and amino acid imbalance in insect diets. In: Oliver KM, Russell JA (eds) Advances in Insect Physiology. Academic Press, pp 161–205

Hongoh Y, Sharma VK, Prakash T, et al (2008) Genome of an endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. Science 322:1108–1109.

Isanapong J, Goodwin L, Bruce D, et al (2012) High-quality draft genome sequence of the *Opitutaceae* bacterium strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes*. J Bacteriol 194:2744–2745.

Kotak M, Isanapong J, Goodwin L, et al (2015) Complete genome sequence of the *Opitutaceae* bacterium strain TAV5, a potential facultative methylotroph of the wood-feeding termite *Reticulitermes flavipes*. Genome Announc 3.

Kotak M, Lin JY, Isanapong J, Rodrigues JLM (2020) Draft genome sequences of strains TAV3 and TAV4 (*Verrucomicrobia*: *Opitutaceae*), isolated from a wood-feeding termite, and in silico analysis of their polysaccharide-degrading enzymes. Microbiol Resour Announc 9.

Leadbetter JR, Schmidt TM, Graber JR, Breznak JA (1999) Acetogenesis from H_2 plus CO_2 by *Spirochetes* from termite guts. Science 283:686–689.

Lin JY, Rodrigues JLM (2020) Geminisphaera. In: Bergey's Manual of Systematics of Archaea and Bacteria. Wiley, pp 1–7

Mobley HL, Hausinger RP (1989) Microbial ureases: Significance, regulation, and molecular characterization. Microbiol Rev 53:85–108

Newbrook K, Staton GJ, Clegg SR, et al (2017) *Treponema ruminis* sp. nov., a spirochaete isolated from the bovine rumen. Int J Syst Evol Microbiol 67:1349–1354.

Noda S, Ohkuma M, Yamada A, et al (2003) Phylogenetic position and *in situ* Identification of ectosymbiotic *Spirochetes* on protists in the termite gut. Appl Environ Microbiol 69:625–633.

Ohkuma M, Iida T, Kudo T (1999) Phylogenetic relationships of symbiotic spirochetes in the gut of diverse termites. FEMS Microbiol Lett 181:123–129.

Ohkuma M, Noda S, Hattori S, et al (2015) Acetogenesis from H_2 plus CO_2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. Proc Natl Acad Sci USA 112:10224

Potrikus CJ, Breznak JA (1980) Anaerobic degradation of uric acid by gut bacteria of termites. Appl Environ Microbiol 40:125–132

Potrikus CJ, Breznak JA (1981) Gut bacteria recycle uric acid nitrogen in termites: A strategy for nutrient conservation. Proc Natl Acad Sci USA 78:4601–4605

Stanton TB, Canale-Parola E (1980) *Treponema bryantii* sp. nov., a rumen spirochete that interacts with cellulolytic bacteria. Arch Microbiol 127:145–156.

Thong-On A, Suzuki K, Noda S, et al (2012) Isolation and characterization of anaerobic bacteria for symbiotic recycling of uric acid nitrogen in the gut of various termites. Microbes Environ 27:186–192

Weatherby JH (1929) Excretion of nitrogenous substances in *Protozoa*. Physiol Zool 2:375–394.

Wertz JT, Béchade B (2020) Symbiont-mediated degradation of dietary carbon sources in social herbivorous insects. In: Oliver KM, Russell JA (eds) Advances in Insect Physiology. Academic Press, pp 63–109

Wertz JT, Breznak JA (2007) *Stenoxybacter acetivorans* gen. nov., sp. nov., an acetate-oxidizing obligate microaerophile among diverse O2-consuming bacteria from termite guts. Appl Environ Microbiol 73:6819–6828.

Wertz JT, Kim E, Breznak JA, et al (2012) Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. Appl Environ Microbiol 78:1544–1555.

TABLES AND FIGURES

Table B-1. Abundance and proportion of *ureC* genes in termite guts.

Sample	Colony	*ureC* copies ng DNA^{-1}	16S rRNA copies ng DNA^{-1}	*ureC* | 16S copies	*ureC* relative abundance (%)
LT1	TH1	310	40678	0.0076	0.76
LT2	TH1	371	59087	0.0063	0.63
LT3	TH1	648	67205	0.0096	0.96
LT4	TH1	910	103702	0.0088	0.88
LT5	TH1	178	41056	0.0043	0.43
LT6	TH1	437	64671	0.0068	0.68
LT7	TH1	851	78132	0.0109	1.09
LT8	TH1	439	32391	0.0136	1.36
LT9	TH1	361	54237	0.0067	0.67
LT10	TH1	413	67608	0.0061	0.61
LT11	TH1	131	30121	0.0044	0.44
LT12	TH1	653	87496	0.0075	0.75
WT1	WTH1	1872	127482	0.0147	1.47
WT2	WTH1	833	54301	0.0153	1.53
WT3	WTH1	502	43315	0.0116	1.16
WT4	WTH1	1033	42190	0.0245	2.45
WT5	WTH1	1153	39813	0.0290	2.90
WT6	WTH1	2171	84579	0.0257	2.57
WT7	WTH1	995	43408	0.0229	2.29
WT8	WTH1	724	50891	0.0142	1.42
WT9	WTH1	1413	56497	0.0250	2.50
WT10	WTH1	1486	76989	0.0193	1.93
WT11	WTH1	1314	44906	0.0293	2.93

Figure B-1. a) Map showing the collection sites for the termite colonies used in this study. b) Phylogenetic tree (maximum-likelihood) based on termite mitochondrial cytochrome oxidase II (COII) gene sequences showing the identification of the termite colonies. Colors correspond with sampling locations, numbers at branch nodes indicate bootstrap support (500 replicates) above 50%, and the bar shows 0.1 substitutions per position. **c)** Alluvial plot showing the distribution of classified gut ureolytic microbial communities based on the *ureC* gene in all termites across different taxonomic ranks. The relative abundance (%) for the taxa at the genus level are shown in parentheses.

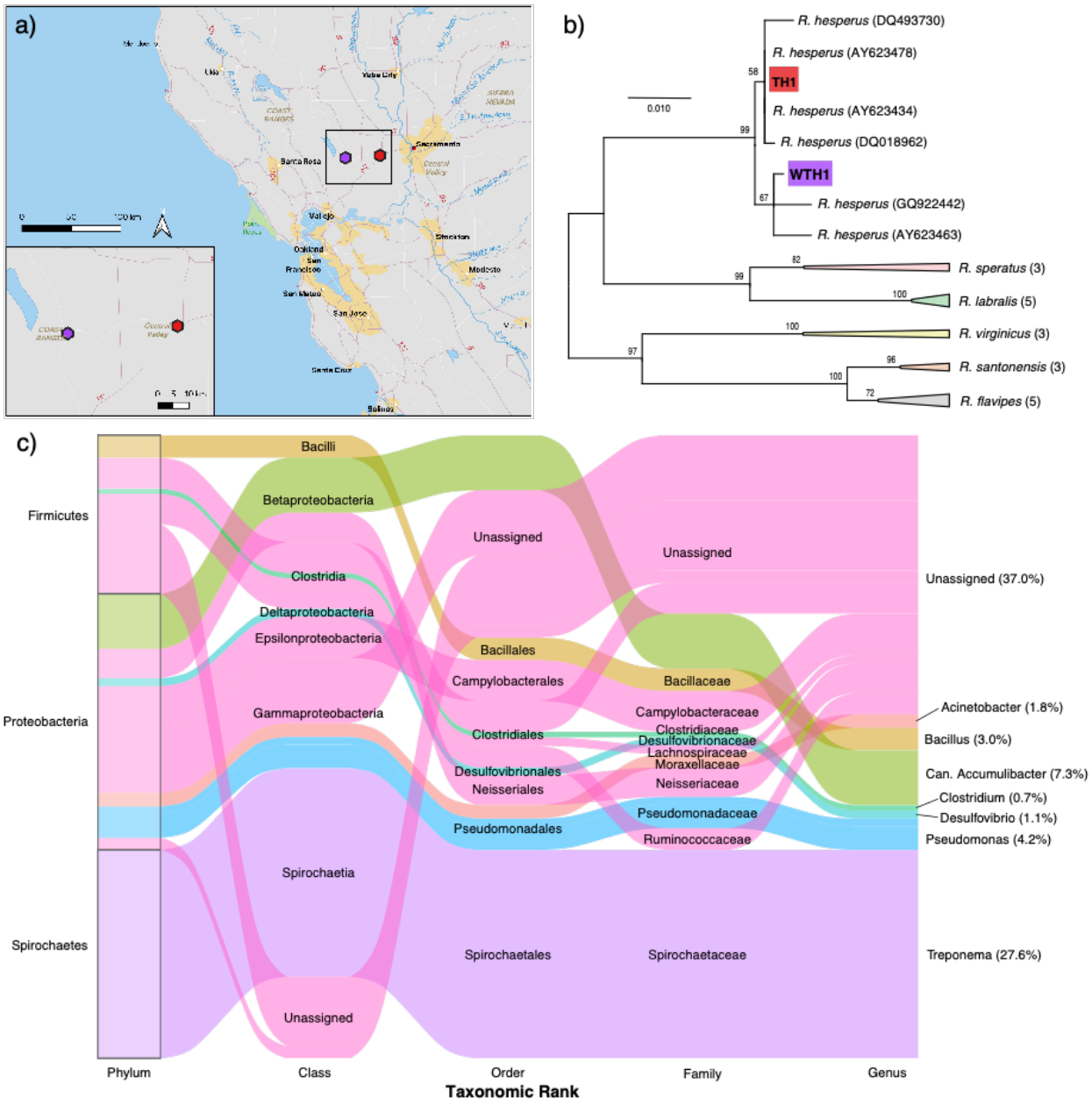
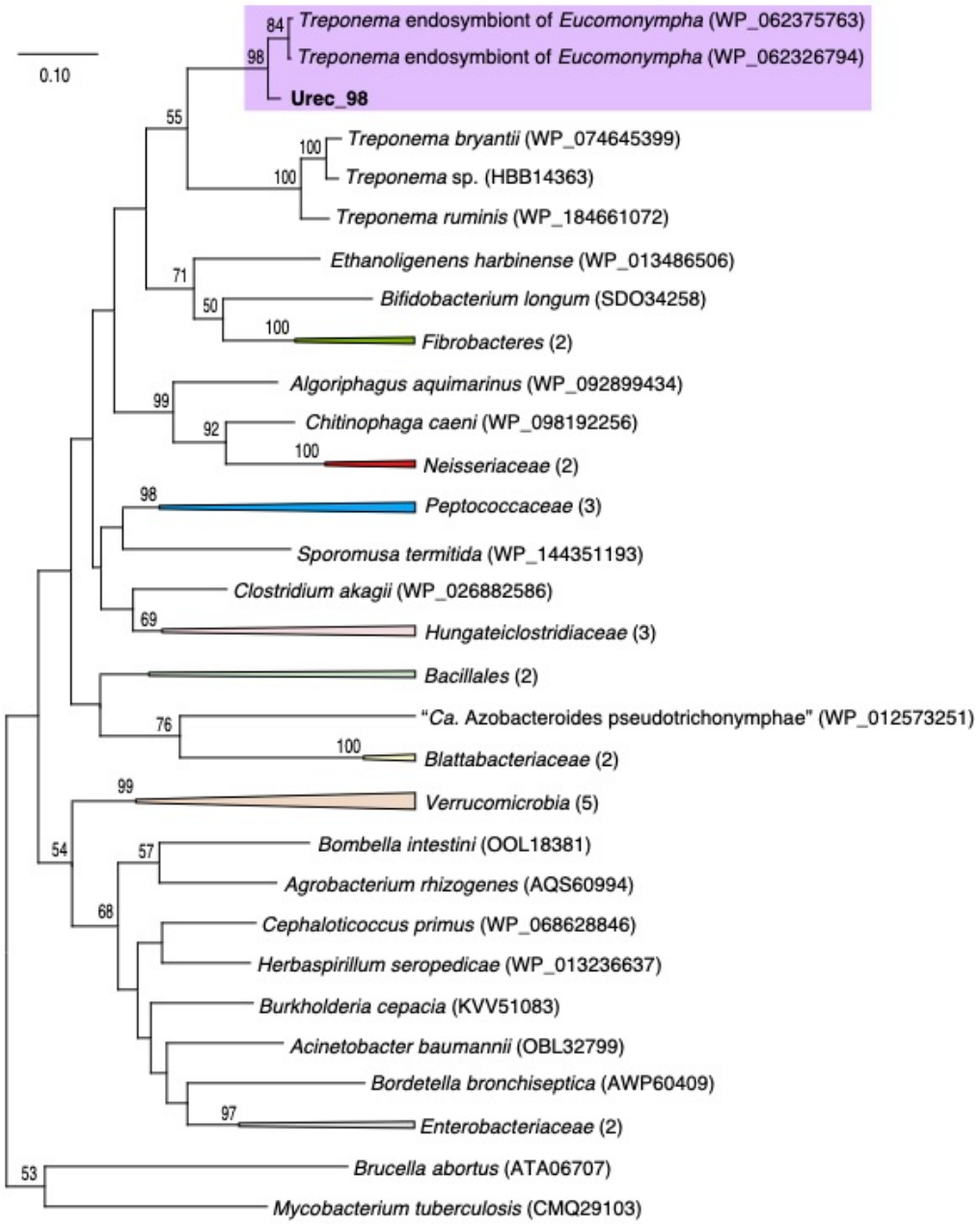


Figure B-2. Phylogenetic characterization of Urec_98 based on protein-coding sequences of the *ureC* gene. Sequences from other *Treponema* species, bacteria previously isolated from termite guts, and other strains spanning major bacterial phyla for which data are available were included as reference species. The tree was constructed by the maximum-likelihood method using the Jones-Thorton-Taylor model. Numbers at branch nodes indicate bootstrap support (500 replicates) above 50%, sequence accession numbers are given in parentheses, and the bar shows 0.1 substitutions per position.



APPENDIX B

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY MATERIALS AND METHODS

Termite collection

Termites were collected from two sampling locations. The first colony (TH1) was collected from the University of California Davis (UC Davis) Putah Creek Riparian Reserve (38.524° N 121.783° W) in March 2017 and reared in the laboratory as described previously (Odelson and Breznak 1983). The second colony (WTH1) was collected at the UC Davis Stebbins Cold Canyon Reserve (38.507° N 122.097° W) in November 2017 and held for less than 36 hours before degutting. Only termites from the worker caste were used for experiments.

Gut dissection, library preparation, and amplicon sequencing

Hindguts (11-12 per group) were removed by first surface-washing each worker in sterile urea isolation broth (UIB, per liter: 5 g of NaCl, 2 g of peptone, 9.5 g of K_2HPO_4, 9.1 g of KH_2PO_4, 1 mL of 1000x trace elements solution (Table S2), and 10 g of urea added aseptically as a solution after cooling) before pulling the thorax and anus apart using sterile forceps (Matson et al. 2007). For DNA extraction, single, whole hindguts were placed into bead-beating tubes containing Powerbead solution and solution C1 from the DNeasy Powersoil DNA extraction kit (Qiagen, Germantown, MD, USA) and stored at -20°C until extraction. DNA was extracted using a vortex adaptor according to the manufacturer's instructions and was quantified using the Qubit dsDNA HS assay kit (Life Technologies, Carlsbad, CA, USA). The *ureC* gene, which encodes for the catalytic subunit of urease containing several conserved regions, was amplified from each sample in triplicate using modified versions of the primers UreC-F and UreC-R (Reed 2001; Jin et al. 2017). A unique 12 bp barcode sequence was added to the reverse primer for each sample (Table S3). Amplifications were carried out in 20 µL volume reactions containing 10 µL Phusion

Hot-Start II High-Fidelity Master Mix (ThermoFisher, Waltham, MA, USA), 0.5 μM each primer, 10 ng sample DNA, and 4 μL sterile ddH_2O. The resulting amplicons were inspected by gel electrophoresis, pooled in equimolar concentrations, and sequenced (PE 250) on a MiSeq platform (Illumina, San Diego, CA, USA) at the UC Davis DNA Technologies core facility using standard sequencing primers following ligation of Illumina sequencing adaptors.

Termite identification

Termite heads were removed from workers, pooled (15-20 heads per extraction), and DNA was extracted as described above for amplification and sequencing of the mitochondrial cytochrome oxidase II (*COII*) gene using the primers A-tLEU and B-tLYS (Benjamino and Graf 2016). PCR products were purified using the Ultraclean PCR clean up kit (MO-BIO Laboratories, Carlsbad, CA, USA) and sequenced at the UC Davis College of Biological Sciences DNA sequencing facility (Davis, CA, USA) using an ABI 3730 platform (Applied Biosystems, Foster City, CA, USA) to identify the termites.

Quantitative PCR

To determine the abundance of the *ureC* gene in termite guts, quantitative PCR (qPCR) was performed on each sample using non-barcoded versions of the primers UreC-F/UreC-R for the *ureC* gene (Reed 2001) and 515F/806R for the 16S rRNA gene (Rubin et al. 2014). Amplifications for each target gene were performed on a Bio-Rad CFX Connect System (Bio-Rad Laboratories, Hercules, CA, USA) in 20 μL reaction mixtures containing 10 μL SsoAdvanced Universal SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA), 0.5 μM each primer, 10 ng template DNA, and 4 μL sterile ddH_2O. Amplification of the *ureC* gene consisted of an initial denaturation of 95°C for 3 min, followed by 39 cycles of 95°C for 10 s and 52°C for 30 s. Amplification of the 16S rRNA gene consisted of an initial denaturation of 95°C for 3 min, followed by 39 cycles of 95°C for 10 s and 60°C for 30 s. Standard curves (with a detection range

of 10^1-10^9 copies) were generated with the pCR Blunt II-TOPO vector (Invitrogen, Carlsbad, CA, USA) containing PCR-amplified fragments of each target. Coefficient of determination (R^2) values and amplification efficiency percentages for the standard curves were 0.965 and 91.3% for *ureC* gene and 0.993 and 108.8% for the 16S rRNA gene, respectively. Triplicate reactions were performed for each gene per sample and a melting curve analysis was performed after each assay to ensure specificity of the amplified products.

Sequence data processing

Raw reads for the *ureC* gene amplicon sequences were first quality checked using FastQC (Andrews 2010) before paired-end reads were merged with FLASH (Magoč and Salzberg 2011) using default parameters. Merged sequences were demultiplexed in QIIME v.1.9 (Caporaso et al. 2010) using the 'split_libraries_fastq.py' and 'split_sequence_file_on_sample_ids.py' scripts. Then, the forward and reverse primer sequences were trimmed from each file using BBDuk (Bushnell) before chimera detection and removal with the 'identify_chimeric_sequences.py' and 'filter_fasta.py' scripts in QIIME using USEARCH v.6.1. To identify the taxonomy of representative sequences, microbial *ureC* gene sequences were downloaded from the FunGene repository (Fish et al. 2013). In addition, *ureC* gene sequences from bacteria originating from the termite gut were downloaded from NCBI and both sequence datasets were compiled into a custom gene package using the '-create' command implemented in GraftM (Boyd et al. 2018) for protein sequence alignment, Hidden Markov Model (HMM) construction, and phylogenetic tree building. Taxonomy was assigned by using the '-graft' command against the compiled *ureC* gene package in GraftM, which places query sequences onto the *ureC* reference tree with pplacer (Matsen et al. 2010) using a default likelihood cut off value of 0.75.

Cultivation and screening procedures

To confirm whether *ureC* gene sequences obtained from termites originated from viable microbial symbionts potentially capable of degrading urea, we cultivated bacteria from hindgut samples and screened for the presence of the *ureC* gene. Termite hindguts (10-15) were removed from each colony and homogenized in sterile UIB using a pestle. The homogenate was then serially diluted and spread onto plates containing Urea Isolation Agar (UIA), which contains the same components as UIB but with the addition of agar (15 g per L). Since the termite hindgut is spatially stratified with respect to oxygen concentration (Brune et al. 1995), replicate extraction and plating procedures were performed under an atmospheric O_2 concentration (20.9% at 1 atm) and inside a 2% O_2 atmosphere-controlled glove box fitted with an oxygen sensor and automated controller (Coy Labs, Grass Lake, MI, USA) using a gas mixture of 5% CO_2 and 95% N_2 at room temperature for approximately 1 month. Individual colonies were streaked onto fresh UIA plates as they appeared and reinoculated onto plates three subsequent times to confirm isolation. Genomic DNA was extracted from cells by using a QIAamp DNA mini kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions and used for repetitive element palindromic (rep) PCR fingerprinting analysis to identify unique isolates (Dombek et al. 2000). The 16S rRNA gene was amplified from each unique isolate using the universal primers 63F and 1389R (Baker et al. 2003). PCR products were purified using ExoSAP-IT (USB Corporation, Cleveland, OH, USA) and sequenced at the UC Berkeley DNA sequencing facility using BigDye terminator chemistry. The resulting sequences were quality-trimmed using 4Peaks v.1.8 (Nucleobytes, Amsterdam, NL) and BLAST searches in the NCBI database (Altschul et al. 1990) were performed to identify the closest matches. PCR using the UreC-F and UreC-R primers was performed on DNA from each isolate and visualized by gel electrophoresis to determine the presence of a *ureC* gene fragment. DNA from *Proteus vulgaris* strain ATCC 6380 (Microbiologics, St. Cloud, MN, USA) and *Escherichia coli* strain BL21 were used as the positive

and negative controls, respectively. All strains were archived by suspending cultures in UIB supplemented with 20% (v/v) glycerol and storing them in cryovials at -80°C.

SUPPLEMENTARY REFERENCES

Altschul SF, Gish W, Miller W, et al (1990) Basic local alignment search tool. J Mol Biol 215:403–410.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data.

Baker GC, Smith JJ, Cowan DA (2003) Review and re-analysis of domain-specific 16S primers. J Microbiol Methods 55:541–555

Benjamino J, Graf J (2016) Characterization of the core and caste-specific microbiota in the termite, *Reticulitermes flavipes*. Front Microbiol 7:171.

Boyd JA, Woodcroft BJ, Tyson GW (2018) GraftM: A tool for scalable, phylogenetically informed classification of genes within metagenomes. Nucleic Acids Res 46:e59–e59.

Brune A, Emerson D, Breznak JA (1995) The termite gut microflora as an oxygen sink: Microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl Environ Microbiol 61:2681–2687

Bushnell B BBMap. In: SourceForge. https://sourceforge.net/projects/bbmap/

Caporaso JG, Kuczynski J, Stombaugh J, et al (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7:335–336.

Dombek PE, Johnson LK, Zimmerley ST, Sadowsky MJ (2000) Use of repetitive DNA sequences and the PCR To differentiate *Escherichia coli* isolates from human and human and animal sources. Appl Environ Microbiol 66:2572–2577.

Fish JA, Chai B, Wang Q, et al (2013) FunGene: The functional gene pipeline and repository. Front Microbiol 4.

Jin D, Zhao S, Zheng N, et al (2017) Differences in ureolytic bacterial composition between the rumen digesta and rumen wall based on *ureC* gene classification. Front Microbiol 8.

Magoč T, Salzberg SL (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27:2957–2963.

Matsen FA, Kodner RB, Armbrust EV (2010) pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinform 11:538.

Matson E, Ottesen E, Leadbetter J (2007) Extracting DNA from the gut microbes of the termite (*Zootermopsis nevadensis*). J Vis Exp e195.

Odelson DA, Breznak JA (1983) Volatile fatty acid production by the hindgut microbiota of xylophagous termites. Appl Environ Microbiol 45:1602–1613

Reed KE (2001) Restriction enzyme mapping of bacterial urease genes: using degenerate primers to expand experimental outcomes. Biochem Mol Biol Edu 29:239–244.

Rubin BER, Sanders JG, Hampton-Marcell J, et al (2014) DNA extraction protocols cause differences in 16S rRNA amplicon sequencing efficiency but not in community profile composition or structure. Microbiologyopen 3:910–921.

SUPPLEMENTARY TABLES

Table SB-1. Identity, abundance, and ureolytic potential of isolates cultivated from termite guts. +, positive result; - negative result; NS, no sequence retrieved; NA, not available.

Strain ID	Colony	Condition	*ureC* Gene	No. Isolates	16S Sequence Length (bp)	NCBI BLAST Results				
						Closest Cultivated Match	Accession No.	Percent Identity (%)	Query Cover (%)	Isolation Source
WTUA-1	WTH1	Air	+	1	1091	*Pseudomonas* sp. P15 SM-2017	LC230077	99	100	Soil
WTUA-2	WTH1	Air	+	1	1085	*Pseudomonas* sp. S3Bt38y	MH463752	98	100	Soil
WTUA-3	WTH1	Air	+	4	1002	*Acinetobacter* sp. HX09	MH368121	100	100	Wastewater
WTUA-5	WTH1	Air	+	1	1004	*Pseudomonas* sp. YJJ-50	KJ093367	100	100	River water
WTUA-6	WTH1	Air	+	2	1028	*Pseudomonas baetica* Ghg51-2	MH201226	99	100	Soil
WTUA-7	WTH1	Air	+	1	1022	*Pseudomonas* sp. S3Bt38y	MH463752	98	100	Soil
WTUA-8	WTH1	Air	+	1	1004	*Pseudomonas putida* 7B1	MH379791	99	100	Plant
WTUA-9	WTH1	Air	+	1	1021	*Chryseobacterium* sp. PCH75	KY628893	100	100	NA
WTUA-10	WTH1	Air	+	2	1098	*Pseudomonas* sp. S3Bt38y	MH463752	98	100	Soil
WTUA-11	WTH1	Air	+	2	1021	*Chryseobacterium* sp. StRB294	AB581560	100	100	Plant
WTUA-12	WTH1	Air	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUA-13	WTH1	Air	+	13	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUA-14	WTH1	Air	+	4	994	*Pseudomonas putida* 7B1	MH379791	99	100	Plant
WTUA-17	WTH1	Air	+	2	1023	*Pseudomonas baetica* Ghg51-2	MH201226	99	100	Soil
WTUA-20	WTH1	Air	+	1	1023	*Chryseobacterium* sp. StRB294	AB581560	100	100	Soil
WTUA-21	WTH1	Air	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUA-22	WTH1	Air	+	2	1048	*Acinetobacter* sp. HX09	MH368121	99	100	Wastewater
WTUA-23	WTH1	Air	+	1	1018	*Pseudomonas* sp. P24	KY084465	99	100	Plant
WTUA-25	WTH1	Air	+	1	1003	*Pseudomonas* sp. P24	KY084465	99	100	Plant
WTUA-26	WTH1	Air	+	2	999	*Acinetobacter* sp. HX09	MH368121	99	100	Wastewater
WTUA-28	WTH1	Air	+	1	1004	*Pseudomonas putida* 7B1	MH379791	99	100	Plant
WTUA-31	WTH1	Air	+	1	1099	*Pseudomonas* sp. YJJ-50	KJ093367	99	100	River water
WTUA-36	WTH1	Air	+	1	1006	*Stenotrophomonas* sp. Mong-4	KY962735	99	100	Soil
WTUA-38	WTH1	Air	+	3	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUA-40	WTH1	Air	+	1	978	*Pseudomonas* sp. YJJ-50	KJ093367	100	100	River water
WTUA-41	WTH1	Air	+	1	1018	*Acinetobacter* sp. HX09	MH368121	99	100	Wastewater
WTUA-43	WTH1	Air	+	1	1021	*Acinetobacter* sp. HX09	MH368121	100	100	Wastewater
WTUA-44	WTH1	Air	+	3	1020	*Acinetobacter* sp. HX09	MH368121	100	100	Wastewater
WTUA-49	WTH1	Air	+	2	1023	*Pseudomonas* sp. P24	KY084465	99	100	Plant
WTUA-51	WTH1	Air	+	1	932	*Pseudomonas baetica* Ghg51-2	MH201226	99	100	Soil
WTUH-1	WTH1	Hypoxia	+	1	982	*Acinetobacter* sp. HX09	MH368121	99	100	Wastewater
WTUH-3	WTH1	Hypoxia	+	1	509	*Acinetobacter* sp. HX09	MH368121	100	100	Wastewater
WTUH-5	WTH1	Hypoxia	+	1	510	*Acinetobacter* sp. HX09	MH368121	100	100	Wastewater
WTUH-9	WTH1	Hypoxia	+	2	467	*Pseudomonas* sp. RW149	MH591593	100	100	River water
WTUH-12	WTH1	Hypoxia	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUH-13	WTH1	Hypoxia	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUH-14	WTH1	Hypoxia	+	1	489	*Acinetobacter* sp. HX09	MH368121	99	100	Wastewater
WTUH-15A	WTH1	Hypoxia	-	3	613	*Acinetobacter* sp. HX09	MH368121	100	100	Wastewater
WTUH-16	WTH1	Hypoxia	+	1	483	*Chryseobacterium* sp. StRB294	AB581560	100	100	Plant
WTUH-18	WTH1	Hypoxia	+	1	994	*Pseudomonas* sp. B	FJ546072	99	100	Soil
WTUH-19	WTH1	Hypoxia	+	1	1032	*Pseudomonas baetica* Ghg51-2	MH201226	99	100	Soil
WTUH-20	WTH1	Hypoxia	+	1	1023	*Chryseobacterium* sp. StRB294	AB581560	99	100	Soil
WTUH-25	WTH1	Hypoxia	-	1	979	*Acinetobacter* sp. HX09	MH368121	99	100	River water
WTUH-26	WTH1	Hypoxia	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUH-27	WTH1	Hypoxia	+	1	996	*Acinetobacter* sp. HX09	MH368121	99	100	River water
WTUH-28	WTH1	Hypoxia	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
WTUH-29	WTH1	Hypoxia	+	1	1028	*Chryseobacterium* sp. StRB294	AB581560	100	100	Soil
WTUH-30	WTH1	Hypoxia	+	1	1015	*Chryseobacterium* sp. StRB294	AB581560	100	100	Soil
WTUH-31	WTH1	Hypoxia	+	1	869	*Acinetobacter* sp. HX09	MH368121	100	100	River water
WTUH-32	WTH1	Hypoxia	+	1	960	*Acinetobacter* sp. HX09	MH368121	100	100	River water
WTUH-33	WTH1	Hypoxia	+	1	905	*Citrobacter farmeri* SSA-1555	MF186607	99	100	*Coptotermes formosanus*
WTUH-34	WTH1	Hypoxia	+	1	956	*Chryseobacterium* sp. TV93Nov	KJ482906	100	100	Soil
WTUH-35	WTH1	Hypoxia	+	1	973	*Acinetobacter* sp. HX09	MH368121	100	100	River water
WTUH-36	WTH1	Hypoxia	+	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
TUA-1A	TH1	Air	+	1	984	*Bordetella petrii* DSM 12804	NR_074291	98	100	*NA*
TUA-1B	TH1	Air	-	1	831	*Lactococcus lactis* IMAU50403	MF623200	99	100	Food
TUA-2	TH1	Air	+	1	1004	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-3	TH1	Air	+	27	998	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-5	TH1	Air	+	33	1013	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-18	TH1	Air	+	1	1031	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-21	TH1	Air	+	1	1016	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-26	TH1	Air	-	2	1023	*Salmonella enterica* subsp. Arizonae	AB273736	98	100	*NA*
TUA-31	TH1	Air	-	1	1049	*Bacillus cereus* MS15	MH635295	99	100	Soil
TUA-32A	TH1	Air	+	1	929	*Bacillus* sp. IMAU61003	MF803665	99	100	Food
TUA-35	TH1	Air	+	1	994	*Chryseobacterium* sp. PCH235	MH096030	99	100	NA
TUA-36	TH1	Air	-	2	1022	*Acinetobacter calcoaceticus* HIW3200905	MG011543	99	100	Water
TUA-37	TH1	Air	+	3	982	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-40	TH1	Air	+	5	998	*Citrobacter farmeri* IHB B 1003	KF475827	99	100	Plant-Associated
TUA-42	TH1	Air	+	1	1016	*Citrobacter* sp. D1	LN995817	99	100	Oil Tank
TUA-43	TH1	Air	+	1	1015	*Citrobacter* sp. 36-4CPA	JF812082	99	100	Soil
TUA-44	TH1	Air	+	1	999	*Citrobacter* sp. D1	LN995817	99	100	Oil Tank
TUA-46	TH1	Air	-	1	338	*Tsukamurella pulmonis* B2-3	KP235204	99	100	Soil
TUA-49	TH1	Air	-	1	1047	*Bacillus cereus* MS15	MH635295	99	100	Soil
TUA-51	TH1	Air	+	1	996	*Citrobacter* sp. 36-4CPA	JF812082	99	100	Soil
TUA-54	TH1	Air	+	8	1024	*Citrobacter* sp. 36-4CPA	JF812082	99	100	Soil
TUA-57	TH1	Air	-	1	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
TUA-58	TH1	Air	-	2	*NS*	*Unidentified*	*NA*	*NA*	*NA*	*NA*
TUH-3	TH1	Hypoxia	+	1	876	*Pseudomonas koreensis* NES-CAP-3	MF079283	100	100	Soil
TUH-5	TH1	Hypoxia	+	1	665	*Citrobacter farmeri* SSA-1555	MF186607	99	100	*Coptotermes formosanus*
TUH-6	TH1	Hypoxia	+	1	886	*Citrobacter farmeri* SSA-1555	MF186607	99	100	*Coptotermes formosanus*
TUH-10B	TH1	Hypoxia	-	1	870	*Citrobacter farmeri* SSA-1555	MF186607	99	100	*Coptotermes formosanus*
TUH-12	TH1	Hypoxia	-	1	914	*Bacillus cereus* MS15	MH635295	99	100	Soil
TUH-13	TH1	Hypoxia	+	1	885	*Bacillus toyonensis* FS37	MH635275	100	100	Soil
TUH-19	TH1	Hypoxia	+	1	960	*Delftia acidovorans* B9	MF679148	100	100	Soil
TUH-23	TH1	Hypoxia	+	1	865	*Chryseobacterium* sp. zzx05	KJ009402	99	100	Plant
TUH-42	TH1	Hypoxia	+	1	869	*Citrobacter farmeri* SSA-1555	MF186607	99	100	*Coptotermes formosanus*

Table SB-2. Components of trace elements solution, 1000X

Compound	Concentration (per liter)
Hydrochloric acid (HCl), 20 mM	1.7 mL
Iron (III) chloride hexahydrate ($FeCl_3 \cdot 6H_2O$)	2.027 g
Boric acid (H_3BO_3)	30 mg
Manganese (II) chloride heptahydrate ($MnCl_2 \cdot 4H_2O$)	100 mg
Cobalt (II) chloride hexahydrate ($CoCl_2 \cdot 6H_2O$)	190 mg
Nickel (II) chloride hexahydrate ($NiCl_2 \cdot 6H_2O$)	24 mg
Copper (II) chloride dihydrate ($CuCl_2 \cdot 2H_2O$)	2 mg
Zinc chloride ($ZnCl_2$)	68 mg
Sodium selenite (Na_2SeO_3)	4 mg
Sodium molybdate (Na_2MoO_4)	30.9 mg

Table SB-3. Barcodes used in the library preparation of *ureC* gene amplicons for sequencing

Sample	ureC Reverse Primer (UreC-R)	Barcode Sequence	Barcode + Reverse Primer (5' - 3')
WTH1-1	5'-*barcode*-GGTGGTGGCACACCATNANCATRTC-3'	TGAAAGCGGTCC	GGACCGCTTTCA GGTGGTGGCACACCATNANCATRTC
WTH1-2		CATAGGACCGTG	CACGGTCCTATG GGTGGTGGCACACCATNANCATRTC
WTH1-3		CAAACGTCATTC	GAATGACGTTTG GGTGGTGGCACACCATNANCATRTC
WTH1-4		CGTGGCGTAAGT	ACTTACGCCACG GGTGGTGGCACACCATNANCATRTC
WTH1-5		TAAGAAAGGCGT	ACGCCTTTCTTA GGTGGTGGCACACCATNANCATRTC
WTH1-6		CACAGGCACCAA	TTGGTGCCTGTG GGTGGTGGCACACCATNANCATRTC
WTH1-7		TCAGATCCGATG	CATCGGATCTGA GGTGGTGGCACACCATNANCATRTC
WTH1-8		ATGGAAGACATG	CATGTCTTCCAT GGTGGTGGCACACCATNANCATRTC
WTH1-9		GCCAACTGTAAC	GTTACAGTTGGC GGTGGTGGCACACCATNANCATRTC
WTH1-10		GTAACGAGTCCG	CGGACTCGTTAC GGTGGTGGCACACCATNANCATRTC
WTH-11		TCCAGTGCGAGA	TCTCGCACTGGA GGTGGTGGCACACCATNANCATRTC
TH1-1	5'-*barcode*-GGTGGTGGCACACCATNANCATRTC-3'	ACAAGACCAGAA	TTCTGGTCTTGT GGTGGTGGCACACCATNANCATRTC
TH1-2		GTCCAAGTGGAC	GTCCACTTGGAC GGTGGTGGCACACCATNANCATRTC
TH1-3		AGCCTCTAAATC	GATTTAGAGGCT GGTGGTGGCACACCATNANCATRTC
TH1-4		TTAACGGCTGAC	GTCAGCCGTTAA GGTGGTGGCACACCATNANCATRTC
TH1-5		TCCAGAAACCGT	ACGGTTTCTGGA GGTGGTGGCACACCATNANCATRTC
TH1-6		CAATATGGCTGC	GCAGCCATATTG GGTGGTGGCACACCATNANCATRTC
TH1-7		TAGCACACCTAT	ATAGGTGTGCTA GGTGGTGGCACACCATNANCATRTC
TH1-8		CACTAGCTAGGT	ACCTAGCTAGTG GGTGGTGGCACACCATNANCATRTC
TH1-9		CAGTGTCAGGAC	GTCCTGACACTG GGTGGTGGCACACCATNANCATRTC
TH1-10		TTAGTTGAGTCC	GGACTCAACTAA GGTGGTGGCACACCATNANCATRTC
TH1-11		ACGAACCCGTAT	ATACGGGTTCGT GGTGGTGGCACACCATNANCATRTC
TH1-12		ACAGGTGAAAGG	CCTTTCACCTGT GGTGGTGGCACACCATNANCATRTC