AN ABSTRACT OF THE DISSERTATION OF

Trang T. H. Nguyen for the degree of Doctor of Philosophy in Soil Science presented on March 19, 2019.

Title: Diversity and Activity of Microbial Extracellular Peptidases in Soil

Abstract approved: _____

David D. Myrold

Soil nitrogen exists largely as organic matter, including plant liter, dead animal matter, and microbial necromass. About 90% of soil organic nitrogen is proteinaceous material that is too large for plants and microorganisms to assimilate directly. Protein depolymerization therefore plays a critical role in mobilizing this organic source of nitrogen, producing lower molecular weight molecules that are bioavailable for both microorganisms and plants. The decomposition of proteins in soils serves as the rate-limiting step of the nitrogen cycle. The ability of microorganisms to access and break down proteinaceous material depends largely on their production of extracellular peptidases, but it involves a trade-off with the energetic cost of producing and secreting these enzymes into the environment, including the risk that other microorganisms can compete with the peptidase-producing organisms for the products released through depolymerization. Consequently, in order to optimize this energy investment, there

might be a tight connection between soil environmental conditions and microbial proteolytic activity. Despite its ecological importance, there is a lack of understanding about the diversity of these extracellular peptidases and their activity as an important factor influencing the protein degradability in soils.

In this dissertation, I first assessed the genetic potential for microorganisms to produce extracellular enzymes, and second, I developed and applied a novel approach to measure the activities of different classes of peptidases in soil. In my first two chapters, I evaluated the abundance and diversity of microbial extracellular peptidases, their evolutionary conservation, and distribution as a function of environmental habitat and lifestyle. Chapter 2 focuses on the secreted peptidases of prokaryotes (*Archaea*, *Bacteria*); chapter 3 focuses on *Fungi*, the dominant soil eukaryote. In both chapters, I analyzed secreted peptidases across microbial lineages using their genomic information and corresponding annotated protein sequences assembled from several databases, including MEROPS, Silva, JGI Genome Portal, and MycoCosm. Peptidase gene sequences of 147 archaeal, 2,191 bacterial and 612 fungal genomes were screened for secretion signals, resulting in 55,072 prokaryotic and 31,668 eukaryotic genes coding for secreted peptidases. I found that *Archaea*, *Bacteria*, and *Fungi* possess unique complements of secreted peptidases and there are differences in the number of secreted peptidases per genome, indicating potential differential abilities for organic nitrogen acquisition. The majority of secreted peptidase families not only follow the phylogenetic evolutionary distribution, but also segregate based on the microbial lifestyles and microbial habitats. This suggests that microorganisms optimize their secreted peptidases to match their surrounding environments.

In Chapter 4, I incorporated the use of selective inhibitors to block the activity of different classes of peptidases. I designed a protocol with these peptidase inhibitors to use directly in natural soils. I validated and optimized this protocol with pure enzymes and peptidase-supplemented soils. This research revealed that the profile of extracellular peptidase activities belonging to different catalytic types varies among soils and correlates with both soil chemical and microbial properties. This is in line with our assumption that soil microorganisms respond to their environmental conditions by investing in peptidases that can optimize their activity.

Collectively, this work provides a comprehensive and foundational understanding about the contribution of different catalytic types of microbial extracellular peptidases to organic nitrogen turnover in soils.

Diversity and Activity of Microbial Extracellular Peptidases in Soil

by
Trang T. H. Nguyen

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented March 19, 2019
Commencement June 2019

Doctor of Philosophy dissertation of Trang T. H. Nguyen presented on March 19, 2019

APPROVED:

Major Professor, representing Soil Science

Head of the Department of Crop and Soil Science

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Trang T. H. Nguyen, Author

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my major professor, Dr. David D. Myrold. Dr. Myrold has always been my academic father and the best mentor that anyone could ever ask for. He has supported me a tremendous amount towards my path to become an independent researcher. Dr. Myrold once asked me how I learned to bike when I was little. He said mentoring a young soul of a science lover like me was just the same way, that he would take his hands off my bike and let me peddle on my own but he would always be there and watch me to make sure I am on the right path. I learned so much from him for not only life but also science lessons, that sometimes there would be obstacles along the way that might require a detour to get to the destination, but giving up will never solve the problem. Thanks to Dr. Myrold, every day, I remind myself to stay focused, think outside the box, maintain self-discipline and move forwards. Thank you so much for thousands of hours discussing science with me for the past nearly 6 years, and more to come!

I would like to thank my sponsor, Vietnam Education Foundation (VEF), a US government agency that provided me funding for the first two years of my PhD. Being a fellow of Vietnam Education Foundation Fellowship program is a privilege. VEF made me feel like home during the time I studied abroad in the United States, especially at the annual conferences. Thanks to VEF, I get to know so many young talented Vietnamese scientists, a generation that hopefully will bring Vietnam to be one of the science and technology leading countries in the world. Thank you Sandy Dang for being a great executive director during my time, a wonderful sister to the VEF family, and a person who always reminds me to believe in myself and motivates me to be better every day.

I would like to thank my colleagues at Oregon State University for all the collaboration and for being extremely helpful with all my work. Dr. Ryan Mueller taught me to become a microbial informatician from a person who has zero experience in such field. Thank you for your patience and dedication. Dr. Markus Kleber guided me to think critically like a scientist and become a better science writer. Dr. Kristin Kasschau, my lab manager, always gave me helpful advice in the lab, guided me how

to be successful in graduate school, and reminded me every day that self-confidence would always be the key to success and independence. I would like to thank Dr. Peter Bottomley, Dr. Anne Taylor, Dr. Brett Mellbye, Dr. Armanda Roco for providing me valuable feedback to improve my research during our Tuesday journal clubs. I appreciate the continuous support from my committee members for answering my questions and guiding me along the way, Dr. Julie Pett-Ridge, Dr. Walt Ream. Other special thanks are for Kelsey Martin, Hayley Peter-Contesse, Ashley Waggoner, Christopher Burgess, Stephany Chacon, Josh Neuman, Rachel Danielson, Nate Tisdell, Andrew Giguere, Clara Weidman, Burl Carpenter, Danielle Runion, Gina O'Relley, Pedro Martinez, Yulin Sun, Fumi Funahashi, Xinda Lu, Shannon Cappellzazi, Ivan de Souza, Gloria Ambrowiak, Vance Almquist, Kris Osterloh, Elizabeth Seeno and an endless list of other graduate students for always being there for me and helping me to answer all the questions I had.

I would like to thank the Department of Crop and Soil Science for administrative support to make sure that everything runs smoothly, that goes to Dr. Jay Noller, Dr. Russ Karrow, Dr. Tom Chastain, Kristin Rifai, Emmalie Goodwin, Tracy Mitzel, Shauna Gutierrez. I also thank Dr. Kate Lajtha and the Sustainability program for offering me teaching assistantship. I thank the Graduate School at Oregon State University for the international and domestic travel award so that I can share my work in many national and international conferences. I especially like to thank Sandarshi Gunawardena from VEF and Emiko Christopherson from the International Office for assisting me with the paperwork that allow me to be here and maintain my legal visa status.

I thank my friends in Corvallis, Jonathan Marcotte, Anh Ha, Aron Boettcher, Graham Kerwin, Trung Vu, Duong Nguyen, Thai Duong, Sea Chu, Linh Nguyen, Truc Vo, Phung Lai, Tam Nguyen, Thang Nguyen, Thuan Nguyen, Khoi Nguyen, Ni Trieu, Anh Ninh, Scott Jordan, Melanie Jordan, and many others for always being there for me and for all the emotional support.

Last but not least, I would like to thank my parents and my sister for always believing in me, supporting me with all your love, and especially distilling in me the love of science at a young age.

CONTRIBUTION OF AUTHORS

Dr. Ryan Mueller assisted with the data extraction and computational analysis approach in Chapter 2. Dr. David Myrold and Dr. Ryan Mueller supervised the findings of the work and contributed to the writing of Chapter 2 publication. Dr. Markus Klebler provided samples for Chapter 4 research and contributed to the writing of this chapter. Dr. David Myrold supervised the findings of the work in Chapter 4 and also contributed to the data analysis and writing of this chapter.

TABLE OF CONTENTS

Page

Chapter 1: General Introduction 1

1.1. Protein turnover as the rate-limiting step of the terrestrial nitrogen cycle 2

1.2. Diversity of microbial extracellular peptidases as the regulating factor for protein turnover 3

1.3. The activity of catalytic types of peptidases as a factor of soil properties 5

1.4. Reference 7

Chapter 2: Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny and Habitat 10

2.1. Abstract 11

2.2. Introduction 12

2.3. Materials And Methods 15

2.3.1. Collection of secreted archaeal and bacterial peptidases and signal sequence identification 15

2.3.2. Comparison of genomic complements of secreted peptidases 17

2.3.3. Phylogenetic analysis 17

2.3.4. Distance matrices comparisons 18

2.3.5. Phylogenetic conservation and clustering 18

2.4. Results 19

2.4.1. Distribution of secreted peptidases across prokaryotic kingdoms 19

2.4.2. Distribution of secreted peptidases within prokaryotic kingdoms 21

2.5. Discussion 24

2.6. Appendix for Chapter 2 44

2.7. Reference 61

TABLE OF CONTENTS (Continued)

Page

Chapter 3: Diversity and Distributions of Extracellular Peptidases Across Fungal Genomes 73

3.1. Abstract 74

3.2. Introduction 75

3.3. Materials And Methods 80

3.3.1. Collection of fungal secreted peptidases 80

3.3.2. Fungal ecological guild annotation 82

3.3.3. Comparison of genomic complements of secreted peptidases 82

3.3.4. Phylogenetic analysis 83

3.3.5. Distance matrices comparisons 83

3.3.6. Phylogenetic conservation and clustering 84

3.3.7. Secreted peptidase distribution in association with fungal ecological functions. 85

3.4. Results 86

3.4.1. Relative abundance of secreted peptidase super-families across fungal kingdom 86

3.4.2. Distribution of secreted peptidases across fungal phylogeny 86

3.4.3. Distribution of secreted peptidases across fungal ecological groups 89

3.5. Discussion 91

3.5.1. Secreted peptidase distribution correlate with fungal taxonomy 91

3.5.2. Secreted peptidase distribution correlate with fungal ecological functions and lifestyles 94

3.6. Appendix for Chapter 3 111

3.7. Reference 116

TABLE OF CONTENTS (Continued)

Page

Chapter 4: Contribution of different catalytic types of peptidases to soil proteolytic activity 123

4.1. Abstract 124

4.2. Introduction 125

4.3. Material and Methods 128

4.3.1. Site characteristics and soil sampling 128

4.3.2. Peptidase assays 129

4.3.3. Effectiveness of peptidase inhibitors and inhibitor concentrations using model peptidases 130

4.3.4. Confirmation of the effectiveness and duration of inhibition in peptidase-supplemented soils 133

4.3.5. Potential interference of other organic compounds or abiotic proteolysis 135

4.3.6. Determination the relative pools of proteolytic enzymes in soils 136

4.3.7. Statistical analyses 137

4.4. Results 138

4.4.1. Effective peptidase inhibitor concentrations 138

4.4.2. Inhibition confirmed in peptidase-supplemented soils 138

4.4.3. Specificity of Folin-Ciocalteau assay confirmed for proteolytic activity 140

4.4.4. Abiotic proteolysis is negligible 141

4.4.5. Extracellular peptidase profiles are related to soil properties 141

4.5. Discussion 142

4.5.1. Substrate-induced proteolytic assay: initial control, length of incubation, and linearity 142

4.5.2. Modifications to measure activities of different peptidase classes 144

4.5.3. Biotic proteolysis: Correlation with soil biochemical properties 146

4.5.4. Metallopeptidase activity dominant and reflective of soil pH 146

TABLE OF CONTENTS (Continued)

Page

4.5.5. Fungi contribute to soil extracellular aspartic peptidase activity 147
4.5.6. Serine peptidases associated with soils under different tree species 148
4.5.7. Uncharacterized proteolytic activity 148
4.6. Appendix for Chapter 4 160
4.6.1. Folin-Ciocalteau Total Proteolytic Activity Assay 160
4.6.2. OPAME Total Free Amino Acid Analysis 163
4.7. Reference 173
Chapter 5: Conclusion 180

LIST OF FIGURES

Figure Page

Figure 2.1. Relative abundance of secreted peptidase super-families in 147 archaeal and 2,191 bacterial genomes (asparagine, aspartic, cysteine, glutamic, metallo-, mixed, serine, threonine, and unknown peptidase super-families). Different colors represent different peptidase super-families. 34

Figure 2.2. Secreted peptidase gene content (per genome) of archaeal and bacterial phyla. Secreted peptidases were grouped into super-families: Total secreted peptidases (including genes from all peptidase super-families); serine, metallo-, cysteine peptidases. The number of analyzed genomes from each prokaryotic phylum is presented next to the phylum names. 35

Figure 2.3. Bipartite association network of shared peptidase families between *Archaea* and *Bacteria*. Node sizes indicate the relative abundance of the secreted peptidases. Node shapes represent different peptidase families: triangle = aspartic; octagon = cysteine; diamond = glutamic; rectangle = metallo-; pentagon = serine; parallelogram = threonine; hexagon = asparagine, mixed, and unknown. Node colors are coded by unique or shared peptidase families between microbial kingdoms (blue = *Bacteria*, red = *Archaea*, gray = shared between *Bacteria* and *Archaea*). Edges denote associations between microbial kingdoms and peptidase families. Edge colors are coded by microbial kingdoms. 36

Figure 2.4. Principal coordinate analysis of secreted peptidase families based on Bray-Curtis dissimilarity of proportions of secreted peptidase families encoded in each genome. Symbol shapes are coded by microbial kingdoms; symbol colors represent some abundant bacterial and archaeal phyla. Significant differences of the secreted peptidase profiles were observed between archaeal and bacterial species (p-value < 0.001, F-statistic = 92.9, PERMANOVA) and among different bacterial phyla (p-value < 0.001, F-statistic = 19.0, PERMANOVA)..... 37

Figure 2.5. Distribution of secreted peptidase super-families across the archaeal 16S rRNA phylogenetic tree. Outer tracks show the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered. 38

Figure 2.6. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in archaeal genomes. Symbol shapes and colors are coded by reported optimal growth conditions (pH, temperature, and salt concentration). Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). The composition of secreted peptidase genes of *Archaea* varied significantly based on their optimal growth conditions (p-value < 0.001, F-statistic = 6.11, PERMANOVA)..... 39

LIST OF FIGURES (Continued)

Figure Page

Figure 2.7. Distribution of secreted peptidases super-families across the bacterial phylogenetic tree. Outer tracks show the copy numbers of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered. 40

Figure 2.8. Distribution of strain-specific secreted peptidases across (A) *Bacteroidetes* and (B) *Actinobacteria* taxa. The outer two tracks represent the habitat each taxon is associated with (orange for aquatic/soil environment, purple for animal environment). The middle tracks show the copy number of genes from each secreted peptidase super-family in each genome. The figure on the top left corner represent the average numbers of each peptidase super-family (S for serine, M for metallo-, C for cysteine, T for threonine, A for aspartic, N for asparagine, U for unknown, P for mixed and G for glutamic) per genome; purple box plots report mean and standard deviation of the peptidase content of genomes that are commonly found in animal-associated environments, and orange box plots report mean and standard deviation of the peptidase content of genomes from soil/aquatic environments, with p-values of: *** = 0-0.001, ** = 0.001-0.01, and * = 0.01-0.05. 42

Figure 3.1. Relative abundance of secreted peptidase super-families in 612 fungal genomes (aspartic, cysteine, glutamic, metallo-, serine, threonine, and unknown peptidase super-families) and in different fungal phyla. Different colors represent different peptidase super-families. 97

Figure 3. 2. Secreted peptidase gene content (per genome) of fungal phyla. Secreted peptidases were grouped into super-families: Total secreted peptidases (including genes from all peptidase super-families); serine, aspartic, metallo- peptidases as more abundant families; cysteine, glutamic, threonine, unknown peptidases as less abundant families. The number of analyzed genomes from each fungal phylum is presented next to the phylum names. Letters on top of each box represent statistical differences of secreted peptidases between three fungal phyla using Tukey's HSD analysis ($p \leq 0.05$). 98

Figure 3.3. Bipartite association network of shared peptidase families between *Ascomycota, Basidiomycota* and *Mucoromycota* fungal phylum. Node sizes indicate the relative abundance of the secreted peptidases. Node shapes represent different peptidase families: octagon = aspartic; hexagon = cysteine; diamond = glutamic; circle = metallo-; square = serine; triangle = threonine and unknown. Node colors are coded by unique or shared peptidase families between microbial kingdoms (green = *Ascomycota*, yellow = *Basidiomycota*, pink = *Mucoromycota*, gray = shared between phyla). Edges denote associations between fungal phyla and peptidase families. Edge colors are coded by fungal phyla. 99

Figure Page

Figure 3.4. Principal coordinate analysis of fungal genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in fungal genomes. Symbol shapes and colors are coded by fungal phyla. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). The composition of secreted peptidase genes of *Fungi* varied significantly based on their phyla ($p = 0.001$, F-statistic =61.044)..... 100

Figure 3.5. Distribution of secreted peptidase super-families across the fungal phylogenetic tree (pruned from Mycocosm fungal tree). The most outer tracks show functional guilds that each fungal genome is generally assumed (solid circle means the genome belongs to the certain guild, the open circle means the genome does not belong to the certain guild). The second outer tracks show the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered..... 101

Figure 3.6. Mantel correlogram between phylogenetic distance and secreted protease profile dissimilarities for fungal taxa based on Pearson's product-moment correlations (p-value < 0.05, filled squares; not significant, open squares). 102

Figure 3.7. Phylogenetic distributions of secreted peptidase families across fungal taxa. 103

Figure 3.8. Secreted peptidase gene content (per genome) of fungal guilds. Secreted peptidases were grouped into super-families: Total secreted peptidases (including genes from all peptidase super-families); serine, aspartic, metallo- peptidases as more abundant families; cysteine, glutamic, threonine, unknown peptidases as less abundant families. The number of analyzed genomes from each fungal phylum is presented next to the phylum names. Letters on top of each box represent statistical differences of secreted peptidases between fungal functional guilds using Tukey's HSD analysis ($p \leq 0.05$). 104

Figure Page

Figure 3.9. (A) Distribution of secreted peptidase super-families across the phylogenetic tree of 86 symbiotrophic fungi (pruned from Mycocosm fungal tree). The most outer tracks represent the trophic mode of these fungi (pathotrophic-symbiotrophic in dark green circles and symbiotroph, mostly representing ectomycorrhizal fungi, in yellow circles). The second outer tracks show the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered. (B) Secreted peptidase gene content (per genome) of 86 symbiotrophic fungal groups. Secreted peptidases were grouped into super-families. The number of analyzed genomes from each fungal group is presented next to the fungal guild classification. The green box plots report mean and standard deviation of the peptidase content of genomes that can be either pathotrophic or symbiotrophic, and the orange box plots report mean and standard deviation of the peptidase content of symbiotrophic genome with p-values. 105

Figure 3.10. (A) Distribution of secreted peptidase super-families across the phylogenetic tree of 95 saprotrophic fungi (pruned from Mycocosm fungal tree). The 95 saprotrophic fungi in this paper was simply classified into the “Saprotrophic” group of fungal guild. The outer most set of tracks represents the main classification of these saprotrophs (white-rot vs. brown-rot fungal species). The second set of outer tracks shows the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered. (B) Secreted peptidase gene content (per genome) of brown-rot and white-rot saprotrophic fungal groups. Secreted peptidases were grouped into super-families. The number of analyzed genomes from each fungal group is presented next to the fungal trait classification. The brown box plots report mean and standard deviation of the peptidase content of genomes that belonging to brown-rot fungi, and the white box plots report mean and standard deviation of the peptidase content of white-rot fungi genomes, with p-values. 107

LIST OF FIGURES (Continued)

Figure Page

Figure 3.11. (A) Distribution of secreted peptidase super-families across the phylogenetic tree of 90 pathotrophic fungi (pruned from Mycocosm fungal tree). The 90 pathographic fungi in this paper were simply classified into the "Others" group of fungal guild. The outer most set of tracks represents the main classification of these pathogens (plant, animal and parasite fungal species). The second set of outer tracks shows the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered. (B) Secreted peptidase gene content (per genome) of pathotrophic fungal groups. Secreted peptidases were grouped into super-families. The number of analyzed genomes from each fungal group is presented next to the fungal guild classification. The green box plots report mean and standard deviation of the peptidase content of genomes that belonging to plant pathogenic fungi, and the purple box plots report mean and standard deviation of the peptidase content of animal pathogenic fungi, and orange box plots for fungal parasite genomes, with p-values. 109

Figure 4.1. Proteolytic activities over 48 h in peptidase-supplemented soil in response to different peptidase inhibitor treatments (Table 2). Data points are means with standard error bars (n=3). Symbol shapes, colors, and line types represent different peptidase inhibitor treatments. Asterisks denote significant difference from control at 48 h (*) for $0.001 < p < 0.01$; (**) for $p < 0.001$)..... 155

Figure 4.2. Correlation between proteolytic activity measured by the OPAME assay (total free amino acids, leucine as standard) and by the Folin-Ciocalteau assay (using tyrosine as standard). Soil proteolytic activities from four different locations were monitored over 48 hours (n=3). Slopes for individual sites were not statistically different ($p \leq 0.05$). 156

Figure 4.3. Proteolytic activities over 48 h in soils from two different ecoregions (Cascade Head and H.J. Andrews Experimental Forests) and two tree species (Douglas-fir and red alder). Data points are means with standard error bars (n=3). Symbol shapes, colors, and line types represent different peptidase inhibitor treatments. 157

Figure 4.4. Biotic and abiotic sources of proteolytic activity from four different soils were calculated using least square regression. The pie charts (A) represent the relative abundance of each peptidase family in each of the four soils, with the size of each chart proportional to the total proteolytic activity. The table (B) shows proteolytic activity (mean ± standard error, μmol tyrosine g^{-1} dry soil d^{-1}) for different peptidase pools in each soil. The statistical significance of each pool size (n=3) are shown in bold (p-values of $\alpha \leq 0.05$). 158

Figure Page

Figure 4.5. Correlation between some soil properties and different components of soil proteolytic activity of four different soils: (A) total soil peptidase activity (μmol tyrosine g^{-1} dry soil d^{-1}) vs. gross ammonification (μg N g^{-1} dry soil d^{-1}), (B) relative metallopeptidase activity vs. soil pH, and (C) relative aspartic peptidase activity vs. fungal:bacterial PLFA ratio. Different points represent the means of different soils (solid = Cascade Head soils, open = H.J. Andrews, Circle = red alder, Triangle = Douglas-fir) with standard error bars. 159

LIST OF TABLES

Table Page

Table 4.1. Soil chemical properties (Boyle et al., 2008; Lu et al., 2015) and soil texture determined by hydrometer method. Data are mean ± standard error (n=3). 150

Table 4.2. Peptidases and peptidase inhibitors used in peptidase-supplemented soil experiments. Tested combinations have "x". Generally, inhibitors were used against only their target peptidase unless non-target interactions had been observed in pure enzyme studies. 151

Table 4.3. The matrix used for least squares regression to calculate for the relative activity of each peptidase group in soil, based on the relationships between the peptidases and the peptidase inhibitors. One indicates that proteolytic activity in not inhibited by the treatment; zero indicates the treatment completely inhibits proteolytic activity. 152

Table 4.4. The inhibitory efficiency of peptidase inhibitors on model peptidases. The numbers represent the percentage of remaining activity after the peptidase inhibitor additions. Bold represent significant inhibition ($p \leq 0.001$; Supplemental Table S4.3). 153

Table 4.5. Significance of proteolytic activity during the first and second 24-h periods of the incubation for each individual peptidase-supplemented soil combination. The p-values for these multiple Welch's two sample t-test analyses were adjusted using the Hommel procedure. Data are p-values for comparisons between the two time periods, with p-values of $\alpha \leq 0.01$ shown in bold. When the p-values are large (> 0.01), it indicates that proteolytic activity was constant during the entire 48-h incubation... 154

LIST OF APPENDICES

Appendix Page

4.6.1. Folin-Ciocalteau Total Proteolytic Activity Assay 160

4.6.2. OPAME Total Free Amino Acid Analysis 163

LIST OF APPENDIX FIGURES

Figure Page

Figure S2.1. Secreted peptidase gene content of 147 archaeal and 2,191 bacterial genomes (genes/mega base pairs (Mb), natural log transformation). *Archaea* in red and *Bacteria* in blue (Welch two-sample t-test, t=-24.0, p-value < 0.001)..... 44

Figure S2.2. Secreted peptidase gene content (genes/genome) of threonine and aspartic peptidases in prokaryotic phyla. The number of analyzed genomes from each prokaryotic phylum is presented next to the phylum names. 45

Figure S2.3. Mantel correlogram between phylogenetic distance and secreted protease profile dissimilarities for archaeal and bacterial taxa based on Pearson's product-moment correlations (p-value < 0.05, filled symbols; not significant, open symbols). 46

Figure S2.4. Phylogenetic distributions of secreted peptidase families across archaeal and bacterial taxa..... 47

Figure S2.5. Secreted peptidase families showing clumped distributions within bacterial genomes. Outer tracks show the presence/absence of genes from each secreted peptidase family in each genome (white=absence, red=presence). Inner ring colors represent the phylum-level classification of microbial genomes..... 48

Figure S2.6. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in each genome. Symbol shapes are coded by either Gram-negative and Gram-positive cell wall classification. Significant differences were observed between Gram-positive and Gram-negative bacteria based on the relative abundance of their secreted peptidase families (p-value < 0.001, F-statistic = 193.3, PERMANOVA)..... 49

Figure S2.7. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in *Bacteroidetes* genomes. Symbol shapes are coded by the environment where the microorganisms are commonly found (triangle for animal microbiota and circle for aquatic/soil environment); symbol colors represent different taxonomic families. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). Secreted peptidase profiles were strongly correlated with the environment in which each species was associated with (p-value < 0.001, F-statistic = 34.5, PERMANOVA). 50

LIST OF APPENDIX FIGURES (Continued)

Figure Page

Figure S2.8. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in *Actinobacteria* genomes. Symbol shapes are coded by the environment where the microorganisms are commonly found (triangle for animal microbiota and circle for aquatic/soil environment); symbol colors represent different taxonomic families. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). Ecological niches, on the other hand, also played a strong role in differentiating this function among Actinobacteria (p-value < 0.001, F-statistic = 26.4, PERMANOVA)..... 51

Figure S3.1. Relative abundance of secreted peptidase super-families in 612 fungal genomes (aspartic, cysteine, glutamic, metallo-, serine, threonine, and unknown peptidase super-families) and in different fungal ecological guilds. Different colors represented different peptidase super-families…….. 111

Figure S3.2. Principal coordinate analysis of fungal genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in fungal genomes. Symbol shapes and colors are coded by fungal guild. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). The composition of secreted peptidase genes of *Fungi* varied significantly based on their guild (p = 0.001, F-statistic =3.599)..... 112

LIST OF APPENDIX TABLES

Table Page

Table S2.1. Differences of secreted peptidases among bacterial phyla using Tukey's HSD analysis. 52

Table S2.2. Phylogenetic signal strength of secreted peptidase families across 2,191 bacterial genomes. Significance of clustering is based on the Fritz and Purvis index (D) of each peptidase family trait (presence or absence of genes). Estimated D value defined whether the secreted peptidase distribution would follow "strongly clumped" ($D \leq 0$), or "Brownian-like evolutionary" ($0 < D < 1$) or "Random" distribution ($D \geq 1$). 53

Table S2.3. Phylogenetic signal strength of secreted peptidase families across 147 archaeal genomes. Significance of clustering is based on the Fritz and Purvis index (D) of each peptidase family trait (presence or absence of genes). Estimated D value defined whether the secreted peptidase distribution would follow "strongly clumped" ($D \leq 0$), or "Brownian-like evolutionary" ($0 < D < 1$) or "random" distribution ($D \geq 1$)... 57

Table S2.4. Statistical differences of secreted peptidases between *Bacteroidetes* families using Tukey's HSD analysis..... 59

Table S2.5. Statistical differences of secreted peptidases between *Actinobacteria* families using Tukey's HSD analysis..... 60

Table S3.1. Phylogenetic signal strength of secreted peptidase families across 612 fungal genomes. Significance of clustering is based on the Fritz and Purvis index (D) of each peptidase family trait (presence or absence of genes). Estimated D value defined whether the secreted peptidase distribution would follow "strongly clumped" ($D \leq 0$), or "Brownian-like evolutionary" ($0 < D < 1$) or "Random" distribution ($D \geq 1$). consenTRAIT (τD), a phylogenetic metric that evaluates the sequence similarity of clusters of sharing trait organisms. τD for any singleton entry (trait that only presents in one genome) was scored by half the distance to the nearest node (assuming the likelihoods to find a neighbor organism with/without the trait are equal) 113

Table S4.1. Model peptidases used in this study. 167

Table S4.2. Peptidase inhibitors and targeted peptidases. 168

Table S4.3. Dunnett's p-values for comparisons of differences between the peptidase inhibitor treatments and the control for each individual model peptidase. Between-treatment comparison p-values of $\alpha \leq 0.01$ in bold. 169

Table S4.4. Dunnett's p-values for comparisons of differences between the peptidase inhibitor treatments and the control for each individual peptidase-supplemented soil combination. Between treatment comparison p-values of $\alpha \leq 0.01$ in bold..... 170

LIST OF APPENDIX TABLES (Continued)

Table Page

Table S4.5. Summary for regression analyses for the correlation between proteolytic activity measured by the OPAME assay (total free amino acids, leucine as standard) and by the Folin-Ciocalteau assay (using tyrosine as standard). Soil proteolytic activities from four different locations were monitored over 48 hours (n=3). 171

Table S4.6. Comparison of rates of proteolytic activity (μmol tyrosine g^{-1} dry soil h^{-1}) during the first and second 24-h periods of the incubation for each individual soil-inhibitor combination. The p-values for these multiple Welch's two sample t-test analyses were adjusted using the Hommel procedure. Data are p-values for t-test comparisons between the two time periods, with p-values of $\alpha \leq 0.05$ shown in bold. When the p-values are large (> 0.05), it indicates that proteolytic activity was constant during the entire 48-h incubation. 172

DEDICATION

To my parents,
Anh N. Tran and Thuong V. Nguyen,
For always being my learning role model and distilling your love for science in me since my young age,

To my little sister,
Ngoc B. Nguyen,
For always being there for me and reminding me to believe in myself,

To my big family back in Vietnam,
For always checking in on me and supporting me with whatever I do

And all of my friends in every corner of the world,
For all your survival meals, for the pictures, little notes and cards,
For the constant encouragement and emotional supports

Without whom none of my success would be possible

Chapter 1: General Introduction

Trang T. H. Nguyen

1.1. Protein turnover as the rate-limiting step of the terrestrial nitrogen cycle

Soil is a complex and interconnected terrestrial system comprised of organic matter, minerals, liquid, and gases. Soil provides habitat for many living organisms, from archaea, bacteria, fungi to plants, and animals. Organic matter decomposition, acting on plant litter, animal dead matter, microbial necromass, is one of the most dynamic and critical biogeochemical processes that happens in soils, whereby carbon, nitrogen, and other nutrients are recycled to enrich the soil biodiversity.

In natural terrestrial systems, nitrogen is usually limited due to the high biological demands and losses (Galloway et al., 2004) and this implies the constant biological need for organic nitrogen decomposition in the soil environments. Soil nitrogen is largely in the organic matter form and about 90% of the organic nitrogen is proteinaceous material (Kögel-Knabner, 2006; Nannipieri and Paul, 2009). Only a minor fraction of soil organic nitrogen is small enough to be available for plant and microbial nitrogen uptake (Nannipieri and Paul, 2009); as a result, this nitrogen pool is not adequate to satisfy the biological demands for soil organisms. Protein depolymerization therefore plays a critical role in mobilizing this organic source of nitrogen and turning it into lower molecular weight molecules that are bioavailable for uptake and assimilation by both plants and microorganisms. The decomposition of proteins in soils serves as the bottleneck and rate-limiting step of the nitrogen cycle (Jan et al., 2009; Schimel and Bennett, 2004). Following protein depolymerization, nitrogen is either competitively taken up as low molecular weight monomers (oligomeric peptides, amino acids) by plants and microorganisms, or mineralized into bioavailable ammonium or nitrate, which can be assimilated for cell growth, and/or eventually lost from the system as different aqueous or gaseous forms (Schimel and Bennett, 2004). Microorganisms with their diverse enzymatic potential play an important role in this process.

1.2. Diversity of microbial extracellular peptidases as the influencing factor for protein turnover

The ability of microorganisms to access and break down nitrogen from organic matter depends largely on their production of extracellular enzymes, which for nitrogen is primarily through the activity of peptidases. Peptidases are the catalytic enzymes that cleave the peptide bonds in protein molecules. Producing these peptidases and secreting them to the environment, though, is an energy-expensive investment, which must be weighed against the risks that other microorganisms may take advantage of the activity of the peptidases once they are released into the environment (Allison, 2005). To be conservative with this investment, each microbe has likely optimized its peptidases to function well in the environment it inhabits. For example, selection might favor peptidases with a broad substrate specificity or that target the common substrates found in a given environment; or in some other cases, selection might favor the enzymes that optimally function under specific soil conditions (pH, temperature, etc.). Despite the ecological importance, there is a lack of understanding about the diversity of these extracellular enzymes as an important factor influencing the protein degradability in soils.

Peptidases are produced by all forms of life and associate with diverse biochemical functions both inside and outside the cells (Page and Cera, 2008; Rao et al., 1998). Peptidases are considered to have emerged early during biological evolution in order to catalyze the cleavage of the peptide bonds between amino acid residues of proteins (Rao et al., 1998). Some extracellular peptidases are produced for pathogenic functions, such as damaging the host tissues or deactivating host defenses (Semenova et al., 2017). Most extracellular peptidases, however, are produced for the purpose of nutrient acquisition, and in some cases, to benefit the mutualistic relationships. For example, ectomycorrhizal fungi, in the symbiotic relationship with plants, are very active in

mobilizing soil organic nitrogen and converting it to bioavailable nitrogen so that fungi can exchange with plants for photosynthate carbon. The ability of microorganisms to utilize soil organic nitrogen by producing extracellular enzymes is expected to depend on the microbial lifestyles and microhabitats. For instance, some free-living saprotrophic bacteria might have higher demand to utilize organic nitrogen from the soils as the main source of nitrogen, compared to the bacteria that can fix nitrogen and turn it into bioavailable nutrients for cell development. As a result, free-living saprotrophic bacteria might possess a more enriched secreted peptidase capability and higher proteolytic performance.

Peptidases are generally classified into 7 main super-families (classes) based on their catalytic mechanisms: aspartic, asparagine, cysteine, glutamic, metallo-, serine, and threonine peptidases (Rawlings et al., 2018). These classes are further categorized into 255 peptidase families based on the similarity in their amino acid sequences. According to Page and Cera (2008), archaea, bacteria, and fungi can be distinguished from each other by specific peptidase families that are shared within each taxonomic group (including intracellular and extracellular peptidases). This differentiation between microbial proteolytic potential is driven by the number of peptidase coding genes in the genomes as well as their peptidase composition. This suggests that there may be functional specialization and optimization of microbial species depending on the peptidases encoded by each taxon. Because soil microbial community structure varies among soil environments, there are likely differences in relative peptidase activity and diversity, which might cause an impact on organic nitrogen cycling.

The first two research chapters offer a comprehensive study of the extracellular peptidase diversity across different microbial kingdoms (archaea, bacteria, and fungi). These chapters address the question of whether the proteolytic function outside the cells differs among microbial

taxa and if there is a connection between the microhabitats and microbial lifestyles and the distinguishable set of secreted peptidases across the three microbial kingdoms. My approach is to extensively analyze the diversity of secreted peptidases and their distributions across *Archaea*, *Bacteria*, and *Fungi* by using their genomic information and annotated protein sequences assembled in several databases, including MEROPS, JGI Genome Portal, MycoCosm, and Silva (Grigoriev et al., 2014, 2011; Quast et al., 2013; Rawlings et al., 2018). I expected to see a widespread distribution of secreted peptidase coding genes across the prokaryotic and fungal tree of life. I also analyzed if each peptidase family is evolutionarily conserved among phylogenetically-related taxa, to determine whether more closely related microbial species tend to share the same specific type of secreted peptidases. The catalytic efficacy of secreted peptidases from different classes are known to be influenced by environmental conditions, such as pH and temperature, therefore, I hypothesized that the distribution of peptidases also varies as a function of the ecological microhabitats occupied by different microbial taxa. These findings provide essential insights into the complement of secreted peptidases in microorganisms within different environments, which could be further validated using assays of peptidase gene expression and proteolytic activity. More broadly, knowledge about the diversity of secreted peptidases can enhance our understanding about the connection between microbial community composition and their biogeochemical functions in terms of organic nitrogen cycling.

1.3. The activity of catalytic types of peptidases as a factor of soil properties

Due to the energetic cost of extracellular enzyme production, the enzyme activity would be expected to be optimized to soil conditions, and it should reflect the adaptation of the living soil microbes to maximize the organic nitrogen acquisition in the soils. Subsequently, there might be an inherent correlation between soil properties on one hand and the microbial proteolytic activity

on the other hand. Little is known, however, about the relative contributions of different microbial catalytic classes of peptidases to soil proteolytic activity under native soil conditions and whether or not these contributions are influenced by the soil properties.

Protease functional types can be classified by using inhibitors specific to their active sites (Rawlings et al., 2018). Some studies have used inhibitors to distinguish among secreted peptidases, but none reflect the proteolytic activity under native soil conditions. Soil peptidases were either extracted (Hayano, 1993; Kamimura and Hayano, 2000; Watanabe and Hayano, 1995) or produced from isolated microorganisms (Rineau et al., 2016; Shah et al., 2013). By contrast, only a few studies have directly used peptidase inhibitors in soil or aquatic samples (Kumar et al., 2004; Renella et al., 2002).

In the third research chapter, I developed a protocol to optimize the use of peptidase inhibitors at the native soil pH and room temperature in order to characterize the activities of different classes of extracellular peptidases in soils. I focused on using the inhibitors that selectively inhibit the four dominant classes of proteolytic enzymes (aspartic, cysteine, metallo-, and serine peptidases). The specificity and effective concentrations of inhibitors were tested with pure enzymes and peptidase-supplemented soils to confirm the inhibitory efficacy before being applied to soils. After developing the method, I measured the activities of different classes of extracellular peptidases from four soils in Oregon, which represent a gradient of biochemical properties, and evaluated the correlations between these enzymatic activities and different soil properties.

Collectively, this work provides a comprehensive and foundational understanding about the contribution of different catalytic types of microbial extracellular peptidases to organic nitrogen turnover in soils, and broadly to the overall productivity via N recycling rates.

1.4. Reference

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

Galloway, J.N., Dentener, F.J., Capone, D.G., Boyer, E.W., Howarth, R.W., Seitzinger, S.P., Asner, G.P., Cleveland, C.C., Green, P.A., Holland, E.A., Karl, D.M., Michaels, A.F., Porter, J.H., Townsend, A.R., Vöosmarty, C.J., 2004. Nitrogen Cycles: Past, Present, and Future. Biogeochemistry 70, 153–226. doi:10.1007/s10533-004-0370-0

Grigoriev, I.V., Cullen, D., Goodwin, S.B., Hibbett, D., Jeffries, T.W., Kubicek, C.P., Kuske, C., Magnuson, J.K., Martin, F., Spatafora, J.W., Tsang, A., Baker, S.E., 2011. Fueling the future with fungal genomics. Mycology 2, 192–209. doi:10.1080/21501203.2011.584577

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research 42, D699–D704. doi:10.1093/nar/gkt1183

Hayano, K., 1993. Protease activity in a paddy field soil: Origin and some properties. Soil Science and Plant Nutrition 39, 539–546. doi:10.1080/00380768.1993.10419794

Jan, M.T., Roberts, P., Tonheim, S.K., Jones, D.L., 2009. Protein breakdown represents a major bottleneck in nitrogen cycling in grassland soils. Soil Biology and Biochemistry 41, 2272–2282. doi:10.1016/j.soilbio.2009.08.013

Kamimura, Y., Hayano, K., 2000. Properties of protease extracted from tea-field soil. Biology and Fertility of Soils 30, 351–355. doi:10.1007/s003740050015

Kögel-Knabner, I., 2006. Chemical Structure of Organic N and Organic P in Soil, in: Nannipieri,

P., Smalla, K. (Eds.), Nucleic Acids and Proteins in Soil, Soil Biology. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 23–48. doi:10.1007/3-540-29449-X_2

Kumar, K., Rosen, C., Russelle, M., 2004. A novel approach to regulate nitrogen mineralization in soil. In: Controlling Nitrogen Flows and Losses., in: Controlling Nitrogen Flows and Losses. Academic Publishers, The Netherlands.

Nannipieri, P., Paul, E., 2009. The chemical and functional characterization of soil N and its biotic components. Soil Biology and Biochemistry 41, 2357–2369. doi:10.1016/j.soilbio.2009.07.013

Page, M.J., Cera, E.D., 2008. Evolution of peptidase diversity. Journal of Biological Chemistry 283, 30010–30014. doi:10.1074/jbc.M804650200

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research 41, D590-596. doi:10.1093/nar/gks1219

Rao, M.B., Tanksale, A.M., Ghatge, M.S., Deshpande, V.V., 1998. Molecular and biotechnological aspects of microbial proteases. Microbiology and Molecular Biology Reviews 62, 597–635.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Research 46, D624–D632. doi:10.1093/nar/gkx1134

Renella, G., Landi, L., Nannipieri, P., 2002. Hydrolase activities during and after the chloroform fumigation of soil as affected by protease activity. Soil Biology and Biochemistry 34, 51–60. doi:10.1016/S0038-0717(01)00152-3

Rineau, F., Stas, J., Nguyen, N.H., Kuyper, T.W., Carleer, R., Vangronsveld, J., Colpaert, J.V., Kennedy, P.G., 2016. Ectomycorrhizal Fungal Protein Degradation Ability Predicted by Soil Organic Nitrogen Availability. Applied and Environmental Microbiology 82, 1391–1400. doi:10.1128/AEM.03191-15

Schimel, J.P., Bennett, J., 2004. Nitrogen mineralization: Challenges of a changing paradigm. Ecology 85, 591–602. doi:10.1890/03-8002

Semenova, T.A., Dunaevsky, Y.E., Beljakova, G.A., Borisov, B.A., Shamraichuk, I.L., Belozersky, M.A., 2017. Extracellular peptidases as possible markers of fungal ecology. Applied Soil Ecology 113, 1–10. doi:10.1016/j.apsoil.2017.01.002

Shah, F., Rineau, F., Canbäck, B., Johansson, T., Tunlid, A., 2013. The molecular components of the extracellular protein-degradation pathways of the ectomycorrhizal fungus *Paxillus involutus*. New Phytologist 200, 875–887. doi:10.1111/nph.12425

Watanabe, K., Hayano, K., 1995. Seasonal variation of soil protease activities and their relation to proteolytic bacteria and *Bacillus* spp in paddy field soil. Soil Biology and Biochemistry 27, 197–203. doi:10.1016/0038-0717(94)00153-R

Chapter 2: Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny and Habitat

Trang T. H. Nguyen*

David D. Myrold, Ryan S. Mueller

Published In:

Frontiers in Microbiology - Evolutionary and Genomic Microbiology section

DOI: 10.3389/fmicb.2019.00413

Accepted: 18 February 2019

Avenue du Tribunal Fédéral 34, CH – 1005 Lausanne, Switzerland

2.1. Abstract

Proteinaceous compounds are abundant forms of organic nitrogen in soil and aquatic ecosystems, and the rate of protein depolymerization, which is accomplished by a diverse range of microbial secreted peptidases, often limits nitrogen turnover in the environment. To determine if the distribution of secreted peptidases reflects the ecological and evolutionary histories of different taxa, we analyzed their distribution across prokaryotic lineages. Peptidase gene sequences of 147 archaeal and 2 191 bacterial genomes from the MEROPS database were screened for secretion signals, resulting in 55 072 secreted peptidases belonging to 148 peptidase families. These data, along with their corresponding 16S rRNA sequences, were used in our analysis. Overall, *Bacteria* had a much wider collection of secreted peptidases, higher average numbers of secreted peptidases per genome, and more unique peptidase families than *Archaea*. We found that the distribution of secreted peptidases corresponded to phylogenetic relationships among *Bacteria* and *Archaea* and often segregated according to microbial lifestyles, suggesting that the secreted peptidase complements of microbial taxa are optimized for the environmental microhabitats they occupy. Our analyses provide the groundwork for examining the specific functional role of families of secreted peptidases in relationship to the organisms and the corresponding environments in which they function.

2.2. Introduction

Peptidases catalyze the cleavage of the peptide bonds between amino acid residues of proteins and are produced by all forms of life (Rao et al., 1998). These proteolytic enzymes are highly diverse in structure, perform multiple biological functions, and can be found in the cytoplasm within cells, tethered to the cell surface, or secreted into the environment. Secretion of extracellular peptidases represents a significant investment of metabolic energy, carbon, and nitrogen by microbial cells, enabling the acquisition of carbon or nitrogen from the environment (Allison et al., 2010; Chróst, 1991; Geisseler and Horwath, 2008; Kumar and Takagi, 1999; Landi et al., 2011).

Proteinaceous material is the most abundant form of soil organic nitrogen. Protein degradation into oligopeptides and amino acids, which can be directly and rapidly metabolized by microorganisms for nutrients and energy, is a critical strategy used by microorganisms to gain bioavailable nitrogen under nitrogen-limited conditions, especially in boreal and temperate forest soils (Geisseler et al., 2010; Schimel and Bennett, 2004). Some peptidases are secreted constitutively into the environment at low concentrations by microorganisms to initiate the degradation of proteins, although microorganisms can also regulate peptidase production and secretion based on their demands for carbon and nitrogen (Geisseler and Horwath, 2008).

In aquatic ecosystems, proteins and peptides contribute significantly to dissolved organic matter, accounting for 5-20% of dissolved organic nitrogen and 3-4% of dissolved organic carbon (Nagata et al., 1998; Pantoja and Lee, 1999). As in terrestrial ecosystems, microbial utilization of organic nitrogen in aquatic systems is regulated by the hydrolysis of these protein polymers (Chróst, 1991). Due to the more dilute nature of aquatic environments, peptidases bound to

microbial cells and sequestered in microbial biofilms are thought to be primarily responsible for the degradation of proteins, and allow microbes to readily take up the decomposition products for further metabolism (Chróst, 1991; Hoppe, 1991; Nagata et al., 1998; Nunn et al., 2003); however, free proteolytic enzymes also contribute to the available nitrogen pool (Chróst, 1991; Obayashi and Suzuki, 2008).

In animal-associated environments protein degradation can be associated with pathogenicity and host disease (Gibson and Macfarlane, 1988a; Richardson et al., 2013), in addition to having a role in direct nutrient acquisition. In some gut environments, microbial peptidases have been found to be constitutively produced and partially bound to the cell surface (Gibson and Macfarlane, 1988a, 1988b; Macfarlane et al., 1986). The regulated secretion of some extracellular peptidases is proposed to help pathogenic microorganisms competitively colonize and invade host cells and tissues by degrading host proteins, such as mucins, collagens, and other extracellular-matrix components (Duarte et al., 2016; Gibson and Macfarlane, 1988a; Loesche, 1988; Nakjang et al., 2012).

Peptidases are universal across all organisms and are considered to have developed early during biological evolution (Rao et al., 1998). Subsequent diversification has led to the development of several peptidase super-families (asparagine, aspartic, cysteine, glutamic, metallo-, serine, and threonine peptidases) that are grouped based on their mechanism of catalysis (Hartley, 1960; Häse and Finkelstein, 1993; Mooshammer et al., 2014; Rao et al., 1998; Rawlings, 2016; Rawlings et al., 2018; Rawlings and Barrett, 1993). Different classes of peptidases are associated with specific biological pathways, substrates, and catalytic reactions (Page and Cera, 2008; Rao et al., 1998; Rawlings and Barrett, 1993). Peptidase families have been shown to be distributed unevenly among microbial groups (Page and Cera, 2008), leading to broad generalizations about

associations of peptidases—both intracellular and extracellular—with different microbial groups. Aspartic peptidases are mostly encoded by *Fungi*, metallopeptidases are common in *Bacteria*, and cysteine and serine peptidases appear to be universal across microorganisms (Caldwell, 2005). *Bacteria* have consistently been found to be the dominant contributor to protease activity in soils and seawater based on studies using protease activity assays, pure culture protease expression, and approaches targeting peptidase genes (Katsuji et al., 1994; Nagata et al., 1998; Obayashi and Suzuki, 2008; Sakurai et al., 2007; Watanabe et al., 2003; Watanabe and Hayano, 1993, 1994, 1996). However, it is unclear how varying peptidase complements across genomes may impact variations in overall activity when considered at a community level or inferred by metagenomics studies. By developing a better understanding of factors influencing the abundance, diversity, and distribution of extracellular peptidase genes across all curated prokaryotic taxa (Arnosti, 2015; Burns et al., 2013), insights into the relationship between microbial community composition and protein degradation capabilities across environments can be gained.

Our goal was to analyze the diversity of secreted peptidases and their distributions across prokaryotic microorganisms by using annotated peptidase sequences collated in the MEROPS database (Rawlings, 2016; Rawlings et al., 2018). We expected to find secreted peptidases from different proteolytic super-families distributed widely across the prokaryotic tree of life. From this peptidase distribution pattern, we also sought to find evidence of whether each peptidase family is evolutionarily conserved among phylogenetically-related taxa. Because the catalytic efficiencies of secreted peptidases from different super-families are known to be affected by environmental conditions, we expected the distribution of peptidases to also vary as a function of the ecological microhabitats occupied by different microbial taxa. More broadly, the findings from these analyses might provide fundamental insights into the complement of secreted peptidases in microorganisms

within different environments, which could be further validated using assays of peptidase gene expression and proteolytic activity.

2.3. Materials and Methods

2.3.1. Collection of secreted archaeal and bacterial peptidases and signal sequence identification

Annotated peptidase sequences of 147 archaeal and 2 191 bacterial species were extracted from the MEROPS MySQL database release 11.0 (http://merops.sanger.ac.uk; (Rawlings, 2016)). Only completely annotated genomes with available 16S rRNA information existing in SILVA database release 128 were considered. Firstly, data pertaining to the organism name, taxonomy, and genome completeness was extracted from the "organism" and "classification" tables of the MySQL database (i.e., merops_taxonomy_id, taxonomy_id, and complete_genome values). The taxonomy_id values were used to query for matching 16S rRNA sequences from the SILVA database. Unique taxonomic IDs present in both databases for species that encode at least one secreted peptidase were used as the primary genomes of interest for this study.

The MEROPS database classifies peptidases into seven super-families based on the catalytic residue serving at the active site of the enzyme (Hartley, 1960; Rawlings and Barrett, 1993), and further divides these super-families into 255 proteolytic families based on similarities in amino acid sequences (Rawlings, 2016). The merops_taxonomy_id was used as a search query against the "features" and "sequence" tables of the MEROPS MySQL database to obtain all information pertaining to annotated peptidase sequences encoded within each genome of interest.

Exported information included the peptidase DNA sequence, the sequence_id, and the peptidase super-family and family classification.

All downloaded peptidase sequences were analyzed with SignalP 4.1 to identify genes encoding putative signal sequence motifs as defined for Gram-positive and Gram-negative bacteria (Mori and Ito, 2001; Petersen et al., 2011), yielding 55 072 secreted peptidases classified to 148 families. To validate the signal peptide prediction using SignalP, we analyzed these 55 072 sequences with Phobius, a combined transmembrane topology and signal peptide predictor, which has a reported higher sensitivity in discriminating between transmembrane domains and signal peptides (Käll et al., 2004, 2007). More than 98% of the sequences identified by SignalP were also identified as having signal petides by Phobius, 0.6% were identified as not having either signal peptides or transmembrane domains, and 1.3% were identified as transmembrane proteins without any signal peptides. Taking into account the imperfection that exists in all signal prediction models, we concluded that using SignalP was a valid method to identify prokaryotic secreted peptidases in our study. Peptidase sequences of *Firmicutes*, *Actinobacteria*, *Deinococcus-Thermus*, and *Archaea* were screened using a Gram-positive model; the remaining bacterial peptidase sequences were screened using a Gram-negative model. The Welch two-sample t-test, or unequal variances t-test, was used to evaluate significant difference between the means of the total secreted peptidases encoded within archaeal and bacterial genomes. One-way analysis of variance with the Tukey's HSD multiple-range test was used to determine the statistical differences between counts of total secreted peptidases among microbial phyla. Statistical analyses were performed in the 'R' programming environment (R. Core Team, 2016).

2.3.2. Comparison of genomic complements of secreted peptidases

The secreted peptidase complements of all taxa were summarized in a matrices containing the gene copy number counts of secreted peptidases assigned to either family or superfamily classifications (rows) across all analyzed genomes (columns). Bray-Curtis dissimilarity indices between the secreted peptidase complements of genomes were calculated from these matrices and used to generate a secreted peptidase distance matrix, or functional distance matrix, using the 'Vegan' package in 'R' (Oksanen et al., 2018). Principal coordinate analyses (PCoA) was used to explore the data and Permutational Multivariate Analyses of Variance (PERMANOVA) was used to determine the statistical differences of the peptidase complements of archaeal and bacterial genomes at different taxonomic levels. A bipartite association network of shared and unique peptidase families was generated using Cytoscape release 3.4.0 (http://www.cytoscape.org/) (Shannon et al., 2003).

2.3.3. Phylogenetic analysis

The 16S rRNA sequences of the selected archaeal and bacterial genomes were extracted from the SILVA database release 128 (http://www.arb-silva.de/) (Quast et al., 2013) and aligned using the NAST aligner (DeSantis et al., 2006). A 16S rRNA neighbor-joining phylogenetic tree was built from alignments using PHYLIP (Felsenstein, 1989). A phylogenetic distance matrix was also constructed using the F84 model of DNADIST (DeSantis et al., 2006). The phylogenetic tree and distributions of secreted peptidase families across the tree were visualized using iTOL (Letunic and Bork, 2016).

2.3.4. Distance matrices comparisons

Correlations between the phylogenetic distance matrix and the secreted peptidases distance matrix, or functional distance matrix, were evaluated using the Mantel test of 'APE' (Analysis of Phylogenetics and Evolution package) in 'R' (Paradis et al., 2004) based on Pearson's product-moment correlation. Mantel correlograms that report the correlation between phylogenetic and functional distances at defined phylogenetic distance classes for *Archaea* and *Bacteria* were calculated using the 'Vegan' package in 'R'.

2.3.5. Phylogenetic conservation and clustering

Phylogenetic signal strengths (D) contributing to the observed distribution patterns for each peptidase super-family and family were calculated from their binary presence/absence in genomes of all considered taxa (Fritz and Purvis, 2010) using the 'CAPER' package (Comparative Analyses of Phylogenetics and Evolution) in 'R' (Orme et al., 2018). Secreted peptidases are considered phylogenetically conserved when they are shared among the majority of members of branched clades, conforming to a Brownian motion evolutionary model ($D \sim 0$), with a relatively constant gain/retention of traits across taxonomic levels. A strongly clumped distribution ($D < 0$) suggests recent innovation or potential gain via horizontal gene transfer within a clade or subset therein. Peptidases are considered randomly distributed ($D \geq 1$) when their presence/absence is not driven by shared traits (e.g., physiology) of closely related species (Berlemont and Martiny, 2013; Martiny et al., 2013; Zimmerman et al., 2013).

To understand the association between the distribution of secreted peptidases and ecological microhabitats, we examined taxonomic subsets of microorganisms, including genomes of 147 *Archaea*, 275 *Actinobacteria*, and 182 *Bacteroidetes* species. Habitat preferences for

archaeal phyla and proposed optimal growth conditions (e.g., pH, temperature, and salt concentration ranges) and for bacteria were downloaded from the JGI GOLD database (Mukherjee et al., 2017). Habitat preferences were visualized on phylogenetic trees together with secreted peptidase count data using iTOL. For *Actinobacteria* and *Bacteroidetes datasets,* the Welch two-sample t-test was used to determine statistical differences between the mean gene copy number of each secreted peptidase super-family between microbial groups originating from different ecological habitats (e.g., soil and aquatic habitat vs. animal-associated habitat). PCoA was used to visualize the data in multidimensional space and PERMANOVA was used to determine the statistical differences of the peptidase complements of microbial groups originating from different ecological habitats. Vectors for peptidase families capturing a significant amount of variation in the total dataset were derived from Pearson correlations with the first two PCoA axes.

2.4. Results

2.4.1. Distribution of secreted peptidases across prokaryotic kingdoms

When normalized to genome size, *Bacteria* had significantly more secreted peptidase coding genes per Mb than *Archaea* (5.84 vs. 1.71, $p < 0.001$) (Fig. S2.1). In both kingdoms, serine, metallo-, and cysteine super-families contributed more than 80% of the secreted peptidase genes (Fig. 2.1). The numbers of peptidase genes per genome belonging to these abundant super-families were also significantly lower in archaeal than in bacterial phyla, agreeing with the general trends of kingdom-level peptidase super-family repertoires (Fig. 2.2). Conversely, significant biases were observed in some of the less common peptidase super-families: aspartic peptidases were more common in *Archaea* than *Bacteria* (9.4% vs. 0.6%), whereas threonine peptidases were more

commonly found in *Bacteria* than *Archaea* (2.3% vs. 0.6%) (Figs. 2.1 and S2.2). Asparagine, glutamic, mixed, and unknown peptidase super-families were rare (Fig. 2.1).

Observing the distribution of peptidase families, as opposed to super-families, offered finer-scale insights into the differential sets of secreted peptidases encoded by *Archaea* and *Bacteria*. Most of peptidase families encoded by *Archaea* were also common to *Bacteria*: 47 peptidase families of the serine, metallo-, cysteine, threonine, and glutamic super-families were shared between the two kingdoms, and contributed to more than one-third of the total peptidase families present in the dataset (Fig. 2.3). Only six peptidase families were unique to *Archaea*, four belonging to the aspartic super-family, whereas 95 peptidase families were unique to members of the bacterial kingdom (Fig. 2.3).

Principal coordinate analysis clustered phylogenetically distinct sets of microorganisms separate from each other based on the secreted peptidase families they encode (Fig. 2.4). Although each PCoA axis explained a low level of data variance, PERMANOVA tests indicated significant differences between different sets of microorganisms. For example, a strong and significant difference of the secreted peptidase profiles was observed between *Archaea* and *Bacteria* ($p < 0.001$) and between bacterial phyla ($p < 0.001$).

Distributions of peptidase families and corresponding super-families across microbial taxa were compared to the phylogenetic relationships among analyzed genomes using presence/absence profiles of peptidases in comparison to a 16S rRNA phylogenetic tree. The distributions of secreted peptidases were found to be significantly correlated with the 16S rRNA phylogeny within each kingdom (*Archaea* $r_{Mantel} = 0.303$, $p < 0.001$, *Bacteria* $r_{Mantel} = 0.334$, $p < 0.001$), indicating an evolutionary relationship in which subsets of phylogenetically related organisms in each

prokaryotic kingdom shared similar types of secreted peptidases. Mantel correlograms showed that conservation was strongest and most significant between more closely related taxa for both *Archaea* and *Bacteria* (Fig. S2.3). *Archaea* demonstrated a weakly significant relationship at all taxonomic levels examined, whereas relationships within *Bacteria* were weakly significant only between taxa that share ≥ 90% 16S rRNA gene sequence identity, beyond which pairs of taxa share little to no functional similarity in the secreted peptidases they encode (Fig. S2.3).

Distributions of individual secreted peptidase families were also evaluated for their phylogenetic dispersion (D). Most of peptidase families (71%) encoded in bacterial genomes showed evidence of non-random phylogenetic clustering (Fig. S2.4 and Table S2.2). Peptidase families with negative values ($D < 0$) represented those with the strongest clustering patterns across the phylogenetic tree. For example, M73 and M84 are endopeptidases that are predominantly restricted to *Bacillus sp.* and M07 is an endopeptidase found mainly in *Actinobacteria* species (Fig. S2.5). Conversely, 77% of peptidase families found within archaeal genomes exhibited random distribution patterns, devoid of phylogenetic signals (Fig. S2.4 and Table S2.3).

2.4.2. Distribution of secreted peptidases within prokaryotic kingdoms

There was no significant difference between the total number of secreted peptidase genes encoded in known *Crenarchaeal* and *Euryarchaeota* genomes ($p = 0.089$); however, the overall composition did vary significantly between these phyla. For example, the overabundance of aspartic peptidases observed at the kingdom-level could be primarily attributed to taxa belonging to *Crenarchaeaota*, as only 15 of 105 of *Euryarchaeota* genomes encoded these genes (Figs. 2.5 and S2.2). Intriguingly, 40% of the *Euryarchaeota* genomes encoding aspartic peptidases were classified as acidophiles, whereas only 8% of the entire *Euryarchaeota* dataset fell into this

environmental classification. Thus, a distinct enrichment was observed for the presence of aspartic peptidases in acidophilic *Euryarchaeota* genomes.

This link between environment and peptidase content of archaeal genomes was further explored with PCoA of peptidase profile data. The composition of secreted peptidase genes of *Archaea* varied significantly based on optimal growth conditions (pH, temperature, and salinity) ($p < 0.001$). Halophilic and haloalkilophilic archaeal clusters were most strongly correlated with the distribution of the serine S01, S08, S12, aspartic A22, and metallo M79 peptidase families (Fig. 2. 6). Thermophilic and thermophilic/acidophilic archaea also separated from each other and from the rest of archaea (Fig. 2.6). Serine S16 peptidase family was associated with the thermophilic archaea. Following the general trend proposed above, aspartic A05 and A37 families were associated with acidophilic archaea (Fig. 2.6). In addition, the presence of serine S53 family peptidases was strongly correlated with an acidophilic lifestyle.

Secreted peptidase profiles of bacterial genomes varied substantially across taxa (Fig. 2.7), with significant differences between phyla in total peptidase counts (Table S2.1) and in composition (Fig. 2.4). Significant differences were observed between Gram-positive and Gram-negative bacteria based on the relative abundance of their secreted peptidase families ($p < 0.001$) (Fig. S2.6). At the phylum level, *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Proteobacteria* were enriched with secreted peptidases, whereas the deep-branching *Aquificae* and *Thermotogae* taxa encoded fewer secreted peptidases (Fig. 2.2 and Table S2.1). *Acidobacteria* encoded the highest mean number of secreted peptidases per genome, with a large relative increase in the number of secreted metallopeptidases and a concomitant decrease in the number of cysteine peptidases compared to other bacterial phyla (Fig. 2.2).

Genomes of the *Bacteroidetes* and *Actinobacteria* phyla encoded high numbers of secreted peptidases and exhibited strong within-phylum distribution patterns related to finer-scale relationships (Fig. 2.7). For the *Bacteroidetes*, serine and metallopeptidases were dominant and well-conserved in presence and copy number across all species (Fig. 2.8A). By contrast, cysteine peptidases were more abundant (i.e., higher copy number per genome) in more recently evolved families of *Bacteroidetes*, such as *Porphyromonadaceae*, *Bacteroidaceae*, and *Prevotellaceae*, and threonine peptidases were more commonly found in deeper-branching lineages of *Bacteroidetes*, including *Cytophagaceae*, *Sphingobacteriaceae*, and *Flavobacteriaceae* (Fig. 2.8A). The latter group of *Bacteroidetes* also had a significantly higher numbers of secreted peptidases compared to the more recently evolved group ($p < 0.001$). Principal coordinate analysis of *Bacteroidetes* showed a significant separation among *Bacteroidetes* families based on the relative abundances of secreted peptidases ($p < 0.001$), which was strongly correlated with the environment associated with each species ($p < 0.001$) (Fig. S2.7). All species of *Porphyromonadaceae*, *Bacteroidaceae*, and *Prevotellaceae*, which encoded fewer secreted peptidases overall but a higher proportion of cysteine peptidases, were associated with an animal environment, whereas 74% of the *Cytophagaceae*, *Sphingobacteriaceae*, and *Flavobacteriaceae* species, which encode more peptidases overall and more threonine peptidases, were predominantly linked to aquatic or soil environments (Figs. 2.8A and S2.7, Table S2.4).

In the *Actinobacteria*, differences among clades were more strongly related to the numbers of peptidase genes than to the types of secreted peptidases (Fig. 2.8B) but were still highly correlated with the environmental microhabitat in which a taxon was found. For example, *Actinobacteria* families associated with animals, such as *Propionibacteriaceae*, *Coriobacteriaceae*, *Bifidobacteriaceae*, and *Corynebacteriaceae*, possessed a lower overall

abundance of secreted peptidases, predominantly of the serine, metallo-, and cysteine super-families, compared to aquatic or soil *Actinobacteria*, such as *Streptomycetaceae*, *Pseudonocardiaceae*, *Nocardiaceae*, *Micromonosporaceae*, and *Actinoplanaceae* (Fig. 2.8B and Table S2.5) ($p < 0.001$). Notable exceptions were host-associated *Mycobacteriaceae* genomes that encode significantly more secreted peptidases than other *Actinobacteria* associated with animal environments (Fig. 2.8B and Table S2.5) ($p < 0.001$). By contrast, *Frankiaceae*, which form nitrogen-fixing root nodules in several families of plants, possessed a low abundance of secreted peptidases compared to other aquatic or soil *Actinobacteria* ($p < 0.001$). Principal coordinate analysis of *Actinobacteria* peptidases showed a significant separation among taxonomic families in the relative abundance of secreted peptidases ($p < 0.001$) and their environment ($p < 0.001$) (Fig. S2.8).

2.5. Discussion

Serine, metallo-, and cysteine peptidases are the dominant (~90%) intracellular and extracellular proteolytic enzymes of *Archaea* and *Bacteria*, whereas aspartic and threonine peptidases contribute <10% to the total (Page and Cera, 2008). Intracellular peptidases are often involved with protein turnover and regulatory functions, whereas extracellular or secreted peptidases are typically viewed as an energetic investment of the organisms that is returned via the acquisition of carbon and nitrogen through enzymatic degradation of proteinaceous material in the environment (Chróst, 1991; Geisseler and Horwath, 2008).

Secreted peptidase diversity varied between *Archaea* and *Bacteria*, suggesting the potential for specialized peptidase functions and optimization among taxa. This variation may be related to

differences in the catalytic residues of the active site; these biochemical differences may provide specific adaptive advantages to different taxa under varying environmental conditions. As a general hydrolytic mechanism, a nucleophilic amino acid residue or water molecule is activated to attack a peptide carbonyl group, cleaving a peptide bond. In the case of serine, cysteine, and threonine peptidases, the histidine residue of a catalytic triad activates the serine, cysteine, or threonine residue, which then serves as the nucleophile that splits the peptide bond (Rao et al., 1998; Theron and Divol, 2014). Alternatively, for aspartic and metallopeptidases the aspartic acid residue or an enzyme-bound metal cofactor activates a water molecule to act as the nucleophile for the hydrolysis (Theron and Divol, 2014; Wu and Chen, 2011).

Bacterial species generally possess more secreted peptidases per genome and have a more diverse repertoire of secreted peptidase families compared to archaeal species (Figs. 2.2 and 2.3). This may confer greater flexibility on *Bacteria* to generate different types of extracellular proteolytic enzymes in response to specific environmental conditions depending on their demand for carbon and nitrogen, resulting in consistently high levels of overall peptidase activity *in situ*. This also suggests that *Bacteria* could be more competitive in obtaining organic nitrogen from the environment compared to *Archaea*. Empirical studies have implicated *Bacteria* to be the dominant contributor to proteolytic activity in soils (Katsuji et al., 1994; Sakurai et al., 2007; Watanabe et al., 2003; Watanabe and Hayano, 1996, 1993), and bacterial isolates from *Bacillus*, *Pseudomonas*, and *Flavobacterium-Cytophaga* have been shown to be important agents of proteolysis, acting as the main sources of soil peptidase activity (Bach and Munch, 2000; Vranova et al., 2013). Our analysis shows that these genera also have a high richness and abundance of secreted peptidases, consistent with their high soil peptidase activities.

At the super-family level, much of the variability in peptidase profiles between genomes of prokaryotic taxa was linked to differences in counts of less common peptidases, namely aspartic peptidases in *Archaea* and threonine peptidases in *Bacteria* (Figs. 2.1 and 2.3). Differences in the complement of secreted peptidases may reflect their adaptation to environmental conditions, such as temperature or pH. Serine, cysteine, and metallo- peptidases are generally optimized and active at neutral to alkaline pH (Rao et al., 1998; Rawlings, 2016), whereas aspartic peptidases generally exhibit high proteolytic activity in acidic conditions (Rao et al., 1998). Our analyses indicate that these enzymatic pH optima reflect the environments in which they are found. For example, three peptidase families – A05, A37, and S53 – were enriched in archaeal acidophile genomes (Fig. 2.6). All three of these peptidase families have been shown to have optimal endopeptidase activities at low pH (Rawlings et al., 2018). Additionally, peptidases of the S53 family appear to be novel endopeptidases within the serine peptidase super-family. These enzymes encode a catalytic triad consisting of Glu, Asp, and Ser residues, as well as an additional Asp residue in the oxyanion hole of the active site (Wlodawer et al., 2003). This active site arrangement stands in contrast to the traditional Asp, His, Ser triad observed in the more common serine S08 peptidases, and effectively relies on two additional acidic residues for activity. These active site arrangements likely relate to the activities of S08 and S53 peptidases in different pH environments, and may account for the observed strong negative correlation of the presence of S08 peptidases within acidophilic genomes. Therefore, the variation in the diversity of peptidase super-families encoded by microbes appears to be at least partially influenced by optimization of catalytic site to specific environmental conditions.

In *Bacteria*, there is a significant difference between the secreted peptidase composition of Gram-positive and Gram-negative bacteria. These two groups of *Bacteria* differ in cell wall

structure (Brown et al., 2015; Vollmer et al., 2008), which may influence their environmental distributions. In Gram-positive bacteria, extracellular enzymes could either be restricted to the cell wall and/or eventually diffuse into the environment (Chróst, 1991). In Gram-negative bacteria, digestive enzymes need to be secreted beyond the outer membrane in order to stimulate the degradation of polymers (Chróst, 1991). The distinction between these enzyme secretion strategies may influence the types of extracellular proteolytic enzymes encoded by these two groups of *Bacteria*. Practically, Gram-positive bacteria may secrete more free extracellular enzymes to the environment in comparison to Gram-negative bacteria with more membrane-bound secreted enzymes (Brown et al., 2015; Chróst, 1991; Vollmer et al., 2008); however, we did not observe a significant difference in the average number of secreted peptidases encoded in the genomes of taxa from these two bacterial groups ($p = 0.056$).

The conservation of secreted peptidase repertoires between pairs of archaeal and bacterial taxa was found to have a moderate positive correlation with phylogenetic relatedness across the prokaryotic tree of life. For both *Archaea* and *Bacteria* this relationship weakens rapidly as phylogenetic distance increases and, for bacterial taxa at least, the relationship is only significant up to approximately the family-level taxonomic equivalent of phylogenetic similarity. These patterns may be due in some part to horizontal gene transfer of peptidases, which may act to conserve features between more closely related taxa that more commonly exchange genes via this mechanism (Choi and Kim, 2007; Lawrence and Hendrickson, 2003). As discussed below, this conservation may also be partially attributed to a confounding correlation between phylogeny and environmental microhabitat of the taxa considered, given that phylogenetically-related taxa often inhabit grossly similar environments. Thus, our analyses are necessarily limited by the definition of microhabitat used here, which may insufficiently define the true microhabitats of each taxon

and the corresponding relationship to functional specialization of peptidase families within those habitats. Despite this potential limitation, these results stand in contrast to the insignificant relationship observed for glycoside hydrolase (GH) profiles and phylogenies of prokaryotes that was defined using a similar approach (Berlemont and Martiny, 2013). Various technical and analytical reasons could account for this discrepancy (e.g., differences in databases, classification schemes, and methodological details). However, stronger conservation of peptidase vs. GH content in genomes could also indicate biologically-driven differences in selective pressures on the different enzymatic types, despite their similar general functional roles in modifying cellular components and obtaining resources via secreted degradative enzymes. Further work will be needed to better define the roles of these important enzymes in speciation and competition in the environment.

Most secreted peptidase families encoded in bacterial genomes were determined to have significant phylogenetic signals in their distribution patterns across taxa. These findings agree with previous studies that found conservation of prokaryotic traits that are governed by multiple genes or metabolic pathways (e.g., spore formation, oxygenic photosynthesis; (Barberán et al., 2017; Goberna and Verdú, 2016)), suggesting that trait conservation and phylogenetic signal strength is not exclusively linked to increased trait complexity. Peptidase families of archaeal genomes did not show the same level of conservation as in *Bacteria*, a result that is likely due to the scant representation of peptidases from individual families in the available archaeal genome dataset, which is a known limitation of this phylogeny-based trait prediction method (Goberna and Verdú, 2016). Non-random distributions were also observed for most super-families when compared to both archaeal and bacterial phylogenies (Tables S2.4 and S2.5). Here, more negative D-values, which are indicative of extreme phylogenetic clustering, were typically observed for less common

peptidases (e.g., threonine peptidases of *Bacteria* and aspartic acid peptidases of *Archaea*), suggesting specialized adaptive roles for these enzymes based on their distinct catalytic mechanisms.

When the conservation and variation of genome-encoded secreted peptidases was examined within more specific bacterial clades (e.g., *Bacteroidetes* and *Actinobacteria* phyla), it was observed that environmental habitat and microbial lifestyle (i.e., "free-living" vs. "animal-associated") was an important determinant of peptidase content in genomes (Fig. 2.8). Generally speaking, *Bacteroidetes* taxa commonly associated with aquatic or soil environments, such as *Cytophagaceae*, *Sphinobacteriaceae*, and *Flavobacteriaceae*, encoded more total peptidases compared to animal-associated *Bacteroidetes*, such as *Bacteroides* and *Prevotella*. Additionally, although serine and metallo- peptidases were common to all *Bacteroidetes*, threonine peptidases were present almost exclusively in aquatic/soil-derived *Bacteroidetes* taxa, whereas cysteine peptidases were significantly enriched in animal-associated *Bacteroidetes* taxa. Given the nature of nitrogen limitation in most soil and aquatic environments, the ability to readily break down high molecular weight proteinaceous material into amino acid precursors for cell growth or energy generation would be highly favorable (Chróst, 1991; Geisseler and Horwath, 2008; Kolton et al., 2013). The *Flavobacteriaceae* include taxa with different lifestyles and genome sizes, and which are common inhabitants of terrestrial and marine ecosystems. Their ability to successfully compete in such oligotrophic environments may be dependent on their capacity to quickly degrade proteinaceous material to obtain nitrogen as a supplement to their well-established specialization of using carbohydrates for energy and as a carbon source (Bryson et al., 2017). This may account for the enriched proteolytic enzyme repertoire observed for these taxa, which is comprised of many outer membrane-associated and extracellular peptidases (Kolton et al., 2013; Tully et al., 2014).

By contrast, host-associated *Prevotella* species present inside the rumen (Griswold et al., 1999; Wallace et al., 1997) or as periodontal pathogens of humans (Gazi et al., 1997; Mallorquí-Fernández et al., 2008) have fewer secreted peptidases compared to soil/aquatic species in the *Bacteroidetes* phylum.

Similar to these trends, *Actinobacterial* families common to soil and aquatic environments (*Streptomycetaceae*, *Pseudonocardiaceae*, *Nocardiaceae*, *Micromonosporaceae*, and *Actinoplanaceae*) were also found to have a greater diversity and abundance of peptidases encoded in their genomes compared to animal-associated taxa. This observation agrees well with our current understanding of the ecology of *Actinomycetes* and their prodigious role as organic matter decomposers in nutrient-limited environments such as soils and freshwaters (Wink et al., 2017). *Streptomyces* species are abundant in terrestrial ecosystems and are well-known for their ability to use a wide variety of insoluble environmental substrates such as animal, plant, fungal, and microbial biomass by diverse extracellular enzymes, including peptidases (Chater et al., 2010). Interestingly, some secreted peptidases from *Streptomyces*, which are strictly regulated by their own inhibitors, are to cannibalize their own mycelial biomass in order to support aerial growth and sporulation when needed (Chater et al., 2010). With a rich repertoire of keratinases (mostly serine and metallo- peptidases), some *Streptomyces* can degrade keratin, an insoluble structural and highly polymerized protein that is commonly found in the outer covering of many animals (Chater et al., 2010; Gupta and Ramnani, 2006; Lange et al., 2016). In other cases, extracellular peptidases from *Streptomyces* may also play a role as an activating mechanism for other secreted proenzymes, such as nucleases, cellulases, and xylanases (Chater et al., 2010).

An exception within the soil-associated *Actinobacteria* with regard to secreted peptidase content are taxa within the *Frankiaceae* family, which are diazotrophic and can be endosymbionts

of actinorhizal plants. Mastronunzio et al., 2009 also noted the lower number of secreted hydrolases, including peptidases, associated with *Frankia* strains compared to soil *Actinobacteria*, and speculated that this may make them less harmful to their plant hosts, thereby facilitating nodulation.

This pattern does not hold for rhizobia, however, which is a group of diazotrophic *Alphaproteobacteria* that form root nodules in legumes and that encode a much richer collection of secreted peptidases compared to *Frankiaceae* or other non-N_2-fixing *Alphaproteobacteria*. Interestingly, rhizobia are known to fix nitrogen only when in a symbiosis with plants because rhizobia lack an endogenous oxygen protection mechanism for the nitrogenase enzyme that catalyzes the nitrogen fixation (Pawlowski and Bisseling, 1996). Thus, possessing an abundant collection of extracellular peptidases might be a strategy for free-living rhizobia to scavenge organic nitrogen and carbon from proteins. In contrast, *Frankia* species maintain their ability to fix atmospheric nitrogen to meet their nitrogen demand when free-living, potentially obviating their need to secrete peptidases to scavenge organic nitrogen from the environment (Norman and Friesen, 2017; Pawlowski and Bisseling, 1996).

Most animal-associated *Actinobacteria* were found to have a lower abundance of secreted peptidases in comparison with those taxa associated with aquatic or soil environments. Examples of the former are taxa from the *Bifidobacteriaceae*, which are often found as members of the human intestinal microbiota, especially in unweaned infant guts (Lee and O'Sullivan, 2010). In this environment, proteolytic activity predominantly arises from peptidases in human breast milk (e.g., anionic trypsin, anionic elastase, and plasmin) and from the gastric proteases (e.g., pepsin) (Dallas et al., 2012; Heyndrickx, 1963). Our analysis (Fig. 2.8) showed that *Bifidobacteriaceae* taxa have a limited potential to break down proteins, which might reflect the high abundance of

host-derived peptidases that generate bioavailable nitrogen within the gut. Conversely, another group of animal-associated *Actinobacteria*, the *Mycobacteriaceae*, possess large numbers of secreted peptidase genes in their genomes. Most *Mycobacterium* species are pathogenic (e.g., *M. tuberculosis, M. leprae*) (Gagneux, 2018) and their secreted peptidases appear to function in roles other than nutrient acquisition. For example, S01 peptidases, such as MarP or Rv3671c, protect *Mycobacterium* species from high acidic and oxidative conditions inside the host, especially when in a dormant state (Biswas et al., 2010; Botella et al., 2017; Kugadas et al., 2016; Ribeiro-Guimarães and Pessolani, 2007). Other peptidases in *Mycobacteria*, such as MycP1 of the S08 family, cleave proteins of the virulent secretion system as part of the infection process (Abdallah et al., 2007; Ohol et al., 2010; Ribeiro-Guimarães and Pessolani, 2007).

Collectively, these examples support the potential role of environmental microhabitat in selecting for peptidase functions, with the general theme that host-associated bacteria tend to encode fewer secreted peptidases than those taxa that are free-living. There are exceptions to this pattern, however, which appear to be linked to specialized traits of the microbes (e.g., pathogenicity, nitrogen fixation).

Our analysis of peptidase diversity has practical implications for microbial ecology studies of protein degradation. First, our analysis of the microbial potential for secreted peptidase production is a foundation for subsequent research applying transcriptomic or proteomic approaches to determine how this potential is realized by the secretion of peptidases under specific environmental conditions. Second, the current oligonucleotide primers designed to amplify peptidases from environmental DNA using PCR focus on specific peptidase families that are encoded by limited microbial taxa. For example, the *npr* primers are able to detect neutral metallopeptidases of the M04 family primarily associated with *Bacillus* species (23rd most

abundant peptidase family in *Bacteria* and *Archaea*), primers for *sub* detect the subtilisin-like S08 peptidase family associated with *Bacillus* species (9th in abundance), and *apr* primers can identify only alkaline metallopeptidases of the M10 family from *Pseudomonas fluorescens* (57th in abundance) (Bach and Munch, 2000; Tsuboi et al., 2014). Therefore, there is a need to design primer sets that are more universal or target a larger diversity of microbial secreted peptidases and that focus on the more abundant families of secreted peptidases (e.g., S11, C40, M23) in order to better capture the protein depolymerization process in environmental samples.

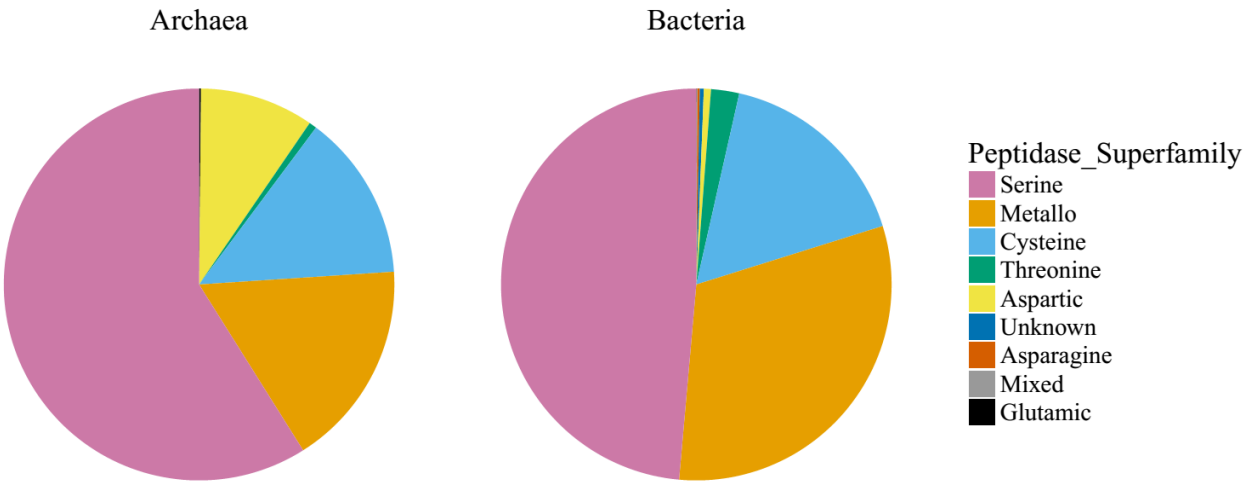


Figure 2.1. Relative abundance of secreted peptidase super-families in 147 archaeal and 2,191 bacterial genomes (asparagine, aspartic, cysteine, glutamic, metallo-, mixed, serine, threonine, and unknown peptidase super-families). Different colors represent different peptidase super-families.

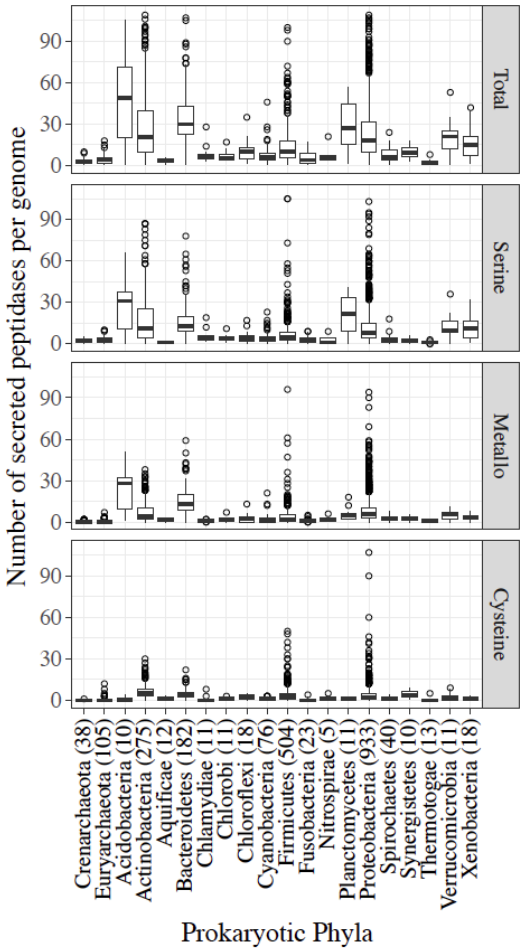


Figure 2.2. Secreted peptidase gene content (per genome) of archaeal and bacterial phyla. Secreted peptidases grouped into super-families: Total secreted peptidases (including genes from all peptidase super-families); serine, metallo-, cysteine peptidases. The number of analyzed genomes from each prokaryotic phylum is presented next to the phylum names.

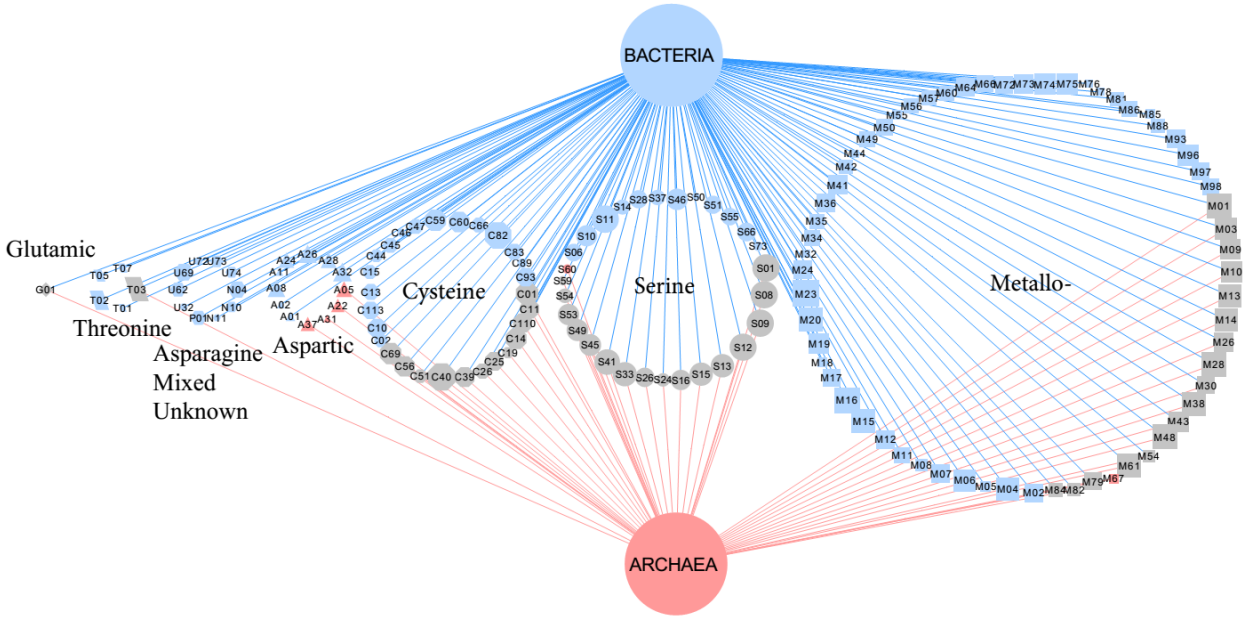


Figure 2.3. Bipartite association network of shared peptidase families between *Archaea* and *Bacteria*. Node sizes indicate the relative abundance of the secreted peptidases. Node shapes represent different peptidase families: triangle = aspartic; octagon = cysteine; diamond = glutamic; rectangle = metallo-; pentagon = serine; parallelogram = threonine; hexagon = asparagine, mixed, and unknown. Node colors are coded by unique or shared peptidase families between microbial kingdoms (blue = *Bacteria*, red = *Archaea*, gray = shared between *Bacteria* and *Archaea*). Edges denote associations between microbial kingdoms and peptidase families. Edge colors are coded by microbial kingdoms.

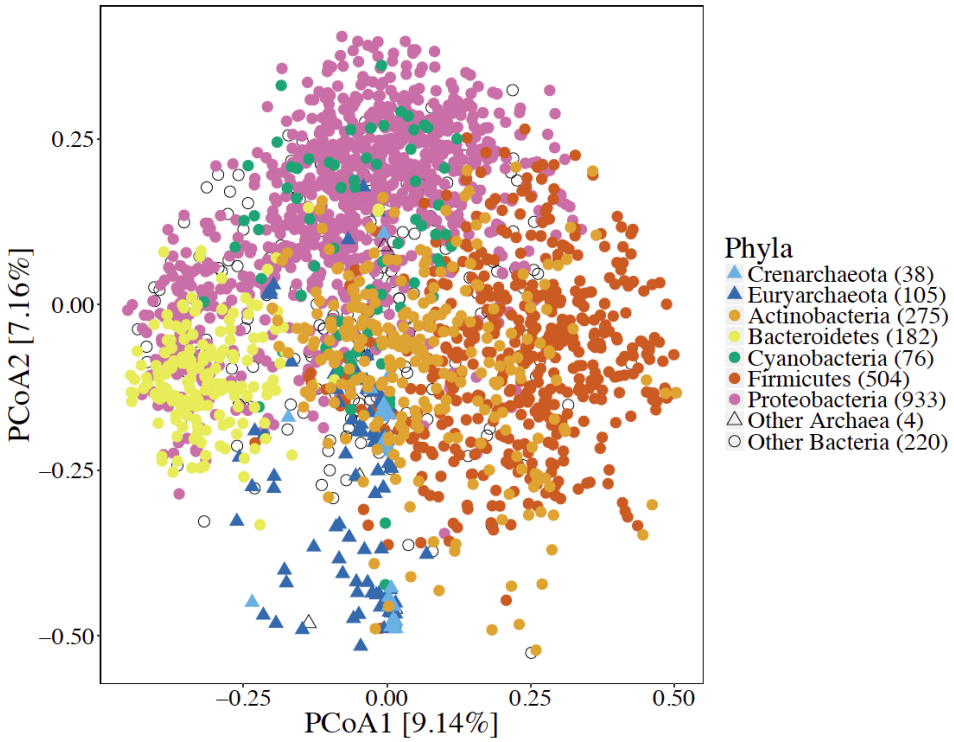


Figure 2.4. Principal coordinate analysis of secreted peptidase families based on Bray-Curtis dissimilarity of proportions of secreted peptidase families encoded in each genome. The number of analyzed genomes from each prokaryotic phylum is presented next to the phylum names. Symbol shapes are coded by microbial kingdoms; symbol colors represent some abundant bacterial and archaeal phyla. Significant differences of the secreted peptidase profiles observed between archaeal and bacterial species (p-value < 0.001, F-statistic = 92.9, PERMANOVA) and among different bacterial phyla (p-value < 0.001, F-statistic = 19.0, PERMANOVA).

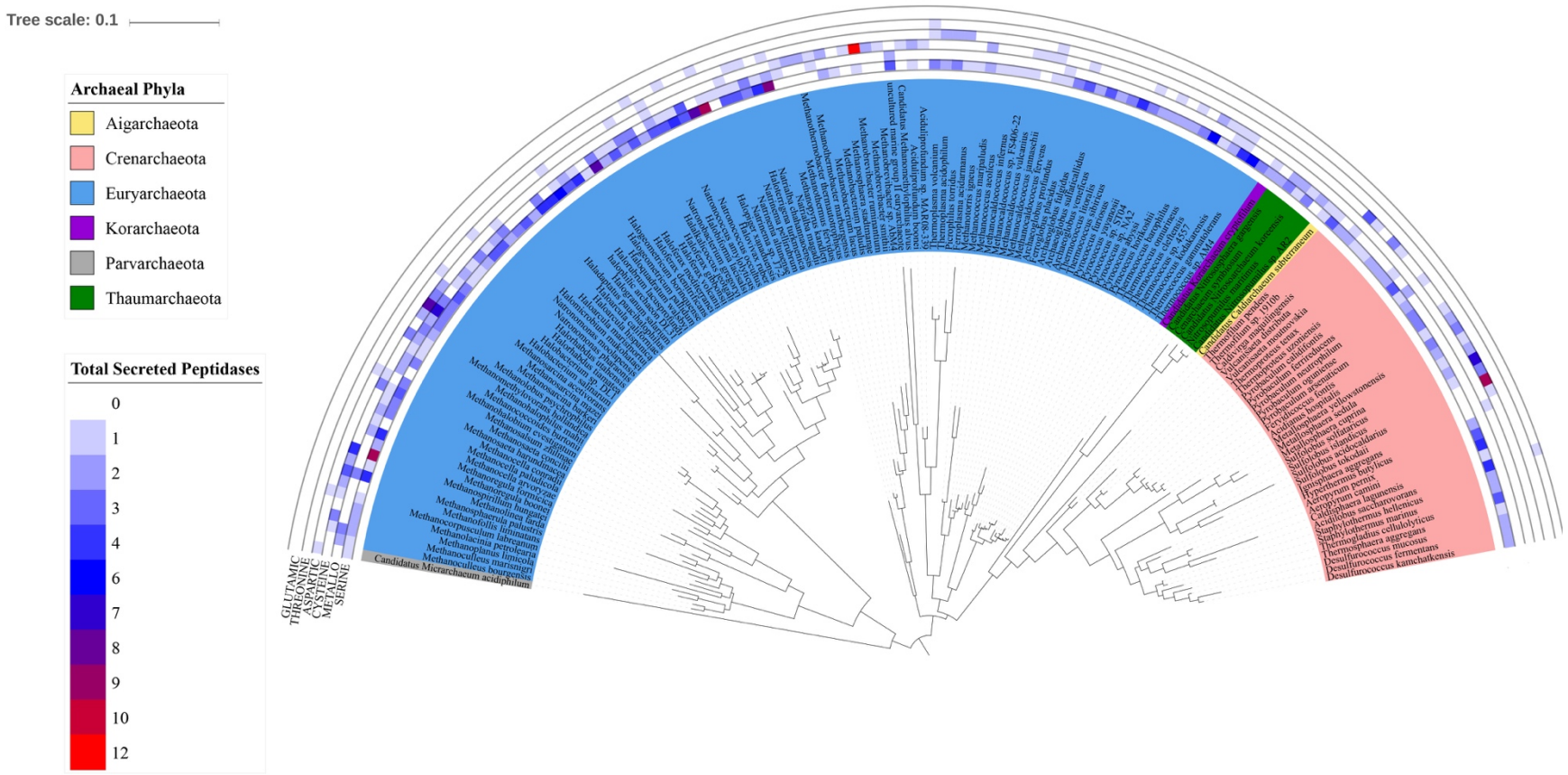


Figure 2.5. Distribution of secreted peptidase super-families across the archaeal 16S rRNA phylogenetic tree. Outer tracks show the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered.

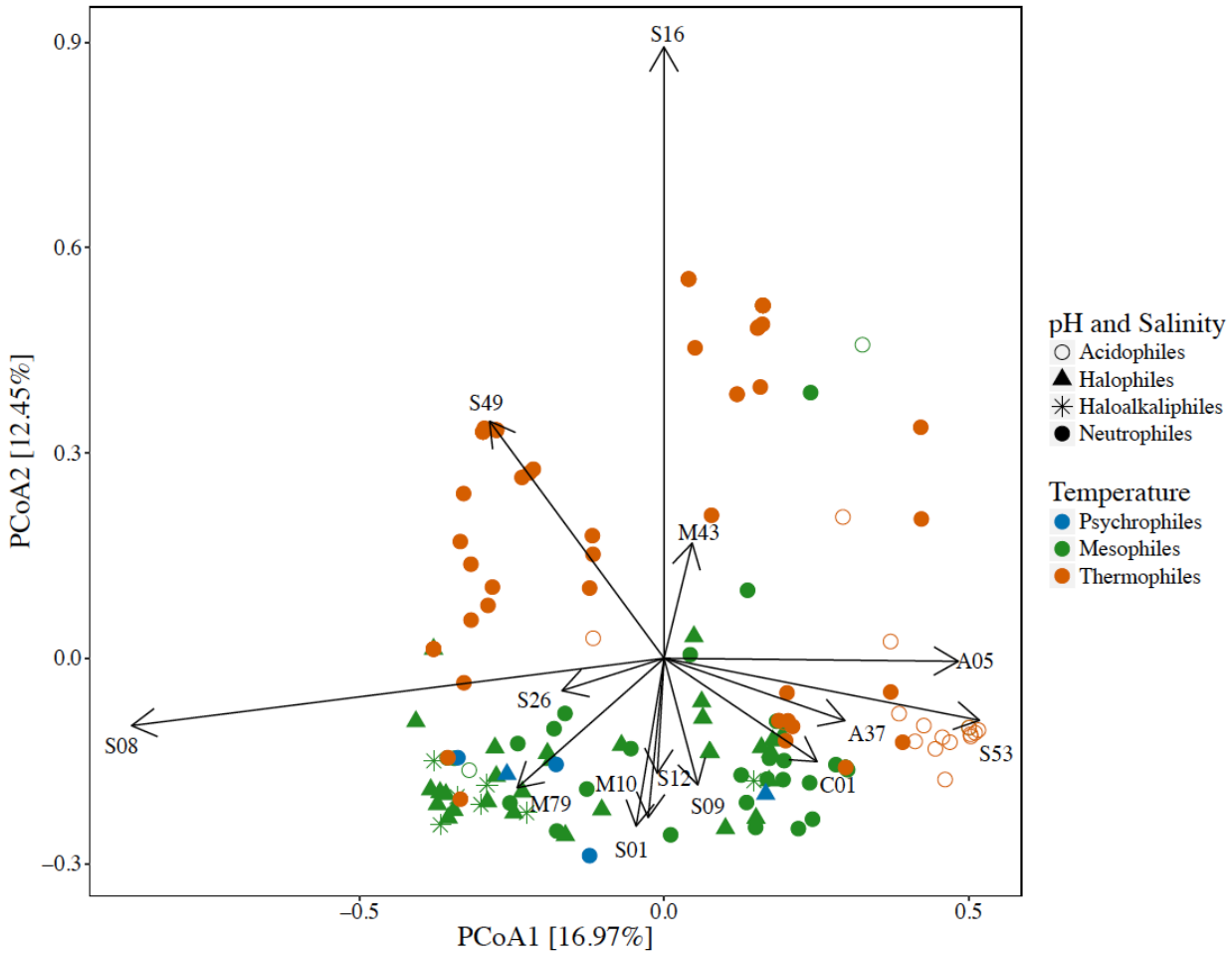


Figure 2.6. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in archaeal genomes. Symbol shapes and colors are coded by reported optimal growth conditions (pH, temperature, and salt concentration). Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). The composition of secreted peptidasegenes of *Archaea* varied significantly based on their optimal growth conditions (p-value < 0.001, F-statistic = 6.11, PERMANOVA).

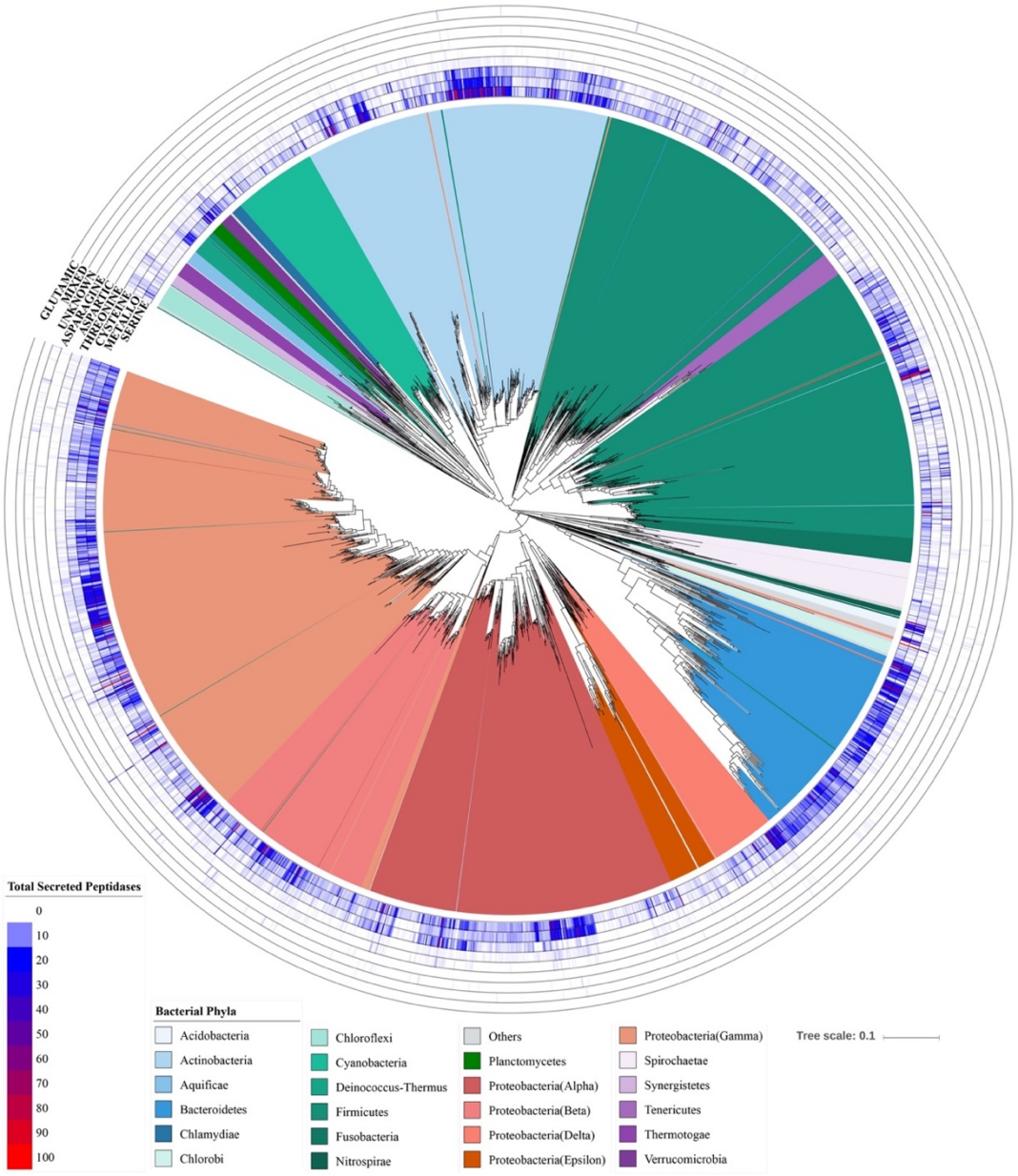
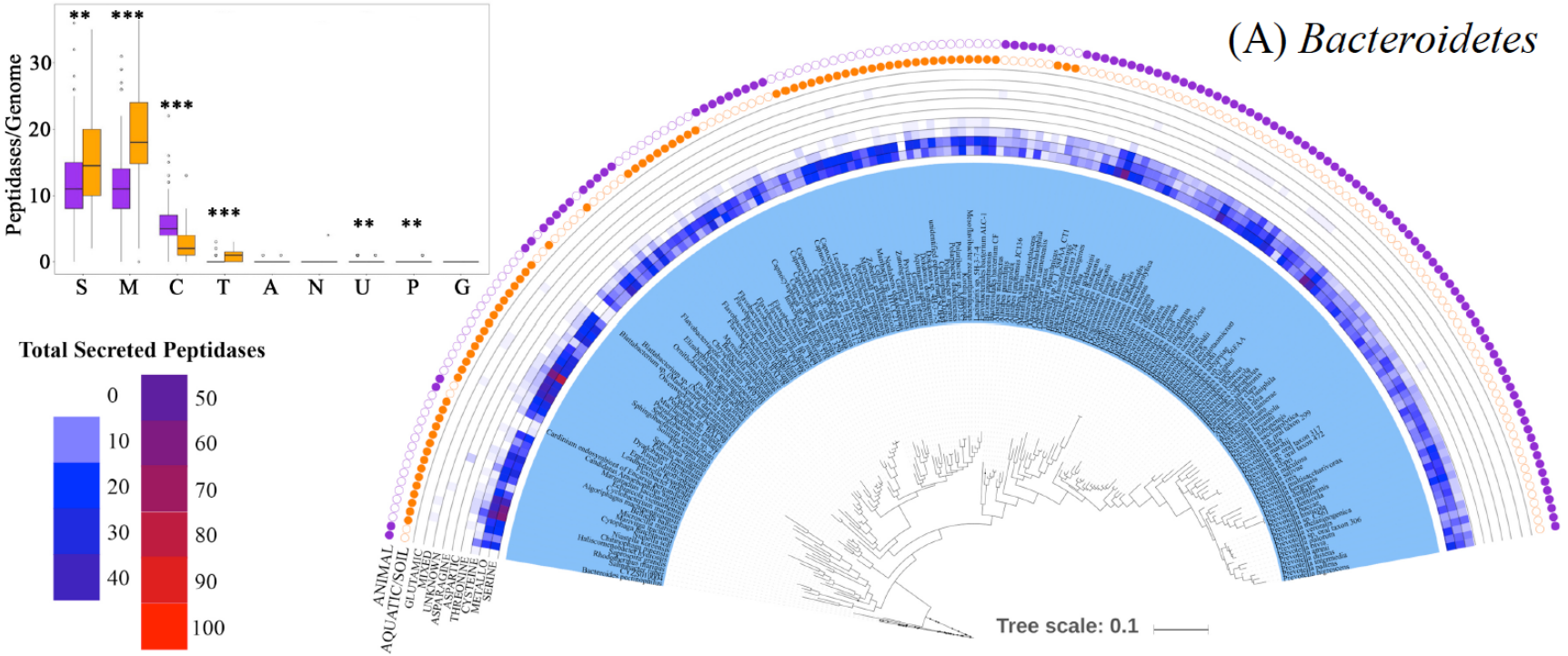


Figure 2.7. Distribution of secreted peptidases super-families across the bacterial phylogenetic tree. Outer tracks show the copy numbers of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered.


(A) Bacteroidetes
Peptidases/Genome
30
20
10
0
**
***
***
***
**
**
S
M
C
T
A
N
U
P
G
Total Secreted Peptidases
0
10
20
30
40
50
60
70
80
90
100
ANIMAL
AQUATIC/SOIL
GLUTAMIC
MIXED
UNKNOWN
ASPARAGINE
ASPARTIC
THREONINE
CYSTEINE
METALLO
SERINE
Tree scale: 0.1

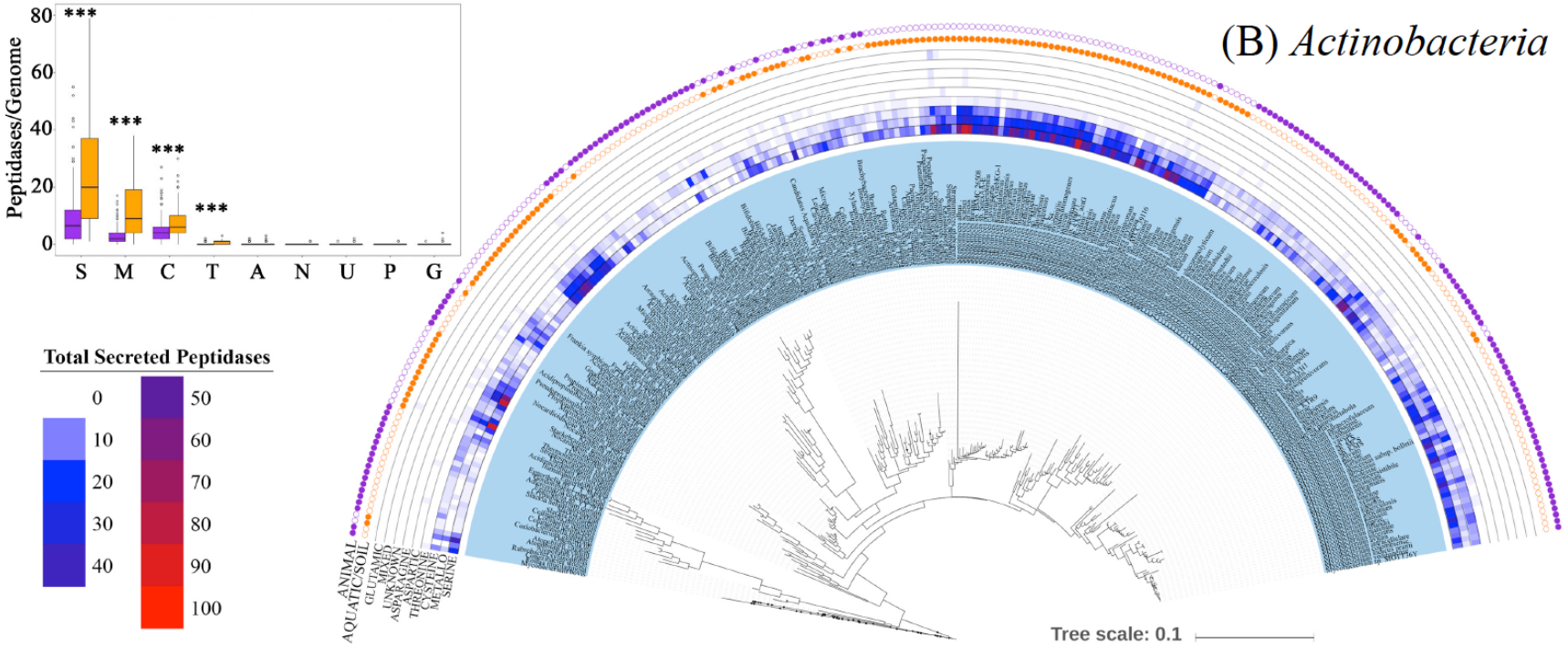


Figure 2.8. Distribution of strain-specific secreted peptidases across (A) *Bacteroidetes* and (B) *Actinobacteria* taxa. The outer two tracks represent the habitat each taxon is associated with (orange for aquatic/soil environment, purple for animal environment). The middle tracks show the copy number of genes from each secreted peptidase super-family in each genome. The figure on the top left corner represent the average numbers of each peptidase super-family (S for serine, M for metallo-, C for cysteine, T for threonine, A for aspartic, N for asparagine, U for unknown, P for mixed and G for glutamic) per genome; purple box plots report mean and standard deviation of

the peptidase content of genomes that are commonly found in animal-associated environments, and orange box plots report mean and standard deviation of the peptidase content of genomes from soil/aquatic environments, with p-values of: *** = 0-0.001, ** = 0.001-0.01, and * = 0.01-0.05.

2.6. Appendix for Chapter 2

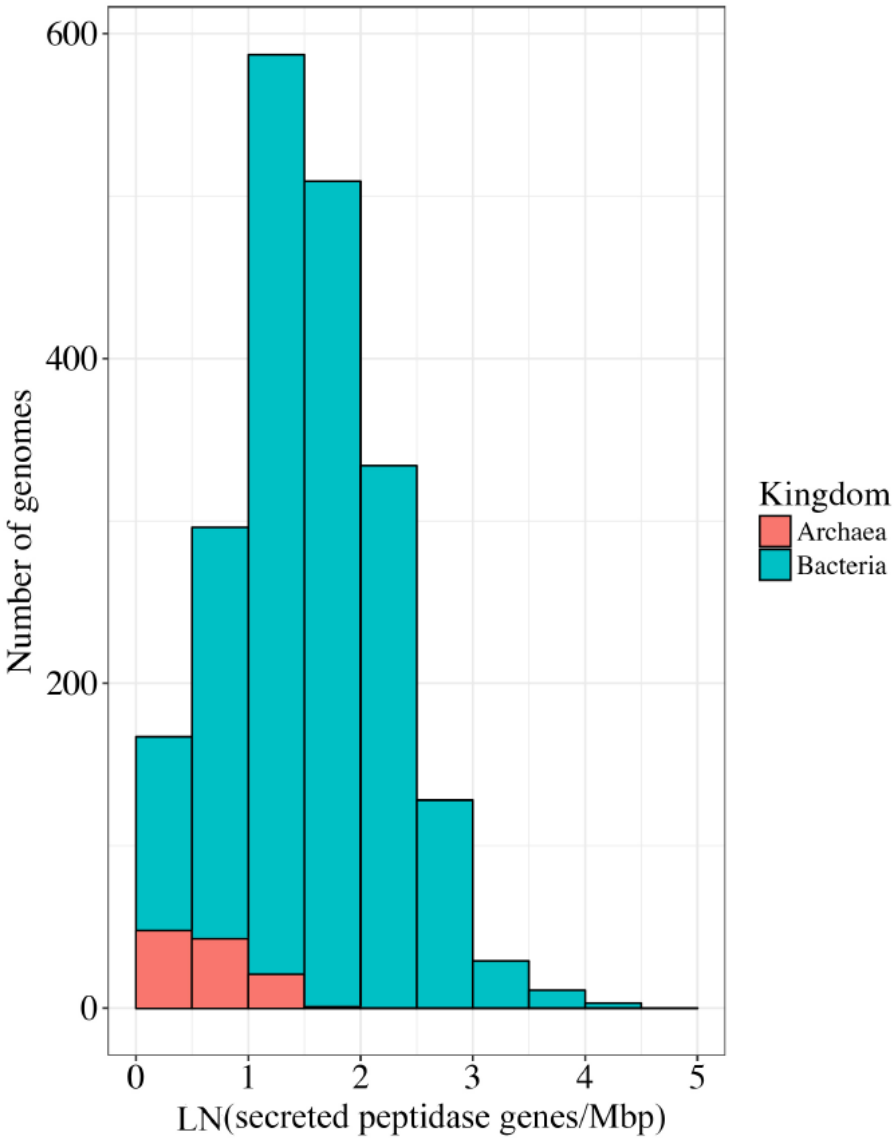


Figure S2.1. Secreted peptidase gene content of 147 archaeal and 2,191 bacterial genomes (genes/mega base pairs (Mb), natural log transformation). *Archaea* in red and *Bacteria* in blue (Welch two-sample t-test, t=-24.0, p-value < 0.001)

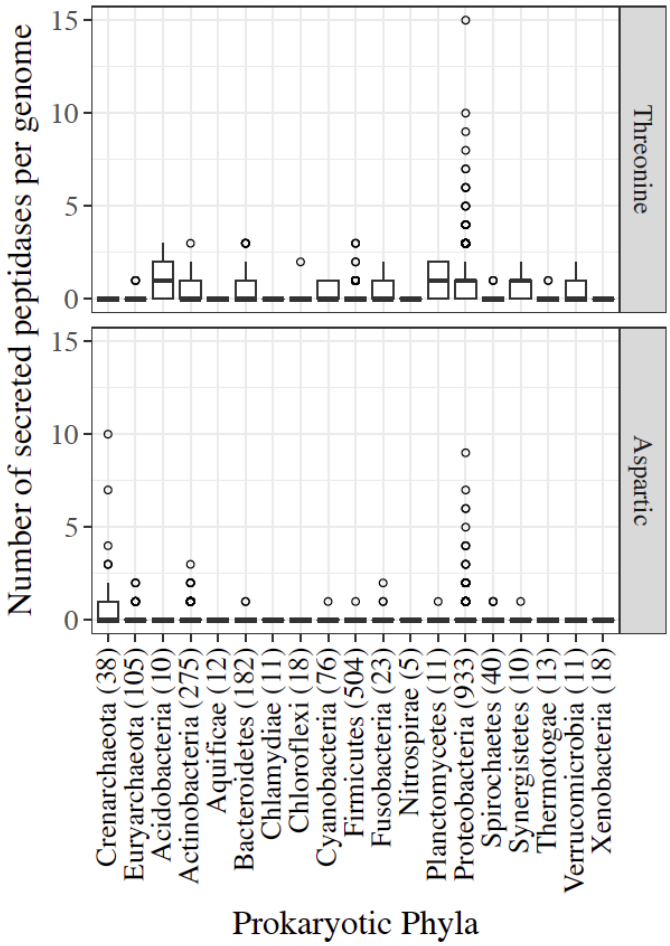


Figure S2.2. Secreted peptidase gene content (genes/genome) of threonine and aspartic peptidases in prokaryotic phyla. The number of analyzed genomes from each prokaryotic phylum is presented next to the phylum names.

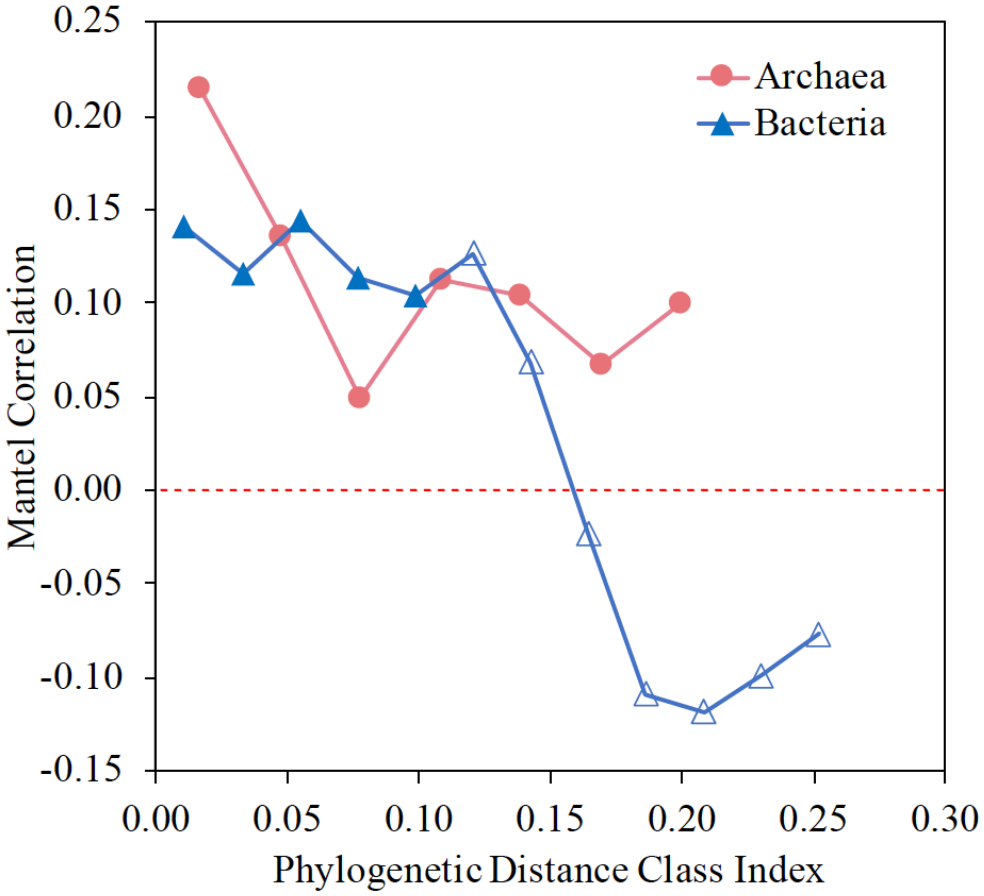


Figure S2.3. Mantel correlogram between phylogenetic distance and secreted protease profile dissimilarities for archaeal and bacterial taxa based on Pearson's product-moment correlations (p-value < 0.05, filled symbols; not significant, open symbols).

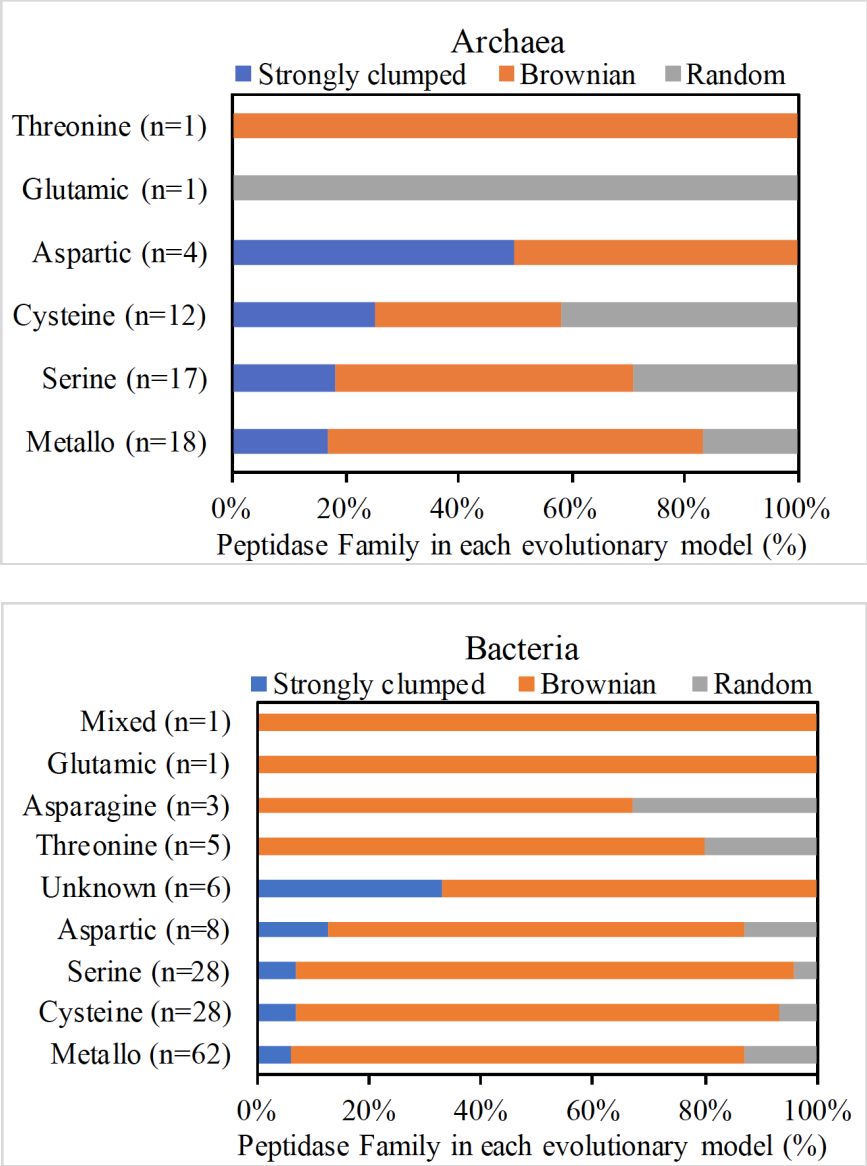


Figure S2.4. Phylogenetic distributions of secreted peptidase families across archaeal and bacterial taxa.

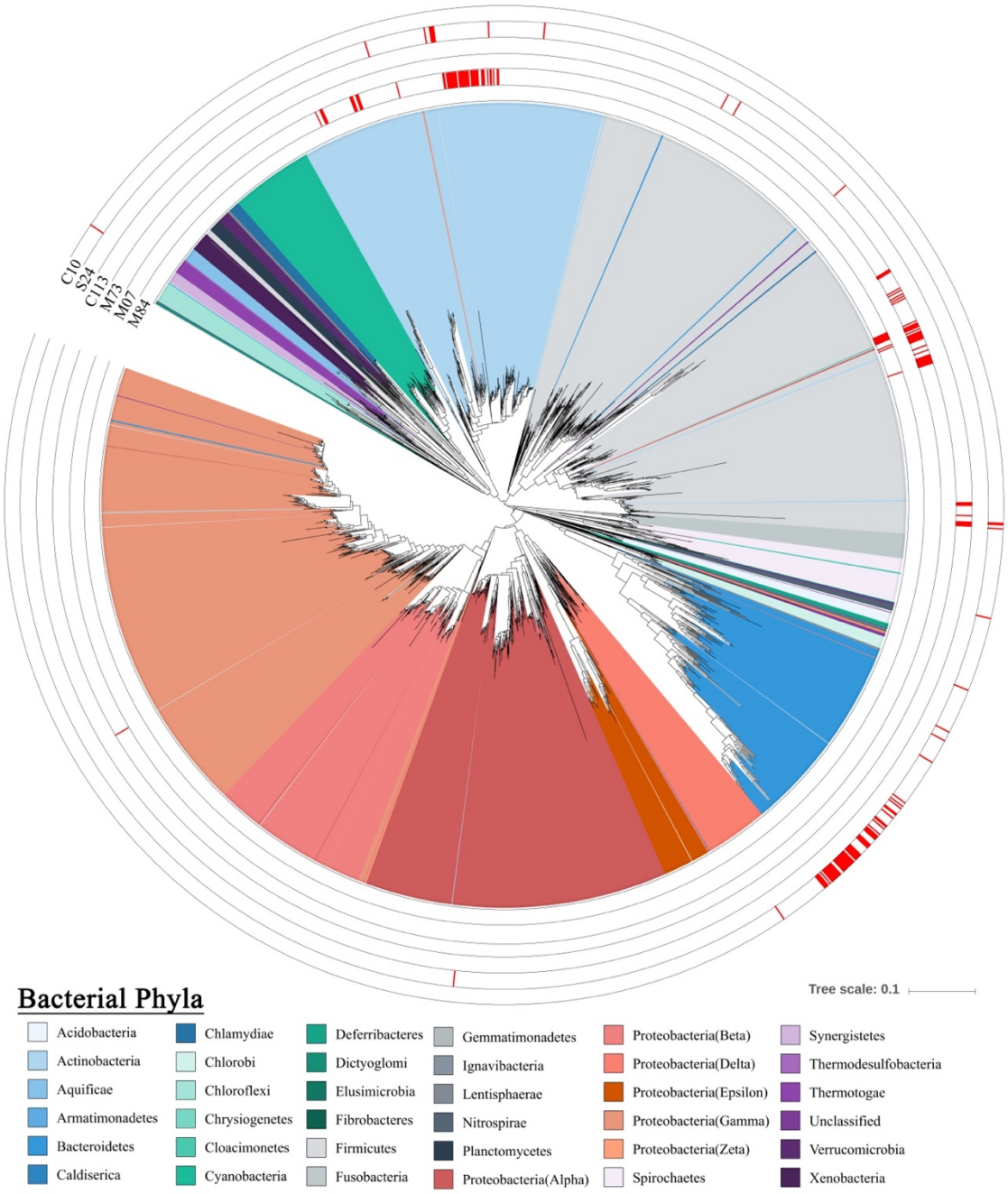


Figure S2.5. Secreted peptidase families showing clumped distributions within bacterial genomes. Outer tracks show the presence/absence of genes from each secreted peptidase family in each genome (white=absence, red=presence). Inner ring colors represent the phylum-level classification of microbial genomes.

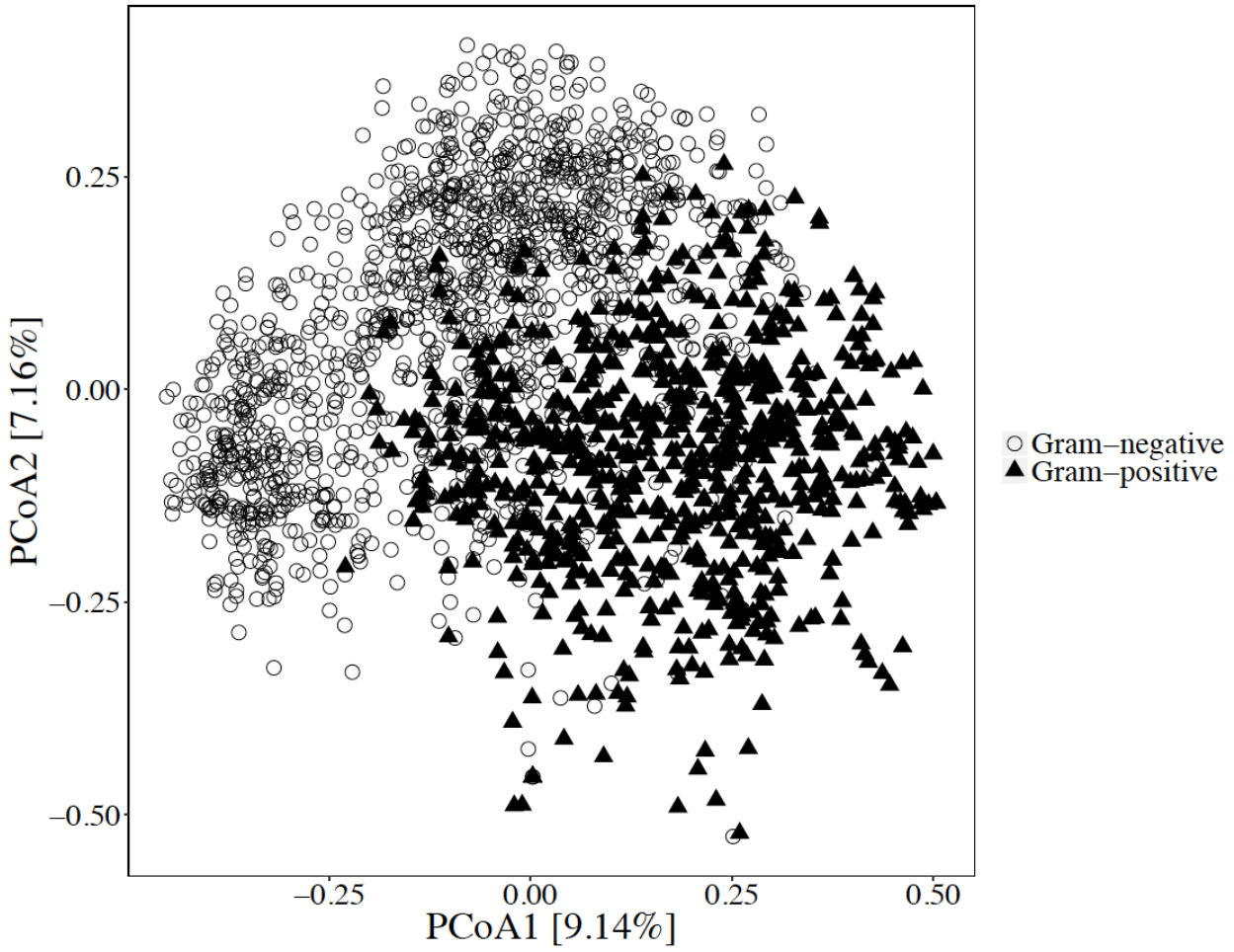


Figure S2.6. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in each genome. Symbol shapes are coded by either Gram-negative and Gram-positive cell wall classification. Significant differences were observed between Gram-positive and Gram-negative bacteria based on the relative abundance of their secreted peptidase families (p-value < 0.001, F-statistic = 193.3, PERMANOVA).

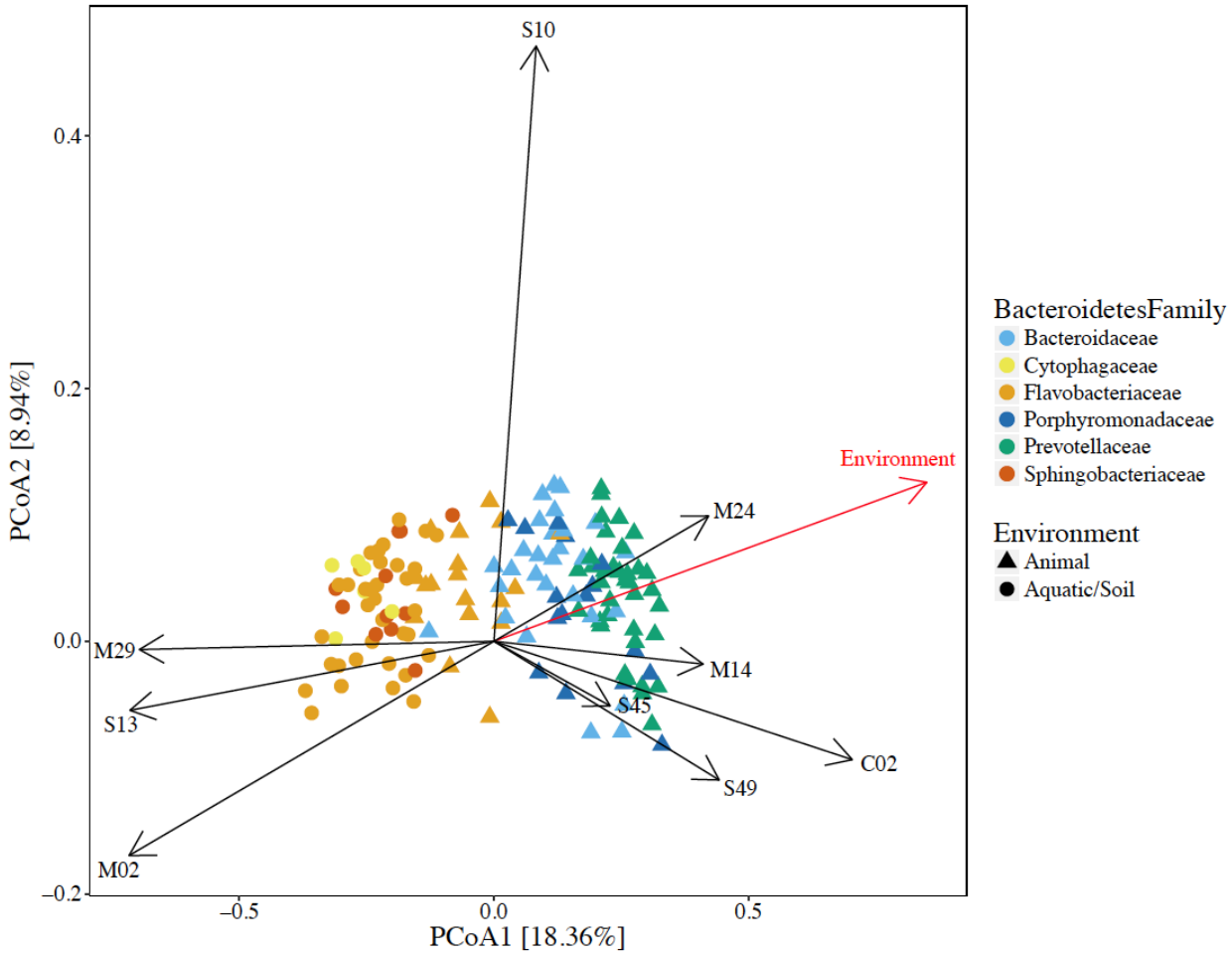


Figure S2.7. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in *Bacteroidetes* genomes. Symbol shapes are coded by the environment where the microorganisms are commonly found (triangle for animal microbiota and circle for aquatic/soil environment); symbol colors represent different taxonomic families. Vector lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). Secreted peptidase profiles were strongly correlated with the environment in which each species was associated with (p-value < 0.001, F-statistic = 34.5, PERMANOVA).

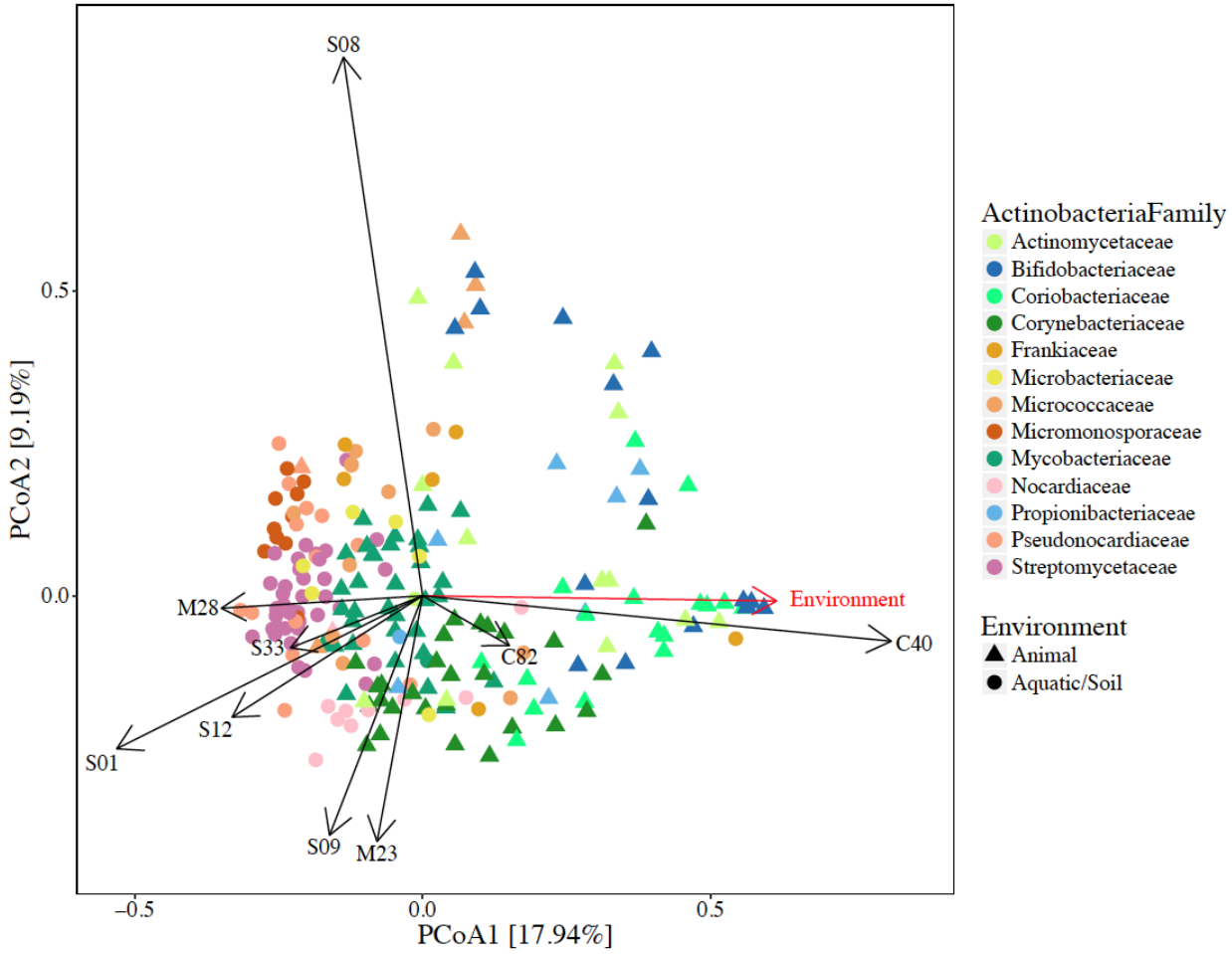


Figure S2.8. Principal coordinate analysis of prokaryotic genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in *Actinobacteria* genomes. Symbol shapes are coded by the environment where the microorganisms are commonly found (triangle for animal microbiota and circle for aquatic/soil environment); symbol colors represent different taxonomic families. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). Ecological niches, on the other hand, also played a strong role in differentiating this function among Actinobacteria (p-value < 0.001, F-statistic = 26.4, PERMANOVA).

Table S2.1. Differences of secreted peptidases among bacterial phyla using Tukey's HSD analysis.

	Acidobacteria	*Actinobacteria*	*Aquificae*	*Bacteroidetes*	*Chlamydiae*	*Chlorobi*	*Chloroflexi*	*Cyanobacteria*	*Firmicutes*	*Fusobacteria*	*Planctomycetes*	*Proteobacteria*	*Spirochaetes*	*Thermotogae*	*Verrucomicrobia*	*Xenobacteria*
Mean number of secreted peptidases	55.3	31.7	3.4	36.3	8.4	7.1	10.4	7.9	17.9	5.7	29.5	28.4	7.4	2.7	20.7	17.1
Acidobacteria			***		*	*	*	***		***			***	***		
Actinobacteria			***	**	*	**	**	***	***	***			***	***		
Aquificae	***	***		***					***		***	***			**	**
Bacteroidetes		**	***		***	***	***	***	***	***		***	***	***		
Chlamydiae	*	*		***												
Chlorobi	*	**		***								*				
Chloroflexi	*	**		***								*				
Cyanobacteria	***	***		***					***		**	***		*		
Firmicutes		***	***	***				***		***		***	***	***		
Fusobacteria	***	***		***					***		***	***			*	**
Planctomycetes			***					**		***			***	***		
Proteobacteria			***	***		*	*	***	***	***			***	***		
Spirochaetes	***	***		***					***		***	***				
Thermotogae	***	***		***				*	***		***	***			***	***
Verrucomicrobia			**							*				***		
Xenobacteria			**							**				***		

p-value: (***) 0-0.001, (**) 0.001-0.01, (*) 0.01-0.05, () > 0.05

Table S2.2. Phylogenetic signal strength of secreted peptidase families across 2,191 bacterial genomes. Significance of clustering is based on the Fritz and Purvis index (D) of each peptidase family trait (presence or absence of genes). Estimated D value defined whether the secreted peptidase distribution would follow "strongly clumped" ($D \leq 0$), or "Brownian-like evolutionary" ($0 < D < 1$) or "Random" distribution ($D \geq 1$).

Peptidase Family	Estimated D	Probability of E(D) resulting from no (random) phylogenetic structure	Probability of E(D) resulting from Brownian phylogenetic structure	Genomes
M07	-0.2259344	0	0.883	47
M73	-0.1694997	0	0.783	32
M84	-0.2318844	0	0.732	10
C113	-0.0912123	0	0.61	11
S24	-0.0186868	0	0.551	9
C10	-0.00685428	0	0.517	59
S46	0.03560422	0	0.401	203
C44	0.1183377	0	0.401	8
M93	0.06470617	0	0.371	51
S28	0.08069442	0	0.316	62
M17	0.09688335	0	0.308	47
M34	0.1920949	0	0.271	16
C15	0.2146679	0	0.262	13
M74	0.08196468	0	0.249	215
C66	0.2367091	0	0.244	12
C02	0.2966722	0	0.244	9
M02	0.108701	0	0.219	86
C25	0.1154511	0	0.213	84
G01	0.3256029	0	0.193	10
U62	0.2398793	0	0.167	26
M88	0.3500109	0	0.161	12
S54	0.3590698	0	0.125	15
M11	0.325439	0	0.115	20
M01	0.09388175	0	0.112	414
S55	0.1978025	0	0.103	59
M19	0.1688879	0	0.078	143
C69	0.1606719	0	0.05	203

M13	0.150051	0	0.048	288
S16	0.2327544	0	0.034	116
N04	0.3161909	0	0.025	55
M86	0.542897	0	0.018	18
S37	0.4474473	0	0.016	24
M28	0.1511733	0	0.015	582
C01	0.1945182	0	0.014	235
A32	0.2915281	0	0.014	85
M35	0.5759979	0	0.014	17
M72	0.4053593	0	0.011	49
M16	0.144704	0	0.009	925
M06	0.2247751	0	0.008	229
M75	0.2263001	0	0.008	210
M66	0.4520076	0	0.008	31
M04	0.2083994	0	0.007	256
S10	0.2441075	0	0.007	180
T02	0.2538085	0	0.007	113
C40	0.1503968	0	0.005	1166
A26	0.3843456	0	0.005	57
S14	0.5553352	0	0.005	23
S06	0.5906663	0	0.004	19
C26	0.6459709	0	0.004	18
M20	0.2435776	0	0.002	277
M49	0.5847863	0	0.002	25
S11	0.2053199	0	0.001	1465
S53	0.3432167	0	0.001	102
M09	0.3744992	0	0.001	88
M64	0.4271057	0	0.001	68
M30	0.4936028	0	0.001	35
A24	0.6858283	0	0.001	24
T03	0.2781851	0	0	772
C82	0.2804334	0	0	910
S13	0.2856551	0	0	649
M03	0.290369	0	0	252
S41	0.3053146	0	0	1163
M14	0.3139296	0	0	335
M38	0.3327629	0	0	304
M96	0.3777077	0	0	192
S12	0.4021022	0	0	1002

S01	0.4097399	0	0	1422
S15	0.4116812	0	0	366
S08	0.4174785	0	0	991
M43	0.4176288	0	0	101
S45	0.4242641	0	0	222
C60	0.4392742	0	0	146
C39	0.4397784	0	0	99
M48	0.4500771	0	0	703
M97	0.4681868	0	0	52
M41	0.4705892	0	0	100
S09	0.4849367	0	0	894
M61	0.493181	0	0	396
C110	0.4960267	0	0	83
M10	0.5026735	0	0	100
C59	0.5149821	0	0	200
S66	0.538241	0	0	84
M15	0.5439112	0	0	543
S49	0.5541239	0	0	269
C14	0.5754452	0	0	110
M26	0.5765666	0	0	41
C93	0.5819775	0	0	73
M24	0.5975717	0	0	74
M36	0.5997786	0	0	64
C13	0.6169344	0	0	29
S33	0.6301359	0	0	426
M23	0.6359832	0	0	1360
M50	0.6435465	0	0	36
S51	0.6527886	0	0	63
C56	0.6592502	0	0	107
A08	0.6680078	0	0	63
C11	0.6738293	0	0	73
C83	0.6841772	0	0	31
M12	0.6982088	0	0	58
U69	0.7769681	0	0	79
S26	0.7917711	0	0	108
M79	0.729761	0.001	0	32
P01	0.7922505	0.003	0	33
C51	0.5217914	0.005	0.05	10
M05	-0.2278856	0.009	0.651	3

T05	0.3934865	0.009	0.216	6
A28	0.7261397	0.009	0	20
M57	0.6577306	0.021	0.015	11
A01	-0.339028	0.029	0.689	2
C47	0.6565088	0.034	0.028	10
M81	0.6174429	0.035	0.065	8
U72	-1.742503	0.052	0.848	1
M60	0.7914669	0.066	0.001	14
U73	0.4583012	0.069	0.233	4
M08	0.7195649	0.116	0.029	7
M18	0.133928	0.128	0.476	2
S73	0.1524141	0.142	0.462	2
M98	0.8741101	0.192	0	13
S59	-0.8006927	0.193	0.672	1
M54	0.7690595	0.243	0.051	5
C19	0.6379069	0.262	0.163	3
U32	-0.04075472	0.316	0.51	1
T07	0.6669495	0.356	0.236	2
N11	0.1935319	0.36	0.428	1
A02	0.1646482	0.371	0.44	1
U74	0.8831385	0.388	0.03	5
M76	0.4561855	0.416	0.392	1
C46	0.7938488	0.47	0.301	1
M56	1.001749	0.511	0	12
N10	1.007016	0.534	0.067	3
M55	1.10351	0.553	0.245	1
S50	1.221841	0.585	0.198	1
M82	1.068431	0.606	0.006	5
M32	1.675042	0.682	0.143	1
M42	1.127865	0.754	0	11
C89	1.484572	0.756	0.028	2
C45	1.096967	0.756	0	15
M44	2.734743	0.819	0.077	1
A11	3.046838	0.874	0.018	1
M78	3.816493	0.915	0.027	1
T01	6.425802	0.982	0	1
M85	3.134598	0.994	0	2

Table S2.3. Phylogenetic signal strength of secreted peptidase families across 147 archaeal genomes. Significance of clustering is based on the Fritz and Purvis index (D) of each peptidase family trait (presence or absence of genes). Estimated D value defined whether the secreted peptidase distribution would follow "strongly clumped" ($D \leq 0$), or "Brownian-like evolutionary" ($0 < D < 1$) or "random" distribution ($D \geq 1$).

Peptidase Family	Estimated D	Probability of E(D) resulting from no (random) phylogenetic structure	Probability of E(D) resulting from Brownian phylogenetic structure	Genomes
C56	-0.8365722	0	0.915	4
A05	-0.3056731	0	0.871	16
S53	-0.293705	0	0.862	18
S16	-0.1277893	0	0.751	44
S12	0.02741457	0	0.447	16
M10	0.04843905	0	0.432	16
M01	0.1706628	0	0.336	8
M79	0.2243264	0	0.179	20
C01	0.3023148	0	0.086	26
S01	0.3590775	0	0.069	20
S49	0.4882171	0	0.004	48
S08	0.4885586	0	0.002	70
M28	0.06883237	0.002	0.478	5
S09	0.545091	0.002	0.012	20
M67	-1.42327	0.004	0.867	2
A37	0.245117	0.005	0.29	6
M48	0.5998375	0.026	0.031	10
C39	0.3796129	0.028	0.241	5
S33	0.4785832	0.028	0.131	7
A22	0.6635692	0.044	0.014	11
A31	-1.788827	0.054	0.911	1
M84	-0.00896638	0.09	0.566	2
M13	0.4058199	0.115	0.317	3
M82	-0.02828574	0.125	0.554	2
C69	0.5660381	0.125	0.185	4
S26	0.7652234	0.125	0.006	11
C110	0.2406773	0.177	0.416	2

T03	0.6611962	0.18	0.145	4
M03	0.3919442	0.244	0.373	2
C19	-0.5283647	0.296	0.558	1
C51	-0.6074144	0.302	0.559	1
S60	-0.4075937	0.321	0.527	1
M43	0.1791917	0.386	0.42	1
S41	0.2983447	0.39	0.381	1
S54	0.7255929	0.406	0.237	2
M30	0.8668889	0.421	0.092	3
M14	0.9940787	0.482	0.002	7
M38	0.6596704	0.496	0.309	1
M09	0.8770507	0.498	0.278	1
S45	1.033513	0.557	0.015	5
S24	1.318157	0.569	0.23	1
M26	1.081546	0.597	0.008	5
S13	1.134787	0.602	0.043	3
C26	1.127768	0.618	0.029	3
M61	1.651654	0.638	0.17	1
C11	1.22681	0.642	0.067	2
M54	1.262134	0.659	0.048	2
G01	2.230843	0.71	0.106	1
C25	2.719368	0.784	0.063	1
C14	1.495708	0.839	0.003	3
C40	3.340966	0.864	0.041	1
S15	1.629052	0.877	0.004	3
S59	3.831459	0.913	0.007	1

Table S2.4. Statistical differences of secreted peptidases between *Bacteroidetes* families using Tukey's HSD analysis.

	Bacteroidaceae	*Cytophagaceae*	*Flavobacteriaceae*	*Porphyromonadaceae*	*Prevotellaceae*	*Sphingobacteriaceae*
Mean. Secreted peptidases	36.91	54.86	33.67	34.94	29.42	54.73
Bacteroidaceae						
Cytophagaceae			*		*	
Flavobacteriaceae		*				**
Porphyromonadaceae						
Prevotellaceae		*				**
Sphingobacteriaceae			**		**	
p-value: (***) 0-0.001, (**) 0.001-0.01, (*) 0.01-0.05, () > 0.05						

Table S2.5. Statistical differences of secreted peptidases between *Actinobacteria* families using Tukey's HSD analysis.

	Actinomycetaceae	*Bifidobacteriaceae*	*Coriobacteriaceae*	*Frankiaceae*	*Microbacteriaceae*	*Micrococcaceae*	*Micromonosporaceae*	*Mycobacteriaceae*	*Nocardiaceae*	*Propionibacteriaceae*	*Pseudonocardiaceae*	*Streptomycetaceae*
Mean. Secreted peptidases	9.42	7.25	7.22	6.67	17.50	17.20	64.70	34.69	43.30	12.13	50.13	72.38
Actinomycetaceae							***	*	**		***	***
Bifidobacteriaceae							***	***	***		***	***
Coriobacteriaceae							***	***	***		***	***
Frankiaceae							***		*		***	***
Microbacteriaceae							***				*	***
Micrococcaceae							***				***	***
Micromonosporaceae	***	***	***	***	***	***		**		***		
Mycobacteriaceae	*	***	***				**					***
Nocardiaceae	**	***	***	*								**
Propionibacteriaceae							***				**	***
Pseudonocardiaceae	***	***	***	***	*	***				**		*
Streptomycetaceae	***	***	***	***	***	***		***	**	***	*	
p-value: (***) 0-0.001, (**) 0.001-0.01, (*) 0.01-0.05, () > 0.05												

2.7. Reference

Abdallah, A. M., Pittius, N. C. G. van, Champion, P. A. D., Cox, J., Luirink, J., Vandenbroucke-Grauls, C. M. J. E., et al. (2007). Type VII secretion — *Mycobacteria* show the way. *Nat. Rev. Microbiol.* 5, 883–891. doi:10.1038/nrmicro1773.

Allison, S. D., Gartner, T. B., Mack, M. C., McGuire, K., and Treseder, K. (2010). Nitrogen alters carbon dynamics during early succession in boreal forest. *Soil Biol. Biochem. 42 1157-1164* 42, 1157–1164.

Arnosti, C. (2015). Contrasting patterns of peptidase activities in seawater and sediments: An example from Arctic fjords of Svalbard. *Mar. Chem.* 168, 151–156. doi:10.1016/j.marchem.2014.09.019.

Bach, H.-J., and Munch, J. C. (2000). Identification of bacterial sources of soil peptidases. *Biol. Fertil. Soils* 31, 219–224. doi:10.1007/s003740050648.

Barberán, A., Velazquez, H. C., Jones, S., and Fierer, N. (2017). Hiding in plain sight: Mining bacterial species records for phenotypic trait information. *mSphere* 2, e00237-17. doi:10.1128/mSphere.00237-17.

Berlemont, R., and Martiny, A. C. (2013). Phylogenetic distribution of potential cellulases in *Bacteria*. *Appl. Environ. Microbiol.* 79, 1545–1554. doi:10.1128/AEM.03305-12.

Biswas, T., Small, J., Vandal, O., Odaira, T., Deng, H., Ehrt, S., et al. (2010). Structural insight into serine protease rv3671c that protects *M. tuberculosis* from oxidative and acidic stress. *Structure* 18, 1353–1363. doi:10.1016/j.str.2010.06.017.

Botella, H., Vaubourgeix, J., Lee, M. H., Song, N., Xu, W., Makinoshima, H., et al.

(2017). *Mycobacterium tuberculosis* protease MarP activates a peptidoglycan hydrolase during acid stress. *EMBO J.* 36, 536–548. doi:10.15252/embj.201695028.

Brown, L., Wolf, J. M., Prados-Rosales, R., and Casadevall, A. (2015). Through the wall: Extracellular vesicles in Gram-positive bacteria, *Mycobacteria* and *Fungi*. *Nat. Rev. Microbiol.* 13, 620–630. doi:10.1038/nrmicro3480.

Bryson, S., Li, Z., Chavez, F., Weber, P. K., Pett-Ridge, J., Hettich, R. L., et al. (2017). Phylogenetically conserved resource partitioning in the coastal microbial loop. *ISME J.* 11, 2781–2792. doi:10.1038/ismej.2017.128.

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

Caldwell, B. A. (2005). Enzyme activities as a component of soil biodiversity: A review. *Pedobiologia* 49, 637–644. doi:10.1016/j.pedobi.2005.06.003.

Chater, K. F., Biró, S., Lee, K. J., Palmer, T., and Schrempf, H. (2010). The complex extracellular biology of *Streptomyces*. *FEMS Microbiol. Rev.* 34, 171–198. doi:10.1111/j.1574-6976.2009.00206.x.

Choi, I.-G., and Kim, S.-H. (2007). Global extent of horizontal gene transfer. *Proc. Natl. Acad. Sci.* 104, 4489–4494. doi:10.1073/pnas.0611557104.

Chróst, R. J. (1991). "Environmental control of the synthesis and activity of aquatic microbial ectoenzymes," in *Microbial Enzymes in Aquatic Environments* Brock/Springer Series in Contemporary Bioscience. (Springer, New York, NY),

29–59. doi:10.1007/978-1-4612-3090-8_3.

Dallas, D. C., Underwood, M. A., Zivkovic, A. M., and German, J. B. (2012). Digestion of protein in premature and term infants. *J. Nutr. Disord. Ther.* 2, 1–9. doi:10.4172/2161-0509.1000112.

DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M., et al. (2006). NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* 34, W394–W399. doi:10.1093/nar/gkl244.

Duarte, A. S., Correia, A., and Esteves, A. C. (2016). Bacterial collagenases – A review. *Crit. Rev. Microbiol.* 42, 106–126. doi:10.3109/1040841X.2014.904270.

Felsenstein, J. (1989). PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5, 164–166.

Fritz, S. A., and Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conserv. Biol. J. Soc. Conserv. Biol.* 24, 1042–1051. doi:10.1111/j.1523-1739.2010.01455.x.

Gagneux, S. (2018). Ecology and evolution of *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* 16, 202–213. doi:10.1038/nrmicro.2018.8.

Gazi, M. I., Cox, S. W., Clark, D. T., and Eley, B. M. (1997). Characterization of protease activities in *Capnocytophaga* spp., *Porphyromonas gingivalis*, *Prevotella* spp., *Treponema denticola* and *Actinobacillus actinomycetemcomitans*. *Oral Microbiol. Immunol.* 12, 240–248. doi:10.1111/j.1399-302X.1997.tb00386.x.

Geisseler, D., and Horwath, W. R. (2008). Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon.

Soil Biol. Biochem. 40, 3040–3048. doi:10.1016/j.soilbio.2008.09.001.

Geisseler, D., Horwath, W. R., Joergensen, R. G., and Ludwig, B. (2010). Pathways of nitrogen utilization by soil microorganisms – A review. *Soil Biol. Biochem.* 42, 2058–2067. doi:10.1016/j.soilbio.2010.08.021.

Gibson, S. A. W., and Macfarlane, G. T. (1988a). Characterization of proteases formed by *Bacteroides fragilis*. *Microbiology* 134, 2231–2240. doi:10.1099/00221287-134-8-2231.

Gibson, S. A. W., and Macfarlane, G. T. (1988b). Studies on the proteolytic activity of *Bacteroides fragilis*. *Microbiology* 134, 19–27. doi:10.1099/00221287-134-1-19.

Goberna, M., and Verdú, M. (2016). Predicting microbial traits with phylogenies. *ISME J.* 10, 959–967. doi:10.1038/ismej.2015.171.

Griswold, K. E., White, B. A., and Mackie, R. I. (1999). Diversity of extracellular proteolytic activities among *Prevotella* species from the rumen. *Curr. Microbiol.* 39, 187–194. doi:10.1007/s002849900443.

Gupta, R., and Ramnani, P. (2006). Microbial keratinases and their prospective applications: An overview. *Appl. Microbiol. Biotechnol.* 70, 21. doi:10.1007/s00253-005-0239-8.

Hartley, B. S. (1960). Proteolytic enzymes. *Annu. Rev. Biochem.* 29, 45–72. doi:10.1146/annurev.bi.29.070160.000401.

Häse, C. C., and Finkelstein, R. A. (1993). Bacterial extracellular zinc-containing metalloproteases. *Microbiol. Rev.* 57, 823–837.

Heyndrickx, G. V. (1963). Further investigations on the enzymes in human milk. *Pediatrics* 31, 1019–1023.

Hoppe, H.-G. (1991). “Microbial extracellular enzyme activity: A new key parameter in aquatic ecology,” in *Microbial Enzymes in Aquatic Environments* Brock/Springer Series in Contemporary Bioscience. (Springer, New York, NY), 60–83. doi:10.1007/978-1-4612-3090-8_4.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *J. Mol. Biol.* 338, 1027–1036. doi:10.1016/j.jmb.2004.03.016.

Käll, L., Krogh, A., and Sonnhammer, E. L. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* 35, W429–W432. doi:10.1093/nar/gkm256.

Katsuji, W., Susumu, A., and Koichi, H. (1994). Evaluation of extracellular protease activities of soil bacteria. *Soil Biol. Biochem.* 26, 479–482. doi:10.1016/0038-0717(94)90180-5.

Kolton, M., Sela, N., Elad, Y., and Cytryn, E. (2013). Comparative genomic analysis indicates that niche adaptation of terrestrial *Flavobacteria* is strongly linked to plant glycan metabolism. *PLOS ONE* 8, e76704. doi:10.1371/journal.pone.0076704.

Kugadas, A., Lamont, E. A., Bannantine, J. P., Shoyama, F. M., Brenner, E., Janagama, H. K., et al. (2016). A *Mycobacterium avium* subsp. *paratuberculosis* predicted serine protease is associated with acid stress and intraphagosomal survival. *Front. Cell. Infect. Microbiol.* 6. doi:10.3389/fcimb.2016.00085.

Kumar, C. G., and Takagi, H. (1999). Microbial alkaline proteases: From a bioindustrial viewpoint. *Biotechnol. Adv.* 17, 561–594. doi:10.1016/S0734-9750(99)00027-0.

Landi, L., Renella, G., Giagnoni, L., and Nannipieri, P. (2011). "Activities of proteolytic enzymes," in *Methods of Soil Enzymology* (Madison, WI), 247–260. Available at: https://dl.sciencesocieties.org/publications/books/abstracts/sssabookseries/methodsofsoilen/247?access=0&view=pdf [Accessed August 9, 2018].

Lange, L., Huang, Y., and Busk, P. K. (2016). Microbial decomposition of keratin in nature—A new hypothesis of industrial relevance. *Appl. Microbiol. Biotechnol.* 100, 2083–2096. doi:10.1007/s00253-015-7262-1.

Lawrence, J. G., and Hendrickson, H. (2003). Lateral gene transfer: When will adolescence end? *Mol. Microbiol.* 50, 739–749. doi:10.1046/j.1365-2958.2003.03778.x.

Lee, J.-H., and O'Sullivan, D. J. (2010). Genomic Insights into *Bifidobacteria*. *Microbiol. Mol. Biol. Rev.* 74, 378–416. doi:10.1128/MMBR.00004-10.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi:10.1093/nar/gkw290.

Loesche, W. J. (1988). The role of *Spirochetes* in periodontal disease. *Adv. Dent. Res.* 2, 275–283. doi:10.1177/08959374880020021201.

Macfarlane, G. T., Cummings, J. H., and Allison, C. (1986). Protein degradation by human intestinal bacteria. *Microbiology* 132, 1647–1656. doi:10.1099/00221287-132-6-1647.

Mallorquí-Fernández, N., Manandhar, S. P., Mallorquí-Fernández, G., Usón, I., Wawrzonek, K., Kantyka, T., et al. (2008). A new autocatalytic activation

mechanism for cysteine proteases revealed by *Prevotella Intermedia* Interpain A. *J. Biol. Chem.* 283, 2871–2882. doi:10.1074/jbc.M708481200.

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

Mastronunzio, J. E., Huang, Y., and Benson, D. R. (2009). Diminished Exoproteome of Frankia spp. in Culture and Symbiosis. *Appl Env. Microbiol* 75, 6721–6728. doi:10.1128/AEM.01559-09.

Mooshammer, M., Wanek, W., Hämmerle, I., Fuchslueger, L., Hofhansl, F., Knoltsch, A., et al. (2014). Adjustment of microbial nitrogen use efficiency to carbon:nitrogen imbalances regulates soil nitrogen cycling. *Nat. Commun.* 5, 3694. doi:10.1038/ncomms4694.

Mori, H., and Ito, K. (2001). The Sec protein-translocation pathway. *Trends Microbiol.* 9, 494–500.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., et al. (2017). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 45, D446–D456. doi:10.1093/nar/gkw992.

Nagata, T., Fukuda, R., Koike, I., Kogure, K., and Kirchman, D. L. (1998). Degradation by *Bacteria* of membrane and soluble protein in seawater. *Aquat. Microb. Ecol.* 14, 29–37. doi:10.3354/ame014029.

Nakjang, S., Ndeh, D. A., Wipat, A., Bolam, D. N., and Hirt, R. P. (2012). A novel extracellular metallopeptidase domain shared by animal host-associated mutualistic and pathogenic microbes. *PLOS ONE* 7, e30287. doi:10.1371/journal.pone.0030287.

Norman, J. S., and Friesen, M. L. (2017). Complex N acquisition by soil diazotrophs: How the ability to release exoenzymes affects N fixation by terrestrial free-living diazotrophs. *ISME J.* 11, 315–326. doi:10.1038/ismej.2016.127.

Nunn, B. L., Norbeck, A., and Keil, R. G. (2003). Hydrolysis patterns and the production of peptide intermediates during protein degradation in marine systems. *Mar. Chem.* 83, 59–73. doi:10.1016/S0304-4203(03)00096-3.

Obayashi, Y., and Suzuki, S. (2008). Occurrence of exo- and endopeptidases in dissolved and particulate fractions of coastal seawater. *Aquat. Microb. Ecol.* 50, 231–237. doi:10.3354/ame01169.

Ohol, Y. M., Goetz, D. H., Chan, K., Shiloh, M. U., Craik, C. S., and Cox, J. S. (2010). *Mycobacterium tuberculosis* Mycp1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* 7, 210–220. doi:10.1016/j.chom.2010.02.006.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). *vegan: Community Ecology Package*. Available at: https://CRAN.R-project.org/package=vegan.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Issac, N., et al. (2018). caper: Comparative Analyses of Phylogenetics and Evolution in R. Available at: ftp://mirror.ac.za/cran/web/packages/caper/vignettes/caper.pdf.

Page, M. J., and Cera, E. D. (2008). Evolution of peptidase diversity. *J. Biol. Chem.* 283, 30010–30014. doi:10.1074/jbc.M804650200.

Pantoja, S., and Lee, C. (1999). Peptide decomposition by extracellular hydrolysis in coastal seawater and salt marsh sediment. *Mar. Chem.* 63, 273–291.

doi:10.1016/S0304-4203(98)00067-X.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinforma. Oxf. Engl.* 20, 289–290.

Pawlowski, K., and Bisseling, T. (1996). Rhizobial and actinorhizal symbioses: What are the shared features? *Plant Cell* 8, 1899–1913. doi:10.1105/tpc.8.10.1899.

Petersen, T. N., Brunak, S., Heijne, G. von, and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi:10.1038/nmeth.1701.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590-596. doi:10.1093/nar/gks1219.

R. Core Team (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015, Vienna, Austria*. ISBN 3-900051-07-0. Available: http://www. R-project. org/(1.12. 2015).

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* 62, 597–635.

Rawlings, N. D. (2016). Peptidase specificity from the substrate cleavage collection in the MEROPS database and a tool to measure cleavage site conservation. *Biochimie* 122, 5–30. doi:10.1016/j.biochi.2015.10.003.

Rawlings, N. D., and Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem. J.* 290 (Pt 1), 205–218.

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R.

D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632. doi:10.1093/nar/gkx1134.

Ribeiro-Guimarães, M. L., and Pessolani, M. C. V. (2007). Comparative genomics of mycobacterial proteases. *Microb. Pathog.* 43, 173–178. doi:10.1016/j.micpath.2007.05.010.

Richardson, A. J., McKain, N., and Wallace, R. J. (2013). Ammonia production by human faecal bacteria, and the enumeration, isolation and characterization of *Bacteria* capable of growth on peptides and amino acids. *BMC Microbiol.* 13, 6. doi:10.1186/1471-2180-13-6.

Sakurai, M., Suzuki, K., Onodera, M., Shinano, T., and Osaki, M. (2007). Analysis of bacterial communities in soil by PCR–DGGE targeting protease genes. *Soil Biol. Biochem.* 39, 2777–2784. doi:10.1016/j.soilbio.2007.05.026.

Schimel, J. P., and Bennett, J. (2004). Nitrogen mineralization: Challenges of a changing paradigm. *Ecology* 85, 591–602. doi:10.1890/03-8002.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303.

Theron, L. W., and Divol, B. (2014). Microbial aspartic proteases: Current and potential applications in industry. *Appl. Microbiol. Biotechnol.* 98, 8853–8868. doi:10.1007/s00253-014-6035-6.

Tsuboi, S., Yamamura, S., Imai, A., Satou, T., and Iwasaki, K. (2014). Linking temporal changes in bacterial community structures with the detection and

phylogenetic analysis of neutral metalloprotease genes in the sediments of a hypereutrophic lake. *Microbes Environ.* 29, 314–321. doi:10.1264/jsme2.ME14064.

Tully, B. J., Sachdeva, R., Heidelberg, K. B., and Heidelberg, J. F. (2014). Comparative genomics of planktonic *Flavobacteriaceae* from the Gulf of Maine using metagenomic data. *Microbiome* 2, 34. doi:10.1186/2049-2618-2-34.

Vollmer, W., Blanot, D., Pedro, D., and A, M. (2008). Peptidoglycan structure and architecture. *FEMS Microbiol. Rev.* 32, 149–167. doi:10.1111/j.1574-6976.2007.00094.x.

Vranova, V., Rejsek, K., and Formanek, P. (2013). Proteolytic activity in soil: A review. *Appl. Soil Ecol.* 70, 23–32. doi:10.1016/j.apsoil.2013.04.003.

Wallace, R. J., McKain, N., Broderick, G. A., Rode, L. M., Walker, N. D., Newbold, C. J., et al. (1997). Peptidases of the rumen bacterium, *Prevotella ruminicola*. *Anaerobe* 3, 35–42. doi:10.1006/anae.1996.0065.

Watanabe, K., and Hayano, K. (1993). Source of soil protease in paddy fields. *Can. J. Microbiol.* 39, 1035–1040. doi:10.1139/m93-157.

Watanabe, K., and Hayano, K. (1994). Estimate of the source of soil protease in upland fields. *Biol. Fertil. Soils* 18, 341–346. doi:10.1007/BF00570638.

Watanabe, K., and Hayano, K. (1996). Seasonal variation in extracted proteases and relationship to overall soil protease and exchangeable ammonia in paddy soils. *Biol. Fertil. Soils* 21, 89–94. doi:10.1007/BF00335998.

Watanabe, K., Sakai, J., and Hayano, K. (2003). Bacterial extracellular protease activities in field soils under different fertilizer managements. *Can. J. Microbiol.*

49, 305–312. doi:10.1139/w03-040.

Wink, J., Mohammadipanah, F., and Hamedi, J. (2017). *Biology and biotechnology of* Actinobacteria. Springer International Publishing Available at: //www.springer.com/us/book/9783319603384 [Accessed August 12, 2018].

Wlodawer, A., Li, M., Gustchina, A., Oyama, H., Dunn, B. M., and Oda, K. (2003). Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases. *Acta Biochim. Pol.* 50, 81–102. doi:035001081.

Wu, J.-W., and Chen, X.-L. (2011). Extracellular metalloproteases from *Bacteria*. *Appl. Microbiol. Biotechnol.* 92, 253. doi:10.1007/s00253-011-3532-8.

Zimmerman, A. E., Martiny, A. C., and Allison, S. D. (2013). Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J.* 7, 1187–1199. doi:10.1038/ismej.2012.176.

Chapter 3: Diversity and Distributions of Extracellular Peptidases across Fungal Genomes

Trang T. H. Nguyen*

David D. Myrold

To Be Submitted In:

Fungal Ecology

xx March 2019

3.1. Abstract

Nitrogen is an important element that fungi always scavenge from the environment. In soils, most of the nitrogen is present as large polymerized nitrogen molecules, such as plant litter or microbial necromass. This study analyzed the diversity and distribution of extracellular peptidases across 612 fungal species as one of the most important enzymatic factors for fungal nitrogen acquisition. We extracted their genomic information and corresponding annotated protein sequences assembled from several databases, including MEROPS, JGI Genome Portal, and MycoCosm. We then evaluated the diversity and abundance of fungal extracellular peptidases, their evolutionary conservation, and distribution as a function of ecological lifestyle. Annotated peptidase gene sequences were screened for secretion signals, resulting in 31,668 eukaryotic secreted peptidase coding genes belonging to 79 peptidase families. We found that *Ascomycota, Basidiomycota,* and *Mucoromycotina* phyla possess distinguishable sets of secreted peptidases, indicating differential abilities for protein degradation. There are differences in the total secreted peptidases per genome and peptidase complements between fungal ecological groups., For example, saprotrophic fungi have fewer secreted peptidases in their genomes compared to the symbiotic or pathogenic fungal species. Some peptidase families were widespread among the studied fungal taxa, whereas some were strictly clustered in certain lineages. Most of the secreted peptidase families in our study strongly followed the evolution of fungi, implying a genetic conservation of proteolytic functions among fungal species. Our study offers an extensive understanding about the diversity and distribution of the extracellular proteolytic enzymes across the fungal kingdom and provides a foundation

for future research applying transcriptomic and proteomic approaches to study fungal capability to degrade proteins.

3.2. Introduction

Fungi must scavenge nitrogen (N) from their environment. Their preferred N sources are ammonium and amino acids, such as glutamate and glutamine (Marzluf, 1997), but these simple N compounds are often in short supply in most environments. In soils, most of the N is present in large biopolymers from remnants of dead plant or microbial biomass, usually in the form of proteins or microbial cell wall materials (Nannipieri and Paul, 2009). To access the N from these biopolymers, fungi rely on the secretion of extracellular enzymes, such as chitinases and peptidases. This analysis focuses on the peptidases secreted by fungi.

Fungi can exist in diverse trophic modes, with different phenotypic traits, and ecological habitats, resulting in different nutrient acquisition strategies in order to overcome the costly energy tradeoffs of enzyme production. Some fungi rely solely on the dead organic material in the environment, some exchange nutrients with plants through symbiotic relationships, and some gain nutrients by colonizing and damaging their living animal or plant hosts. Many of the fungal traits involved in organic carbon acquisition from complex plant material with a diverse array of lignocellulolytic extracellular enzymes that have been shown to link tightly to the fungal evolutionary history (Nagy et al., 2017; Semenova et al., 2017; Talbot et al., 2013). Our objective is to determine if similar patterns exist between the distribution of extracellular peptidase coding genes in fungi and their ecology and taxonomy.

Fungal species are classically grouped into different ecological groups, or functional guilds. Each fungal guild is defined as a group of species that utilizes the same environmental resources in an analogous way whether or not they are phylogenetically related (Nguyen et al., 2016; Root, 1967). For the purpose of this study, we assigned fungal species into four main ecological groups: symbiotrophic fungi, saprotrophic fungi, symbiotrophic/saprotrophic fungi, and other fungi (mostly pathotrophs). Symbiotrophic fungi represent the fungal symbionts that usually associate with plant roots. Saprotrophic fungi represent the free-living fungal decomposers commonly found in terrestrial habitats. Pathotrophic fungi form an antagonistic relationship with their host, either plant or animal. We expected to see that fungi belonging to different ecological groups, possess unique extracellular peptidases, reflecting the fungal adaptation to distinctive environmental resources.

Saprotrophic fungi account for about 90% of total heterotrophic respiration in forest ecosystems (Crowther et al., 2012) and belong to the *Mucoromycotina*, *Basidiomycota*, and *Ascomycota*. With a rich blend of plant-degrading enzymes (lignocellulolytic enzymes), saprotrophic fungi are a major driver of litter decomposition and regulator of soil carbon cycling (Crowther et al., 2012). Saprotrophic fungi colonize the litter horizons where there is more available organic carbohydrate material for carbon and energy (Talbot et al., 2013). Their hyphal network spreads along the soil-litter interface and creates channels to access the carbon sources and translocate nutrients in soils (Crowther et al., 2012). The litter horizon is also a sink of organic nitrogen, mostly from plant and animal debris. Curiously, however, a recent study of 17 fungal isolates found that saprotrophic fungi were generally associated with

low proteolytic activity (Semenova et al., 2017) and possessed fewer secreted peptidase coding genes compared to some pathogenic fungal species (Soanes et al., 2008). Semanova et al. (2017) found exopeptidases that cleave peptide bonds from the N-terminal end were more dominant than endopeptidases, which were suggested be more beneficial for saprotrophic lifestyle in terms of how quickly the organic nitrogen can be broken down and consumed.

Ectomycorrhizal fungi, a group of symbiotic fungal species, perform not only an important ecological role in relationship with plants (Tedersoo et al., 2009), but also a significant role in nutrient cycling. Ectomycorrhizal fungi form an intercellular interface with plant roots without penetrating plant cell walls and develop their extramatrical mycelia to explore nutrient sources (Agerer, 2001). The mutualistic relationship between fungi and plant roots exists as a balanced parasitism in natural conditions. Ectomycorrhizal fungi originally evolved in *Basidiomycota* through many evolutionary events from saprotrophic ancestors with some independently evolved in *Ascomycota* (Pellitier and Zak, 2018). As a result of evolutionary processes, ectomycorrhizal fungi have lost many genes with saprotrophic functions, such as some lignocellulolytic enzymes (Pellitier and Zak, 2018). It is still an open question whether or not ectomycorrhizal fungi may have retained some nitrogen-related enzyme coding genes in order to utilize nitrogen from soil organic matter. According to some studies, the host plants support symbiotic ectomycorrhizal fungi by providing up to 30% of their photosynthate carbon to exchange for about 70% of their nitrogen and phosphorus requirements (Gorka et al., 2019; Martin and Nehls, 2009; Nehls, 2008; Treseder and Lennon, 2015). In line with this, a recent study demonstrated that ectomycorrhizal fungi

have a significant role in breaking down the polyphenol-protein complexes in soils through the joint action of some proteolytic enzymes and by producing oxidative agents (e.g., radical OH groups) based on the Fenton reaction (Beeck et al., 2018). Some pure culture studies also suggest that ectomycorrhizal fungi have the ability to degrade proteins by producing a diverse mixture of extracellular peptidases, dominantly aspartic endopeptidases (Rineau et al., 2016; Shah et al., 2013). Additionally, the ectomycorrhizal fungal species present in older forests are shown to adapt to the most common N source and therefore have much greater protein degradation abilities (Rineau et al., 2016), suggesting a close correlation between their environment and their ability to scavenge for organic nitrogen.

Pathotrophic fungi represent an ecological group that feeds on the living cells of their hosts (animal, plant, or fungi) for nutrients during at least part of their fungal life cycle. Most of the proteolytic enzymes encoded by these fungal pathogens are connected with pathogenesis-related functions, such as damaging host tissues or deactivating the host defense (Semenova et al., 2017). Trypsin-like serine endopeptidases are a unique marker for pathogenic fungi, which are lost in saprotrophs (Dunaevskii et al., 2006; Hu and Leger, 2004; Semenova et al., 2017). Extracellular peptidases that serve as virulence factors in pathogenic fungi include aspartic peptidases belonging to pepsin family (A1), serine peptidases in subtilisin family (S8), and metallopeptidases in M35 and M36 families (Monod et al., 2002).

Our analysis expands upon prior studies of fungal peptidase production. For example, Semanova et al. (2017) empirically examined the secreted peptidases expressed by 17 species of *Ascomycota* and *Basidiomycota* representing saprotrophs

and pathogens. *Ascomycota* produce predominantly serine peptidases (mostly S08 family, subtilisin) with alkaline pH optima, whereas *Basidiomycota* produce mostly metallopeptidases with more neutral pH optima. Muszewska et al. (2017) did an intensive analysis concentrated on serine peptidases, both intracellular and secreted. They analyzed 634 proteomes across the fungal evolutionary tree, finding that serine peptidase abundance scaled with genome size, with a tendency for greater peptidase richness in evolutionarily younger taxa. In addition, Muszewsak et al. (2017) reported relationships between fungal lifestyle and their repertoire of serine peptidases: plant symbionts having fewer, pathogens and saprotrophs having more.

Regarding the diverse lifestyles, taxonomy, and complicated evolutionary history of fungi, we postulate that the distribution of extracellular peptidases will be influenced strongly and segregate between fungal lifestyles and taxonomic groups. To test this idea, we analyzed the diversity of secreted peptidases across a wide range of fungal species and determined the conservation of the proteolytic trait across an extensive list of well-studied fungal genomes. We obtained the fungal genomic information from JGI Genome Portal, MyCosm 1000 Fungal genome project and annotated their secreted peptidase family information using MEROPS Batch Blast (Grigoriev et al., 2014, 2011; Rawlings et al., 2018).

3.3. Materials and Methods

3.3.1. Collection of fungal secreted peptidases

We selected 612 fungal genomes of interest belonging to three fungal phyla (356 genomes from *Ascomycota*, 234 genomes from *Basidiomycota*, 22 genomes from *Mucoromycotina*) from MycoCosm, the DOE JGI's web-based fungal genomics resource (https://genome.jgi.doe.gov/programs/fungi/index.jsf, data accessed March 2017) (Grigoriev et al., 2014, 2011). Only completely annotated genomes existing in the database with available annotated protein information were considered.

First, data pertaining to the organism names and genome completeness were extracted from JGI Genome Portal (Grigoriev et al., 2011) with unique JGI identification (JGI ID). This allowed us to narrow down the list of fungal genomes for our study to only completely sequenced species. Fungal taxonomic classification was queried from UNITE (Nilsson et al., 2019). The taxonomy ID and taxonomy name were extracted from NCBI Taxonomy Browser on Taxonomy Name/ ID Status Report page (https://www.ncbi.nlm.nih.gov/Taxonomy/TaxIdentifier/tax_identifier.cgi).

Next, we extracted the information table of SigP (protein ID, protein names, protein sequences, sequence ID, signal peptide classification) and EC-KEGG (protein ID, enzyme nomenclature assignment) for each individual fungal species from the JGI Genome Portal (Grigoriev et al., 2011). From the SigP table, we obtained all the annotated protein sequences that were previously identified to be genes encoding putative signal sequence motifs as defined for eukaryotes based on SignalP (Mori and Ito, 2001; Petersen et al., 2011, Grigoriev et al 2014). From the EC-KEGG table, we

screened only the proteins that were classified as proteolytic enzyme (EC 3.4.). Combining these two outputs and using the matching protein ID information, we queried for the peptidase sequences that contained signal peptides in each single fungal genome. These secreted peptidase results for each fungal genomes were summarized into a fasta file that comprised their amino acid sequences with unique species JGI ID, species names, protein ID, sequence ID information. The fasta files for each species were then concatenated together into three fasta files for *Ascomycota*, *Basidiomycota,* and *Mucormycotina*.

We obtained the peptidase family information for the peptidase sequences of interest by MEROPS Batch Blast release 11.0 (Rawlings et al., 2018). The MEROPS database classifies peptidases into seven super-families based on the catalytic residue serving at the active site of the enzyme (Hartley, 1960; Rawlings and Barrett, 1993), and further divides these super-families into 255 proteolytic families based on similarities in amino acid sequences (Rawlings et al., 2018). After the Batch Blast run for the dataset from each fungal phylum, all the peptidase sequences that were recognized as protein inhibitors were removed from the three fasta files. We then obtained the comprehensive list with all the important information about the sequences, including: JGI ID, protein ID, sequence ID, peptidase family classification, amino acid sequences. This data mining process yielded 17,500 peptidase sequences from 356 *Ascomycota* primary genomes of interest, 13,415 peptidase sequences from 234 *Basidiomycota* genomes, and 753 peptidase sequences from 22 *Mucoromycotina* genomes.

3.3.2. Fungal ecological guild annotation

The ecological groups, or guilds, for our fungal species were annotated using FUNGuild, an open annotation tool for parsing fungal datasets (Nguyen et al., 2016) (https://github.com/UMNFuN/FUNGuild). We simplified the fungal guild information into four main ecological groups: Saprotroph (381 taxa), Symbiotroph (mostly Ectomycorrhizal Symbiotroph and Pathotroph-Symbiotroph; 86 taxa), Both (Symbiotroph-Saprotroph or Pathotroph-Symbiotroph-Saprotroph; 17 taxa), and Others (mostly Pathotrophs, with the rest being unclassified; 128 taxa).

3.3.3. Comparison of genomic complements of secreted peptidases

The secreted peptidase complements of all taxa were summarized in a matrix containing the gene copy number counts of secreted peptidases assigned to either family or superfamily classifications (rows) across all analyzed genomes (columns), as described in our previous study of prokaryotic secreted peptidases (Nguyen et al., 2019). Bray-Curtis dissimilarity indices between the secreted peptidase complements of fungal genomes were calculated from these matrices and used to generate a secreted peptidase distance matrix, or functional distance matrix, using the 'Vegan' package in 'R' (Oksanen et al., 2018). Principal coordinate analyses (PCoA) was used to explore the data and Permutational Multivariate Analyses of Variance (PERMANOVA) was used to determine the statistical differences of the peptidase complements of fungal genomes at different taxonomic levels and different fungal guild classification.

The Welch two-sample t-test, or unequal variances t-test, was used to test for statistical differences between the means of the total secreted peptidases encoded within fungal genomes. One-way analysis of variance with the Tukey's HSD multiple-

range test was used to determine the statistical differences between counts of total secreted peptidases and secreted peptidases belonging to each peptidase super-family among fungal phyla and among fungal ecological guilds. Statistical analyses were performed in the 'R' programming environment (R. Core Team, 2016).

3.3.4. Phylogenetic analysis

A downloaded phylogenetic tree was constructed by the MycoCosm team based on available protein sequence alignments (Grigoriev et al., 2014). For our analysis, we pruned the tree so that it contained 612 primary fungal genomes of interest using 'APE' package (Analyses of Phylogenetics and Evolution) in R (Paradis and Schliep, 2019). The Newick tree for 612 fungal genomes is included in the Supplementary files. The phylogenetic distances were back calculated based on the phylogenetic tree and these distances were converted into a phylogenetic distance matrix. The phylogenetic tree and distributions of secreted peptidase families across the tree were visualized using iTOL (Letunic and Bork, 2016).

3.3.5. Distance matrices comparisons

Correlations between the phylogenetic distance matrix and the secreted peptidases distance matrix, or functional distance matrix, were evaluated using the Mantel test of 'APE' (Analysis of Phylogenetics and Evolution package) in 'R' (Paradis and Schliep, 2019) based on Pearson's product-moment correlation. Mantel correlograms that report the correlation between phylogenetic and functional distances at defined phylogenetic distance classes for *Fungi* were calculated using the 'Vegan' package in 'R'.

3.3.6. Phylogenetic conservation and clustering

As described by Nguyen et al. (2019), phylogenetic signal strengths (D) contributing to the observed distribution patterns for each peptidase super-family and family were calculated from their binary presence/absence in genomes of all considered taxa (Fritz and Purvis, 2010) using the 'CAPER' package (Comparative Analyses of Phylogenetics and Evolution) in 'R' (Orme et al., 2018). Secreted peptidases are considered phylogenetically conserved when they are shared among the majority of members of deeply branched clades, conforming to a Brownian motion evolutionary model ($D \sim 0$), with a relatively constant gain/retention of traits across taxonomic levels. A strongly clumped distribution ($D < 0$) suggests recent innovation or potential gain via horizontal gene transfer within a clade or subset therein. Peptidases are considered randomly distributed ($D \geq 1$) when their presence/absence is not driven by shared traits (e.g., microhabitat, physiology) of closely related species (Berlemont and Martiny, 2013; Martiny et al., 2013; Zimmerman et al., 2013).

We used consenTRAIT, a consensus analysis of phylogenetic trait distribution (Martiny et al., 2013) to assess the phylogenetic distance at which a group of genes belonging to a given peptidase family cluster among fungal taxa. The phylogenetic distance between the fungal taxa that share the same traits, in this case, the same peptidase family is called τ_D. Any peptidase family that had only one sequence across all fungal taxa (singletons) was assumed to have equal likelihood of finding an adjacent organism with or without the trait and have τ_D equal to half of the distance to the closest node on the phylogenetic tree. This analysis was done in R using the script provided

by Martiny et al. (2012) and ‘APE’ and ‘adephylo’ packages (Jombart et al., 2010; Paradis and Schliep, 2019).

3.3.7. Secreted peptidase distribution in association with fungal ecological functions.

To understand the association between the distribution of secreted peptidases and ecological functions, we examined taxonomic subsets of microorganisms, including genomes of 86 fungal symbiotroph, 95 saprotroph, and 90 pathotroph species. Their trophic mode and guild information was annotated using FUNGuild and visualized on the phylogenetic trees together with secreted peptidase count data using iTOL. For the symbiotrophic dataset, we identified the species based on two ecological groups of ectomycorrhizal symbiotroph and pathotroph-symbiotroph (mostly endophyte-plant/animal pathogen). The saprotrophic subset of data consisted of brown-rot and white-rot fungal genomes, which differ in their ability to degrade lignin, cellulose, and hemicellulose. Pathotrophic group was categorized into three groups based on their hosts: plant pathogen, animal pathogen, and fungal parasite. We used ANOVA analysis to determine the statistical differences between the mean gene copy number of each secreted peptidase super-family between fungal groups and represented these differences using boxplot.

3.4. Results

3.4.1. Relative abundance of secreted peptidase super-families across fungal kingdom

Across all 612 studied fungal genomes, serine peptidases contributed the majority of the extracellular peptidase repertoire (52%), followed by aspartic peptidases (29%), and metallopeptidases (17%) (Fig. 3.1). Cysteine peptidases were a small fraction of the genomic potential (1.3%) and the remaining 0.7% of fungal secreted peptidase genes belonged to glutamic, threonine, and unknown peptidases (Fig. 3.1).

Serine peptidases were dominant in *Ascomycota* and *Basidiomycota* (50-55%), whereas aspartic peptidases made up 50% the secreted peptidase encoding genes in *Mucoromycotina* (Fig. 3.1). Metallopeptidases contributed the most to *Ascomycota* (21%) but represented only 10-11% of extracellular peptidase genes in the other two phyla (Fig. 3.1). *Mucoromycotina* also possessed more cysteine peptidases (4%) compared to *Ascomycota* and *Basidiomycota* (Fig. 3.1). By contrast, when we compared the relative abundance of secreted peptidase super-families between fungal ecological groups, the differences were not very noticeable, except that symbiotrophic fungi had marginally higher relative aspartic peptidases than other groups (Fig. S3.1).

3.4.2. Distribution of secreted peptidases across fungal phylogeny

We observed significant differences of extracellular peptidase super-family complements among the three fungal phyla (Fig. 3.2, p=0.001). *Basidiomycota* coded for more secreted peptidases per genome than *Ascomycota* and *Mucoromycotina*. *Ascomycota* and *Basidiomycota* possessed significantly more serine peptidases than

Mucoromycotina. Metallopeptidases and threonine peptidases were the most abundant in *Ascomycota*. Although only a small fraction of the secreted peptidase genetic potential in *Fungi*, cysteine peptidases were most prevalent in *Mucoromycotina*.

At the peptidase family level, 22 peptidase families belonging to aspartic, cysteine, metallo-, and serine peptidases were common among the three phyla and contributed to more than a quarter of the total peptidase families present in the dataset (Fig. 3.3). *Ascomycota* and *Basidiomycota* shared 31 peptidase families that belong to all peptidase catalytic types; *Mucoromycotina* shared only two peptidase families with each of the other two fungal phyla, besides the 22 families shared among all three phyla. *Mucoromycotina* possessed only one unique peptidase family (M79), whereas *Ascomycota* had 16 and *Basidiomycota* had five unique peptidase families.

Principal coordinate analysis grouped species from the three fungal phyla based on the relative abundance of secreted peptidase families they encode (Fig. 3.4). Although the first two PCoA axes explained only 10-16% of data variance, a strong and significant difference of the secreted peptidase profiles was observed between fungal phyla according to PERMANOVA ($p = 0.001$). The *Mucoromycotina* cluster was most correlated with the A01 (aspartic) and S08 (serine) peptidase families.

Distributions of peptidase families and corresponding super-families across fungal taxa were compared to the phylogenetic relationships among analyzed genomes using relative abundance profiles of peptidases in comparison to the fungal phylogenetic tree that was constructed by aligning protein sequences (Grigoriev et al., 2014). The distributions of secreted peptidases were significantly correlated with fungal phylogeny ($r_{Mantel} = 0.210$, $p = 0.01$), indicating an evolutionary relationship in

which subsets of phylogenetically related organisms shared similar types of secreted peptidases. Some of these evolutionary patterns of the peptidase distribution can be observed when secreted peptidase gene abundance is viewed on the fungal phylogenetic tree (Fig. 3.5). For example, the earlier evolved group of *Basidiomycota* generally had fewer secreted peptidases compared to the later evolved species in the same phyla. This trend was also true for *Ascomycota*. Some *Ascomycota* groups were richer in aspartic and metallopeptidases than other groups, shown with darker outer tracks in Fig. 3.5. Mantel correlograms showed that phylogenetic conservation of secreted peptidases was significant between closely related fungal taxa (Fig. 3.6). The more distant the pairs of taxa were from each other, the higher chance that this correlation would be insignificant, meaning these species might share little functional similarity in their secreted peptidase complements.

Distributions of individual secreted peptidase families were also evaluated for their phylogenetic dispersion (D). Most of peptidase families (84%) encoded in fungal genomes showed evidence of non-random phylogenetic clustering (Fig. 3.7 and Table S3.1). The remaining secreted peptidase families were randomly distributed, potentially initiated by gene loss or gene transfer. Peptidase families with negative values ($D < 0$) represented those with the strongest clustering patterns across the phylogenetic tree. For example, M79, an endopeptidase belonging to metallopeptidase class, is unique to some *Mucoromycotina* species, S24, an endo-serine peptidase is found mainly within *Basidiomycota*, and S51, an exo-serine peptidase, is present in only *Ascomycota* (Tab. S3.1).

Given that the majority of the secreted peptidase were non-randomly phylogenetic clustered, we used consenTRAIT to determine if the τ_D, the clade depth in fungal phylogenetic distance at which a cluster of genes from different taxa occurred, varied among all the peptidase families. Of the 79 peptidase family traits that we analyzed, A01, M14, M28, S08, S09, S10, S28, and S53 were the most abundant peptidase families and their τ_D also yielded the highest values of 1.58 in fungal phylogenetic distances. The rest of the peptidase families covered a wide range of τ_D values from 0.001 to 0.137, depending on their distribution pattern across the fungal taxa. For example: S15 serine peptidase family with τ_D=0.001 was found only in five *Ascomycota* and *Basidiomycota* taxa; M12 metallopeptidase family with τ_D=0.101 was common in 408 fungal taxa among all three phyla.

3.4.3. Distribution of secreted peptidases across fungal ecological groups

We observed significant differences in the extracellular peptidase complements among fungal ecological groups (PERMANOVA, p=0.001, Fig. S3.2). Symbiotrophic fungi possessed significantly more total secreted peptidases, especially aspartic and metallopeptidases, than saprotrophic fungi (Fig. 3.8). There was no difference between the "Saprotrophs" and "Others" groups (pathotrophic fungi), except for the slight variance in their threonine peptidase content. Significant differences were also observed between symbiotrophic and pathotrophic groups in terms of their total secreted peptidase, serine, and aspartic peptidase gene content (more peptidases were found in symbiotrophic species).

The symbiotrophic fungi that belong to *Basidiomycota*, mostly ectomycorrhizal fungi, tended to possess high content of serine and aspartic peptidases, but very low number of metallopeptidases (Fig. 3.9A, 3.9B). The earlier evolved group of symbiotrophic *Ascomycota* had relatively fewer peptidases compared to the latter evolved *Ascomycota*, which are pathotrophic/symbiotrophic fungi, the fungi that can function as a mutualist or antagonist depending on the situation. The group of pathotrophic/symbiotrophic fungi exhibited a significant higher content of metallopeptidases compared to the rest of the symbiotrophic fungi that we investigated (Fig. 3.9A, 3.9B).

We selected a small subset of saprotrophic fungi within the *Basidomycota* to compare their secreted peptidase complements. This group was divided into known brown-rot and white-rot saprotrophic fungi. Brown-rot fungi decompose cellulosic plant material, whereas white-rot fungi are capable of degrading lignin. Interestingly, white-rot fungi had a significantly larger genetic potential to break down proteins compared to brown-rot fungi (Fig. 3.10A, 3.10B). Brown-rot fungi generally possessed lower serine and metallopeptidases compared to white-rot fungi, whereas, the aspartic, cysteine, threonine, and unknown peptidase gene numbers were similar between the two fungal groups (Fig. 3.10B).

We examined the pathogenic fungi within "Other" category and divided them into three groups: plant pathogens, animal pathogens, and fungal parasites. Serine peptidases differed significantly among these three fungal pathogenic groups (Fig. 3.11A). Animal pathogenic fungi were more likely to contain more secreted peptidases compared the other two groups (Fig. 3.11B).

3.5. Discussion

3.5.1. Secreted peptidase distribution correlate with fungal taxonomy

Fungi possess a different collection of extracellular peptidases compared to *Archaea* and *Bacteria*, which implies potentially differential ecological roles in organic nitrogen acquisition. Serine and metallopeptidases are among the most abundant secreted peptidase super-family in prokaryotic genomes (~75-80%) (Nguyen et al., 2019), similarly to serine and metallopeptidase relative abundance in fungal genomes from this study. Aspartic peptidases contribute 30% to the fungal extracellular proteolytic enzymes. This finding is in line with several studies that found secreted aspartic peptidases contribute a high proportion of secreted peptidase coding genes in fungal genomes and their expressed extracellular peptidases (Caldwell, 2005; Rineau et al., 2016; Shah et al., 2013; Theron and Divol, 2014). These aspartic peptidases are usually optimized at acidic pH, compared to serine peptidases (alkaline pH) and metallopeptidases (neutral pH) (Rao et al., 1998; Theron and Divol, 2014; Wu and Chen, 2011). Cysteine peptidases are generally optimal in neutral pHs, but a few of them can function in very acidic environments (Rao et al., 1998). Although cysteine peptidases contributed to about 20% of secreted peptidase repertoires in prokaryotes, in fungal genomes they only contribute less than 2%. Consequently, the variation in fungal peptidase repertoires compared to the prokaryotic peptidases might indicate that *Fungi* can be flexibly active in proteolysis across a wide range of pH conditions, especially with superior adaptation to acidic environments.

The distribution of secreted peptidases across the three fungal phyla generally followed a pattern of phylogenetic conservation (r_{Mantel} = 0.21, p = 0.01, Fig. 3.6). *Mucoromycotina* represents an early evolved group of *Fungi* that lost flagellum and colonized terrestrial ecosystems, whereas the *Ascomycota* and *Basidiomycota* evolved more recently (Muszewska et al., 2017; Spatafora et al., 2016). Each of these phyla have a distinguishable set of secreted peptidase complements (Fig. 3.1 and Fig. 3.2). In our study, *Basidiomycota* possessed the highest number of secreted peptidase coding genes. Interestingly, the difference in the total number of secreted peptidase coding genes between *Basidiomycota* and two other fungal phyla seemed to be driven mainly by the aspartic peptidase content (belonging to the A01, A11, and A22 families) (Fig. 3.2). The A01 family was found in 610 of the 612 studied fungal genomes and was shared among all three fungal phyla. The A22 family occurred in 66 genomes across all phyla, whereas the A11 family was present in just three fungal species from *Basidiomycota* and *Ascomycota*. These aspartic peptidases were well conserved with the fungal phylogeny (Fig. 3.3, Fig. 3.7 and Table S3.1). The ubiquity of A01 peptidases and its strong connection to phylogeny could make it a useful target for developing primers that could be used to assess the genomic potential and diversity of aspartic peptidases in environmental studies.

Although a study of 17 fungal isolates suggested that metallopeptidase expression might be more characteristic of *Basidiomycota* species and serine peptidases (especially S08 subtilisin) with *Ascomycota* species (Semenova et al., 2017), this was not supported by our study of a much wider range of fungi. *Ascomycota* have greater potential for producing secreted serine and metallopeptidases in comparison with

Basidiomycota and *Mucoromycotina*. This is consistent with the gain of peptidase-coding genes as the fungal species evolved to adapt to different environmental conditions and types of substrates (Muszewska et al., 2017). *Mucoromycetes*, as the more ancient group of our studied fungi, had a relatively low content of extracellular serine and metallopeptidases (Figs. 3.2 and 3.5). Many of the more recently evolved *Basidiomycetes* lineages possessed a richer collection of serine and metallopeptidases compared to the earlier branches (Fig. 3.5). This trend generally held with *Ascomycetes*, except that the most recent *Ascomycota* clades had reduced number of metallopeptidase coding genes in their genomes. About a third of the serine and metallopeptidase families were shared among three fungal phyla and two-thirds were shared only between the two younger fungal phyla (*Ascomycota* and *Basidiomycota*) (Fig. 3.3). About 80-85% of the serine and metallopeptidase families followed the fungal evolution (Fig. 3.7, Table S3.1) and were correlated with the differences among fungal phyla (Fig. 3.4). Eight families (M18, M23, M50, M54, M77, S14, S41, S49) were present in only a few randomly distributed taxa across the phylogeny suggesting that the presence of these families could have been initiated by gene loss or gene transfer.

Compared to the repertoires of secreted peptidases in prokaryotes, fungal genomes had a low contribution of extracellular cysteine peptidases (Fig. 3.1). This finding is in agreement with pure culture studies that found no significant contribution of extracellular cysteine peptidases to the fungal proteolytic activity (Rineau et al., 2016; Shah et al., 2013; Silva et al., 2006). Only two of the cysteine peptidase families were shared between the three phyla, half were shared only between *Ascomycota* and *Basidiomycota*, and the rest were unique. Overall, 80% of the cysteine peptidase

families found in fungal genomes followed Brownian-evolutionary pattern, clustering at deep branches (Fig. 3.7 and Table S3.1).

3.5.2. Secreted peptidase distribution correlate with fungal ecological functions and lifestyles

Saprotrophic fungi are well-known for their capacity to break down complex plant materials thanks to their rich collection of lignocellulolytic enzymes. Interestingly, saprotrophic fungi did not possess a larger collection of proteolytic enzymes to break down organic nitrogen, compared to other ecological fungal groups (Fig. 3.8), supporting an earlier study that found low secreted peptidase activity in saprotrophic fungal isolates (Semenova et al., 2017). We further compared the proteolytic potential between brown-rot and white-rot saprotrophic fungi in *Basidiomycota* and found that white-rot fungi had a richer collection of all families of secreted peptidases compared to their brown-rot ancestors. White-rot fungi are associated with their ability to break down lignin in plant materials by producing enzymes that can oxidize these aromatic compounds (Semenova et al., 2017). In decaying litter and soil, proteins often exist in complexes with (poly)-phenolic compounds, which slows degradation (Adamczyk et al., 2008; Northup et al., 1995). The combination of the oxidative ability of white-rot fungi along with their proteolytic enzymes likely enables them to efficiently mine for carbon and nitrogen (Adamczyk et al., 2009).

Most of the fungal symbiotrophs, especially ectomycorrhizal fungi, evolved from free-living fungal groups and were believed to have lost many saptrophic

functions involved in decay of plant materials (Beeck et al., 2018). Surprisingly, we found that these symbiotrophic fungi possessed an abundant collection of extracellular proteolytic enzymes that could help them scavenge nitrogen from the environment. This finding is in line with other research that characterized the proteolytic activity in ectomycorrhizal fungi (Rineau et al., 2016; Shah et al., 2013). Beeck et al. (2018) suggested that even though ectomycorrhizal fungi lacked lignocellulolytic enzymes, they could still oxidize some complex plant materials by using the chemical Fenton reaction that produces the OH radicals for oxidation. They found evidence that secreted peptidase production was delayed and followed the OH production, in order to efficiently break down soil (poly)-phenolic-complexed proteins (Beeck et al., 2018). They also implied this dual oxidative-proteolytic system was adapted from an ancestral brown-rot carbon acquisition strategy and served as a means for ectomycorrhizal fungi to obtain nitrogen in exchange for carbon from their plant host.

Within the symbiotrophic fungi group, there are several fungi that are either pathotrophic and symbiotrophic, mostly *Ascomycota* species. Compared to ectomycorrhizal fungi from the same phylum, these pathotrophic-symbiotrophic fungal species possessed more secreted peptidases, especially serine, metallo-, and cysteine peptidases. This proteolytic enrichment might be linked to pathogenic functions, such as deactivating host defenses or degrading host cells (Monod et al., 2002; Semenova et al., 2017).

The distribution of secreted peptidases in pathotrophic fungi was also unique. Similar to previous findings, we found serine peptidases to be the most dominant in pathotrophic fungi (Fig. S3.1) (Monod et al., 2002; Semenova et al., 2017). Differences

were also recognized between pathotrophic fungal groups based on their guilds: plant pathogen, animal pathogen, and fungal parasites. Serine peptidases drove the main differences between the secreted peptidase repertoires among these fungal species (the most in animal pathogens, the least in fungal parasites), as found by others who identified subtilisin (S08) and trypsin (S01) peptidases as virulence factors involved in pathogenesis (Monod et al., 2002; Muszewska et al., 2017; Semenova et al., 2017). We also found that some more recently evolved *Ascomycota* plant pathogens had more rare secreted peptidases (cysteine, threonine, and unknown peptidases). This genetic gain might indicate gene transfer or convergent evolution.

Based on our observations, we suggest that fungal lifestyles and evolutionary history play significant roles in shaping their proteolytic functions. Fungi are one of the most prevalent eukaryotic groups in terrestrial ecosystems and play a critical role in recycling carbon and nitrogen. This study offered an extensive understanding about the diversity and distribution of the extracellular proteolytic enzymes across fungal kingdom and provided a foundation for future research applying transcriptomic and proteomic approaches to study the fungal ability to break down proteins and consequently enhance our understanding of fungal N cycling in terrestrial ecosystems.

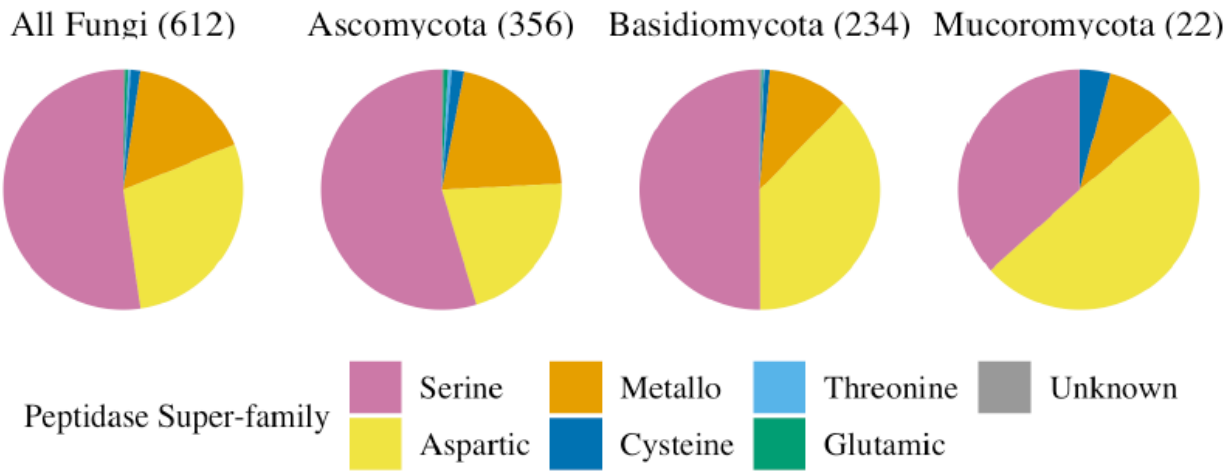


Figure 3.1. Relative abundance of secreted peptidase super-families in 612 fungal genomes (aspartic, cysteine, glutamic, metallo-, serine, threonine, and unknown peptidase super-families) and in different fungal phyla. Different colors represent different peptidase super-families.

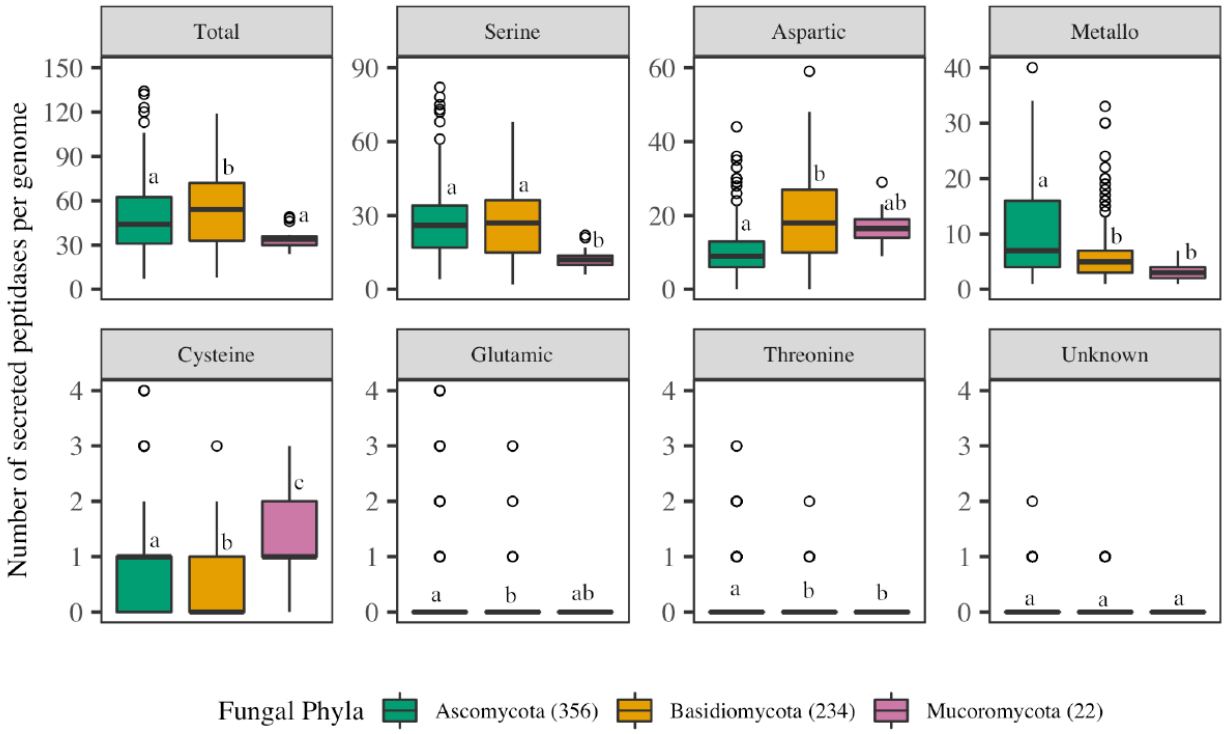


Figure 3.2. Secreted peptidase gene content (per genome) of fungal phyla. Secreted peptidases were grouped into super-families: Total secreted peptidases (including genes from all peptidase super-families); serine, aspartic, metallo- peptidases as more abundant families; cysteine, glutamic, threonine, unknown peptidases as less abundant families. The number of analyzed genomes from each fungal phylum is presented next to the phylum names. Letters on top of each box represent statistical differences of secreted peptidases between three fungal phyla using Tukey's HSD analysis ($p \leq 0.05$).

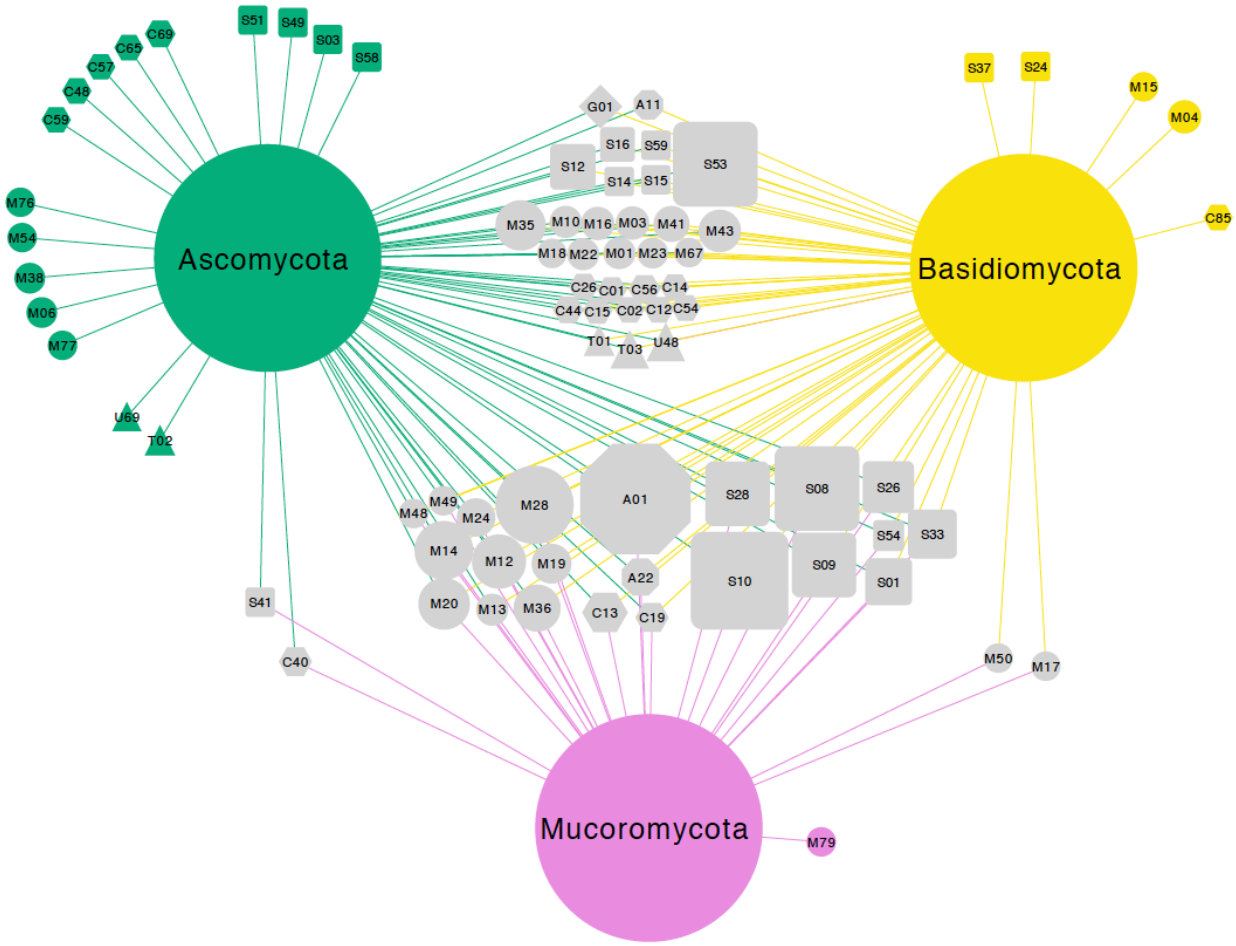


Figure 3.3. Bipartite association network of shared peptidase families between *Ascomycota, Basidiomycota* and *Mucoromycota* fungal phylum. Node sizes indicate the relative abundance of the secreted peptidases. Node shapes represent different peptidase families: octagon = aspartic; hexagon = cysteine; diamond = glutamic; circle = metallo-; square = serine; triangle = threonine and unknown. Node colors are coded by unique or shared peptidase families between microbial kingdoms (green = *Ascomycota*, yellow = *Basidiomycota*, pink = *Mucoromycota*, gray = shared between phyla). Edges denote associations between fungal phyla and peptidase families. Edge colors are coded by fungal phyla.

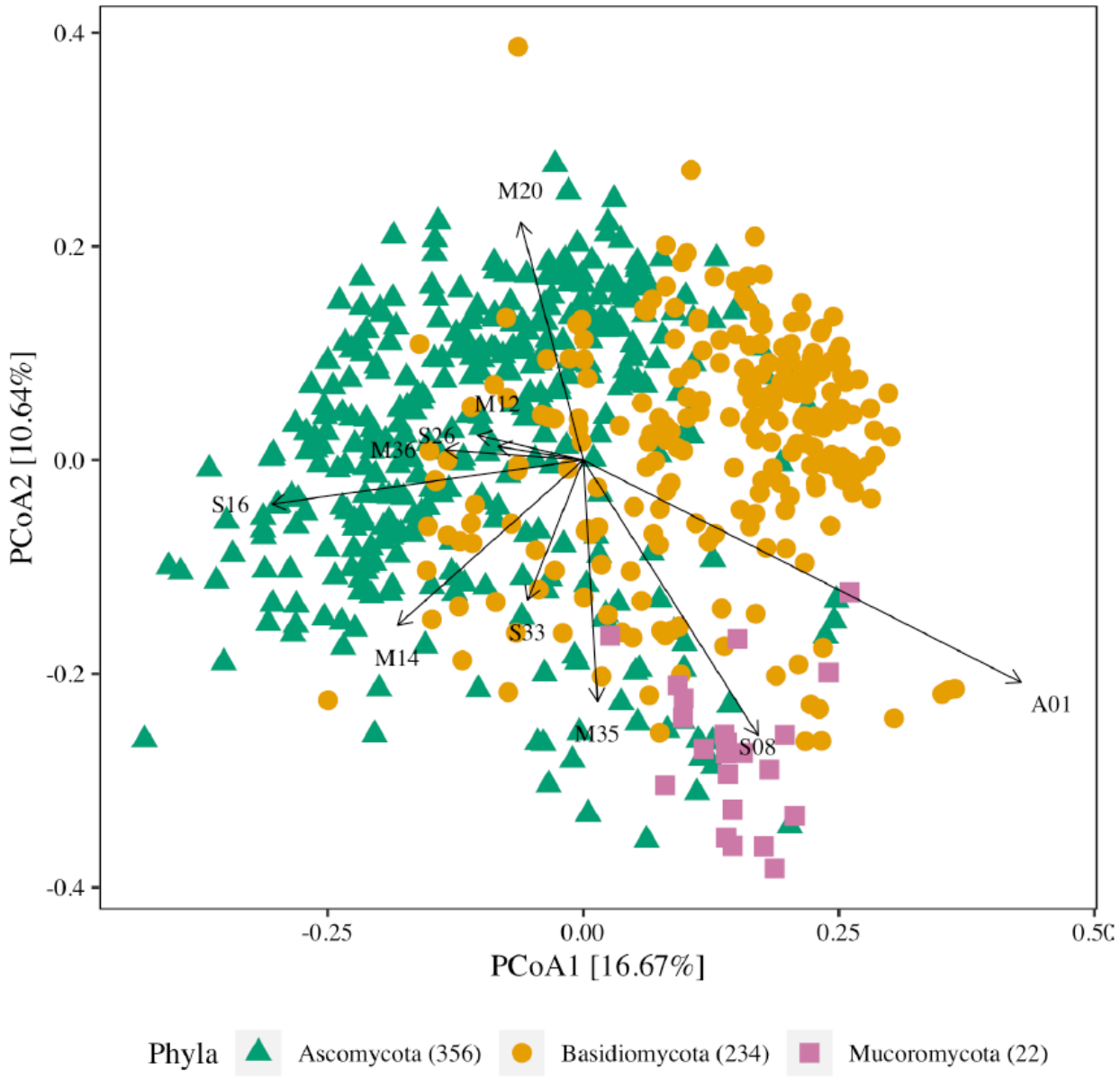


Figure 3.4. Principal coordinate analysis of fungal genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in fungal genomes. Symbol shapes and colors are coded by fungal phyla. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). The composition of secreted peptidase genes of *Fungi* varied significantly based on their phyla ($p = 0.001$, F-statistic =61.044)

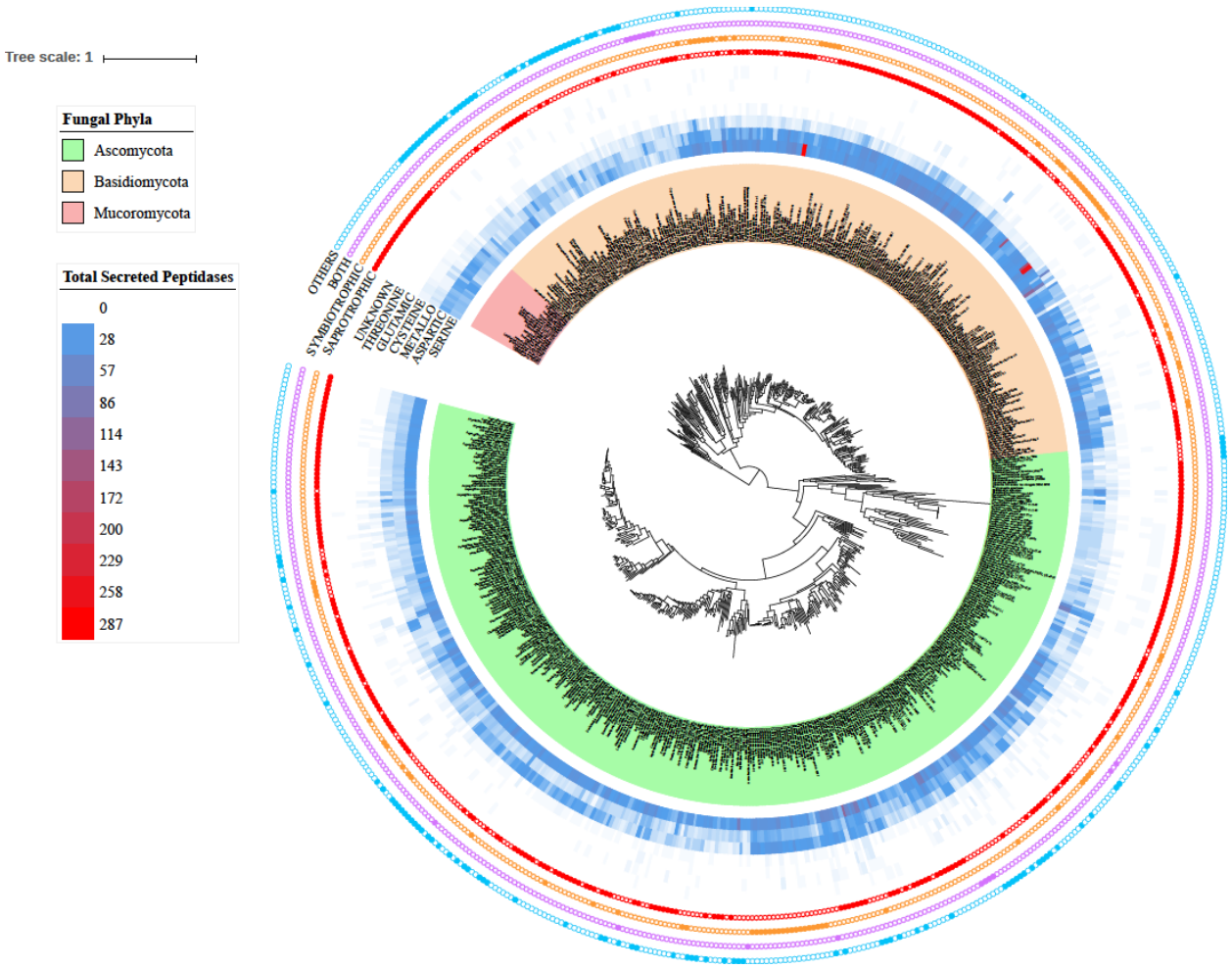


Figure 3.5. Distribution of secreted peptidase super-families across the fungal phylogenetic tree (pruned from Mycocosm fungal tree). The most outer tracks show functional guilds that each fungal genome is generally assumed (solid circle means the genome belongs to the certain guild, the open circle means the genome does not belong to the certain guild). The second outer tracks show the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered.

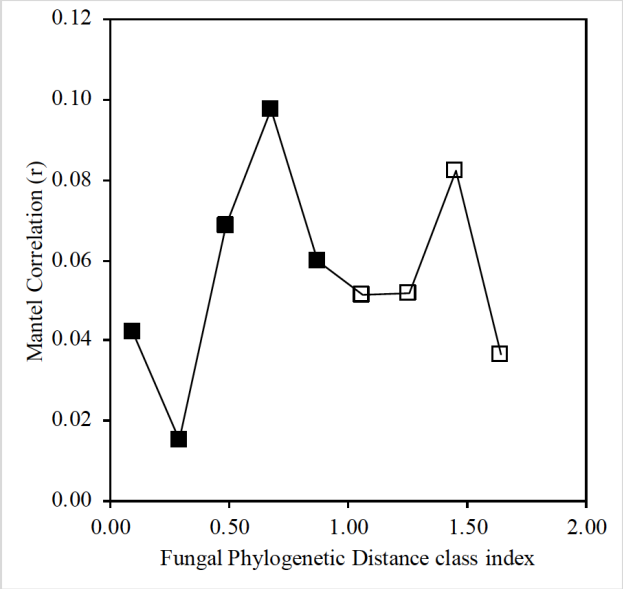


Figure 3.6. Mantel correlogram between phylogenetic distance and secreted protease profile dissimilarities for fungal taxa based on Pearson's product-moment correlations (p-value < 0.05, filled squares; not significant, open squares).

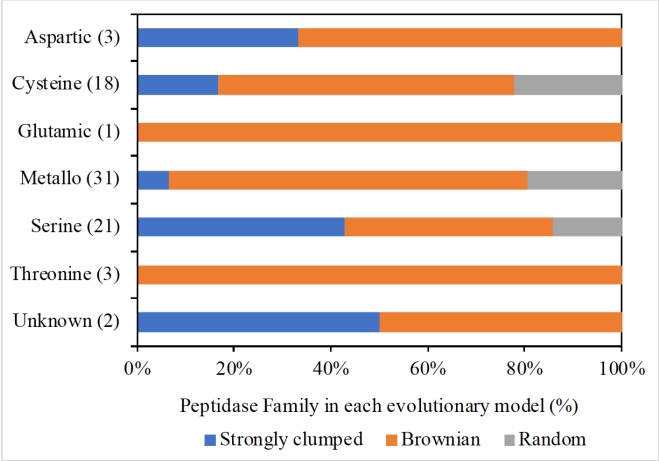


Figure 3.7. Phylogenetic distributions of secreted peptidase families across fungal taxa.

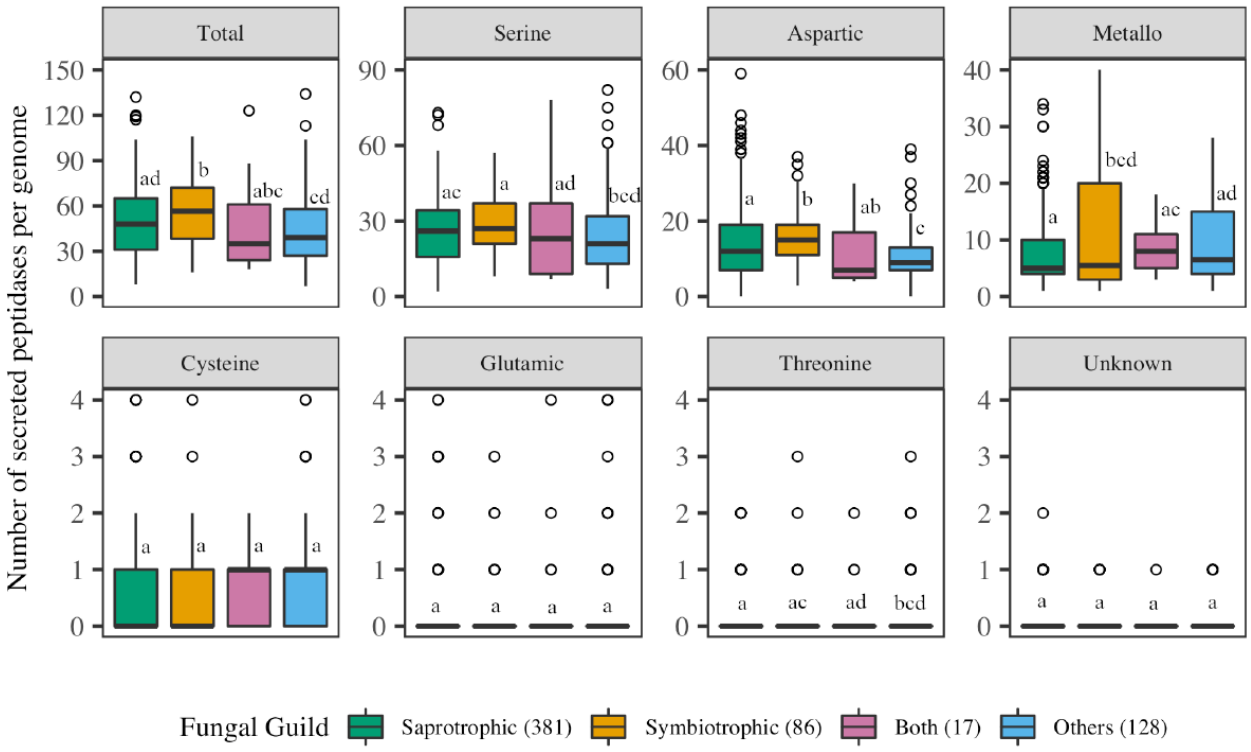
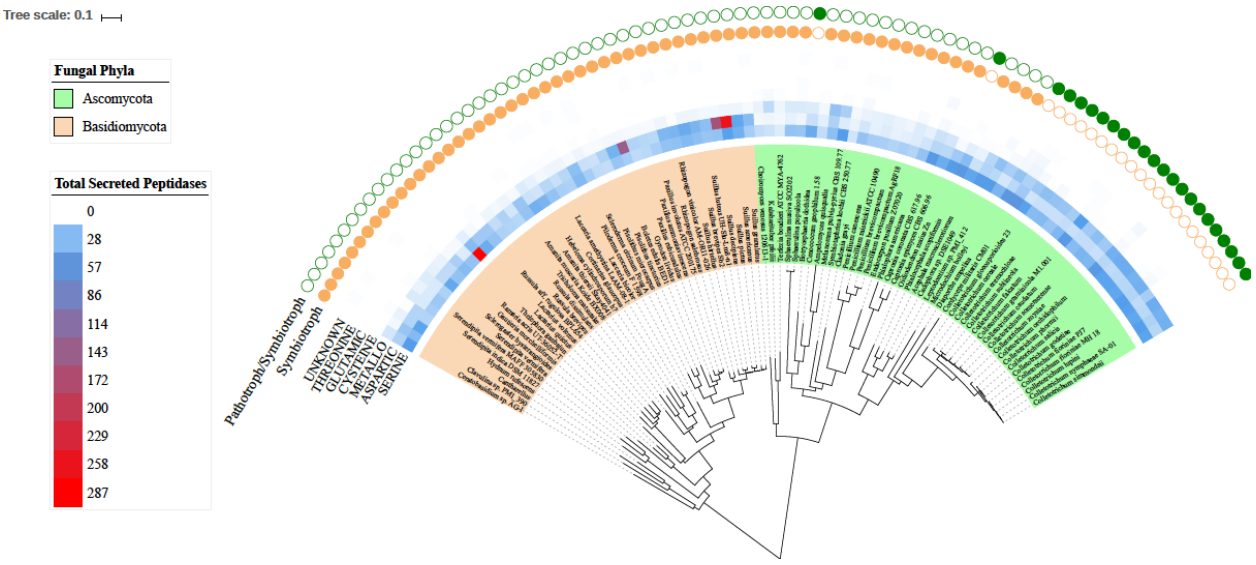


Figure 3.8. Secreted peptidase gene content (per genome) of fungal guilds. Secreted peptidases were grouped into super-families: Total secreted peptidases (including genes from all peptidase super-families); serine, aspartic, metallo- peptidases as more abundant families; cysteine, glutamic, threonine, unknown peptidases as less abundant families. The number of analyzed genomes from each fungal phylum is presented next to the phylum names. Letters on top of each box represent statistical differences of secreted peptidases between fungal functional guilds using Tukey's HSD analysis ($p \leq 0.05$).

(A)



(B)

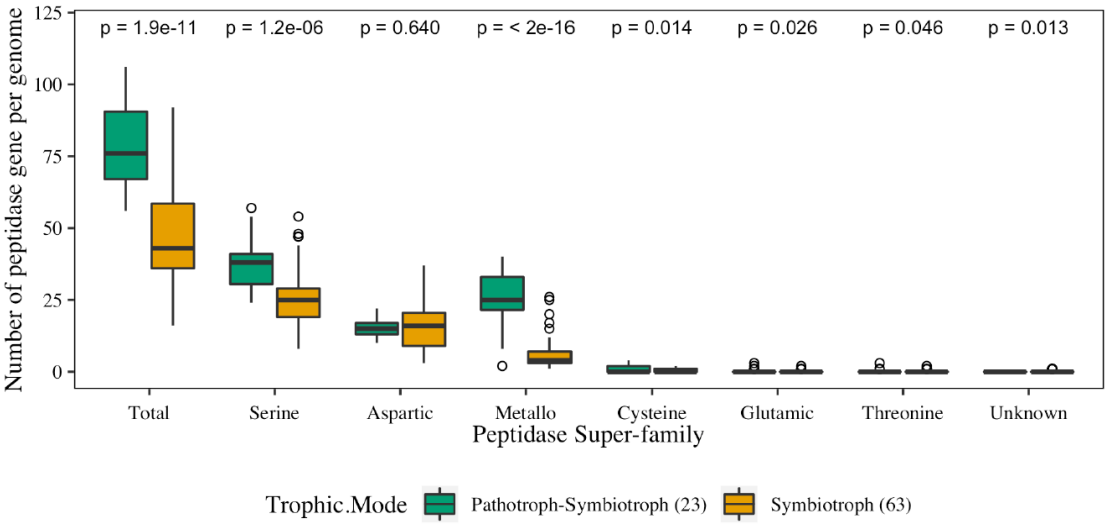
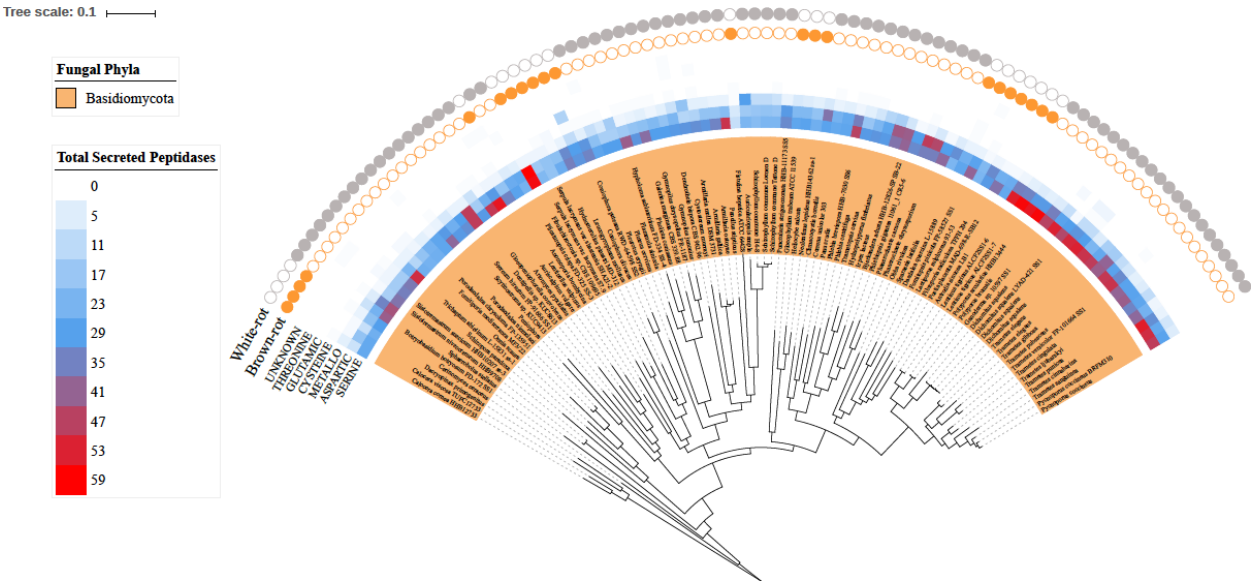


Figure 3.9. **(A)** Distribution of secreted peptidase super-families across the phylogenetic tree of 86 symbiotrophic fungi (pruned from Mycocosm fungal tree). The most outer tracks represent the trophic mode of these fungi (pathotrophic-symbiotrophic in dark green circles and symbiotroph, mostly representing ectomycorrhizal fungi, in yellow circles). The second outer tracks show the copy number of genes from each secreted peptidase super-family in each genome. Inner track

color corresponds to the phylum-level classification of each taxon considered. **(B)** Secreted peptidase gene content (per genome) of 86 symbiotrophic fungal groups. Secreted peptidases were grouped into super-families. The number of analyzed genomes from each fungal group is presented next to the fungal guild classification. The green box plots report mean and standard deviation of the peptidase content of genomes that can be either pathotrophic or symbiotrophic, and the orange box plots report mean and standard deviation of the peptidase content of symbiotrophic genome with p-values.

(A)



(B)

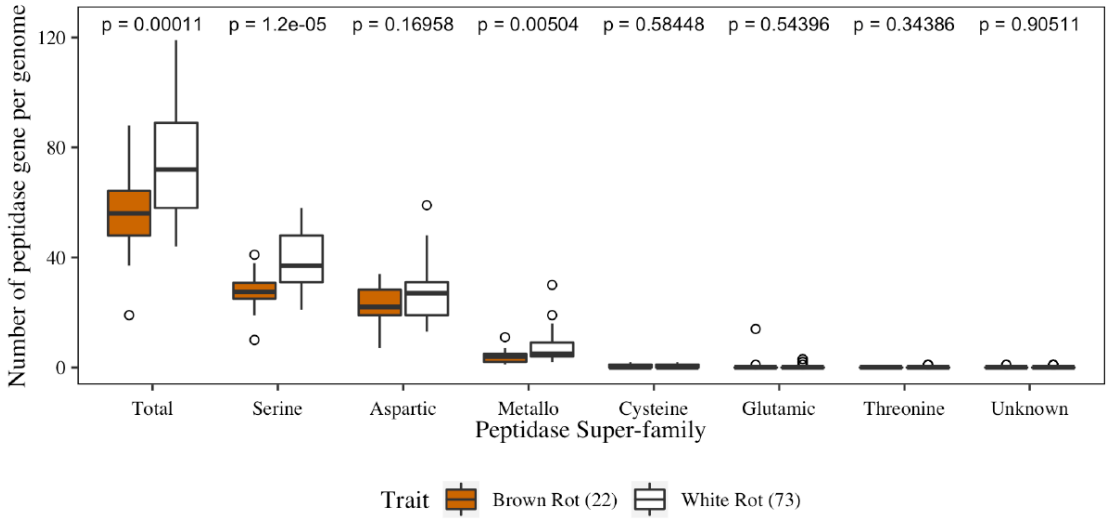
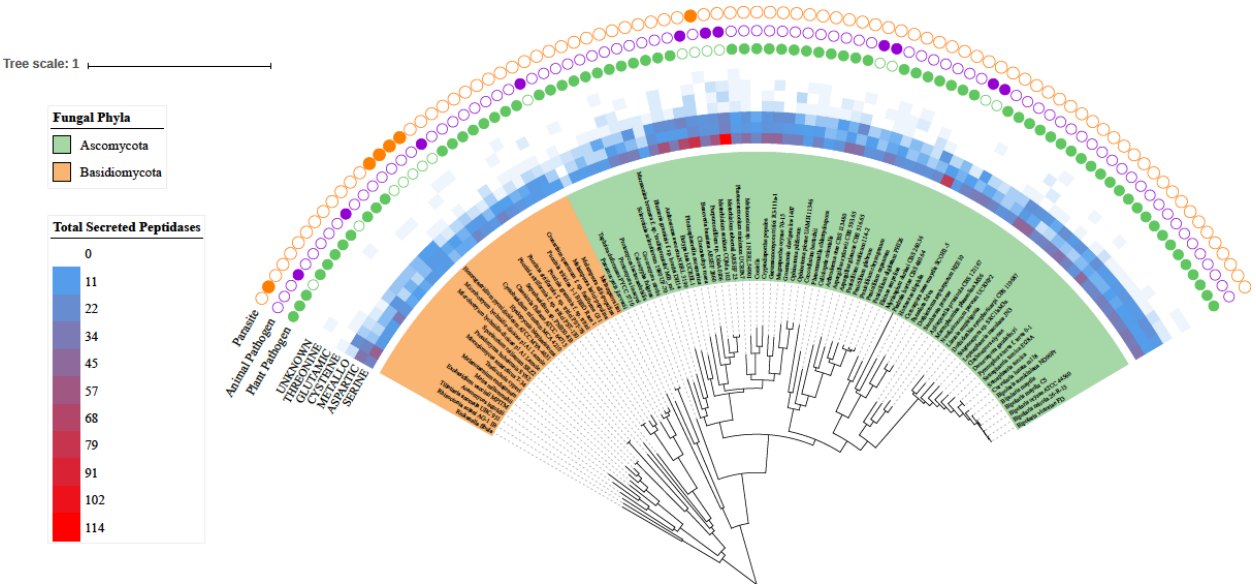


Figure 3.10. (A) Distribution of secreted peptidase super-families across the phylogenetic tree of 95 saprotrophic fungi (pruned from Mycocosm fungal tree). The 95 saprotrophic fungi in this paper were simply classified into the "Saprotrophic" group of fungal guild. The outer most set of tracks represents the main classification of these saprotrophs (white-rot vs. brown-rot fungal species). The second set of outer tracks

shows the copy number of genes from each secreted peptidase super-family in each genome. Inner track colors correspond to the phylum-level classification of each taxon considered. (B) Secreted peptidase gene content (per genome) of brown-rot and white-rot saprotrophic fungal groups. Secreted peptidases were grouped into super-families. The number of analyzed genomes from each fungal group is presented next to the fungal trait classification. The brown box plots report mean and standard deviation of the peptidase content of genomes that belonging to brown-rot fungi, and the white box plots report mean and standard deviation of the peptidase content of white-rot fungi genomes, with p-values.

(A)



(B)

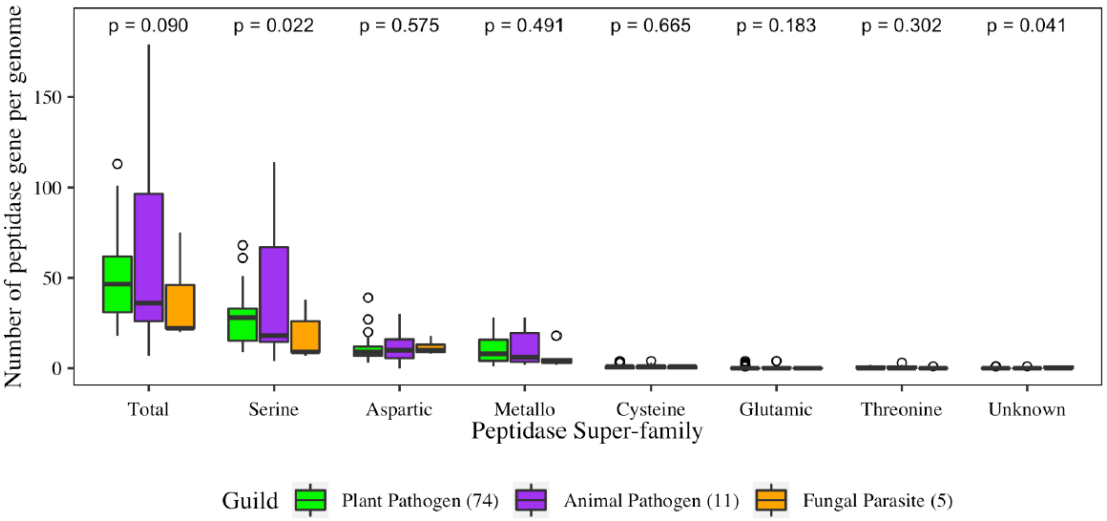


Figure 3.11. (A) Distribution of secreted peptidase super-families across the phylogenetic tree of 90 pathotrophic fungi (pruned from Mycocosm fungal tree). The 90 pathographic fungi in this paper was simply classified into the "Others" group of fungal guild. The outer most set of tracks represents the main classification of these pathogens (plant, animal and parasite fungal species). The second set of outer tracks

shows the copy number of genes from each secreted peptidase super-family in each genome. Inner track color corresponds to the phylum-level classification of each taxon considered. (B) Secreted peptidase gene content (per genome) of pathotrophic fungal groups. Secreted peptidases were grouped into super-families. The number of analyzed genomes from each fungal group is presented next to the fungal guild classification. The green box plots report mean and standard deviation of the peptidase content of genomes that belonging to plant pathogenic fungi, and the purple box plots report mean and standard deviation of the peptidase content of animal pathogenic fungi, and orange box plots for fungal parasite genomes, with p-values.

3.6. Appendix for Chapter 3

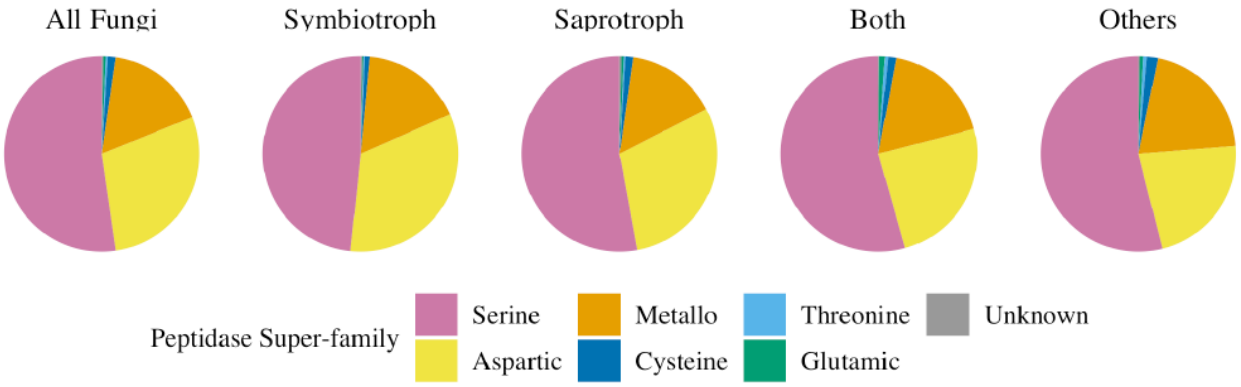


Figure S3.1. Relative abundance of secreted peptidase super-families in 612 fungal genomes (aspartic, cysteine, glutamic, metallo-, serine, threonine, and unknown peptidase super-families) and in different fungal ecological guilds. Different colors represent different peptidase super-families.

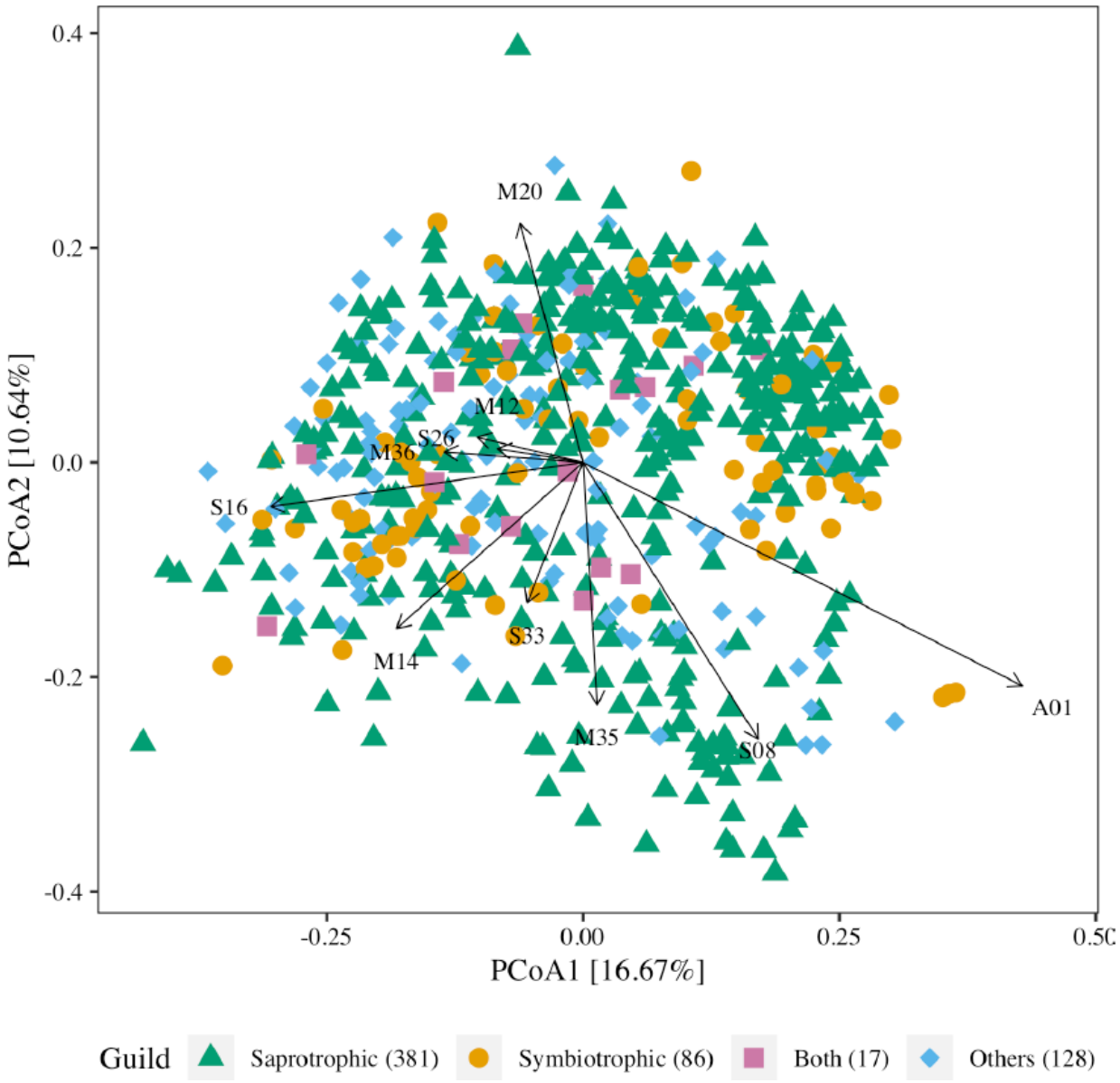


Figure S3.2. Principal coordinate analysis of fungal genomes based on Bray-Curtis dissimilarities of proportions of secreted peptidase families encoded in fungal genomes. Symbol shapes and colors are coded by fungal guild. Vectors lengths are scaled relative to the correlation of individual peptidase families with the two axes shown (Pearson's correlation). The composition of secreted peptidase genes of *Fungi* varied significantly based on their guild ($p = 0.001$, F-statistic =3.599)

Table S3.1. Phylogenetic signal strength of secreted peptidase families across 612 fungal genomes. Significance of clustering is based on the Fritz and Purvis index (D) of each peptidase family trait (presence or absence of genes). Estimated D value defines whether the secreted peptidase distribution would follow "strongly clumped" ($D \leq 0$), or "Brownian-like evolutionary" ($0 < D < 1$) or "Random" distribution ($D \geq 1$). consenTRAIT (τD), a phylogenetic metric that evaluates the sequence similarity of clusters of sharing trait organisms. τD for any singleton entry (trait that only presents in one genome) is scored by half the distance to the nearest node (assuming the likelihoods to find a neighbor organism with/without the trait are equal)

Peptidase Family	Estimated D	τD	Probability of E(D) resulting from no (random) phylogenetic structure	Probability of E(D) resulting from Brownian phylogenetic structure	Genomes
A01	0.717	1.577	0.447	0.322	610
A11	-0.389	0.057	0.050	0.657	3
A22	0.372	0.104	0.000	0.018	66
C01	0.747	0.052	0.165	0.039	9
C02	0.964	0.053	0.486	0.039	6
C12	0.797	0.020	0.198	0.021	10
C13	0.676	0.069	0.000	0.000	237
C14	0.660	0.055	0.222	0.163	5
C15	0.355	0.101	0.242	0.396	3
C19	0.828	0.065	0.035	0.000	36
C26	1.297	0.027	0.810	0.001	7
C40	-0.042	0.070	0.000	0.589	30
C44	0.488	0.063	0.000	0.044	19
C48	1.250	0.055	0.760	0.008	6
C54	-0.421	0.046	0.181	0.597	2
C56	0.911	0.040	0.274	0.000	19
C57	3.419	0.012	0.644	0.306	1
C59	0.567	0.037	0.336	0.329	3
C65	-27.059	0.137	0.008	0.989	1
C69	0.657	0.019	0.030	0.024	14
C85	4.240	0.003	0.703	0.268	1

G01	0.510	0.048	0.000	0.000	74
M01	0.452	0.071	0.000	0.053	23
M03	0.758	0.066	0.008	0.000	35
M04	0.368	0.080	0.000	0.120	20
M06	0.764	0.036	0.155	0.026	11
M10	0.586	0.051	0.000	0.021	20
M12	0.330	0.101	0.000	0.003	408
M13	0.693	0.072	0.003	0.001	26
M14	0.296	1.577	0.000	0.015	384
M15	3.660	0.057	0.609	0.346	1
M16	0.923	0.077	0.231	0.000	36
M17	0.023	0.088	0.000	0.497	7
M18	1.530	0.078	0.968	0.001	8
M19	0.411	0.080	0.000	0.008	72
M20	0.438	0.082	0.000	0.000	266
M22	0.822	0.049	0.110	0.001	20
M23	1.026	0.030	0.576	0.000	17
M24	0.727	0.044	0.000	0.000	65
M28	0.586	1.577	0.000	0.003	582
M35	0.588	0.085	0.000	0.000	116
M36	0.768	0.066	0.000	0.000	155
M38	0.656	0.031	0.015	0.014	18
M41	0.493	0.063	0.000	0.005	53
M43	0.648	0.069	0.000	0.000	61
M48	0.711	0.070	0.250	0.117	6
M49	0.433	0.065	0.040	0.216	7
M50	1.643	0.074	0.916	0.006	5
M54	1.414	0.010	0.635	0.213	2
M67	0.685	0.101	0.085	0.046	10
M76	-1.638	0.016	0.000	0.914	3
M77	12.725	0.016	0.840	0.110	1
M79	-0.216	0.086	0.002	0.637	5
S01	0.006	0.108	0.000	0.488	120
S03	-14.789	0.004	0.155	0.842	1
S08	-0.426	1.577	0.029	0.682	609
S09	0.640	1.577	0.000	0.000	450
S10	0.861	1.577	0.485	0.277	610
S12	0.644	0.058	0.000	0.000	127
S14	1.196	0.071	0.742	0.005	8

S15	-0.290	0.001	0.000	0.673	5
S16	0.504	0.054	0.000	0.005	45
S24	-1.468	0.070	0.000	0.944	4
S26	0.373	0.091	0.000	0.001	297
S28	-0.063	1.577	0.000	0.652	519
S33	0.669	0.082	0.000	0.000	155
S37	-0.426	0.039	0.508	0.506	1
S41	1.119	0.058	0.608	0.021	6
S49	2.452	0.030	0.847	0.065	2
S51	-0.160	0.043	0.001	0.626	6
S53	-0.425	1.577	0.000	0.999	505
S54	0.874	0.058	0.200	0.001	18
S58	-6.075	0.020	0.376	0.606	1
S59	0.325	0.082	0.206	0.389	3
T01	0.894	0.059	0.275	0.002	14
T02	0.555	0.017	0.009	0.061	13
T03	0.616	0.050	0.000	0.000	67
U48	0.661	0.086	0.000	0.000	77
U69	-10.751	0.033	0.605	0.368	1

3.7. References

Adamczyk, B., Kitunen, V., and Smolander, A. (2008). Protein precipitation by tannins in soil organic horizon and vegetation in relation to tree species. *Biol. Fertil. Soils* 45, 55–64. doi:10.1007/s00374-008-0308-0.

Adamczyk, B., Kitunen, V., and Smolander, A. (2009). Polyphenol oxidase, tannase and proteolytic activity in relation to tannin concentration in the soil organic horizon under silver birch and Norway spruce. *Soil Biol. Biochem.* 41, 2085–2093. doi:10.1016/j.soilbio.2009.07.018.

Agerer, R. (2001). Exploration types of ectomycorrhizae. *Mycorrhiza* 11, 107–114. doi:10.1007/s005720100108.

Beeck, M. O. D., Troein, C., Peterson, C., Persson, P., and Tunlid, A. (2018). Fenton reaction facilitates organic nitrogen acquisition by an ectomycorrhizal fungus. *New Phytol.* 218, 335–343. doi:10.1111/nph.14971.

Berlemont, R., and Martiny, A. C. (2013). Phylogenetic distribution of potential cellulases in *Bacteria*. *Appl. Environ. Microbiol.* 79, 1545–1554. doi:10.1128/AEM.03305-12.

Caldwell, B. A. (2005). Enzyme activities as a component of soil biodiversity: A review. *Pedobiologia* 49, 637–644. doi:10.1016/j.pedobi.2005.06.003.

Crowther, T. W., Boddy, L., and Hefin Jones, T. (2012). Functional and ecological consequences of saprotrophic fungus–grazer interactions. *ISME J.* 6, 1992–2001. doi:10.1038/ismej.2012.53.

Dunaevskii, Y. E., Gruban, T. N., Belyakova, G. A., and Belozerskii, M. A. (2006). Extracellular proteinases of filamentous fungi as potential markers of

phytopathogenesis. *Microbiology* 75, 649–652. doi:10.1134/S0026261706060051.

Fritz, S. A., and Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conserv. Biol. J. Soc. Conserv. Biol.* 24, 1042–1051. doi:10.1111/j.1523-1739.2010.01455.x.

Gorka, S., Dietrich, M., Mayerhofer, W., Gabriel, R., Wiesenbauer, J., Martin, V., et al. (2019). Rapid Transfer of Plant Photosynthates to Soil Bacteria via Ectomycorrhizal Hyphae and Its Interaction With Nitrogen Availability. *Front. Microbiol.* 10. doi:10.3389/fmicb.2019.00168.

Grigoriev, I. V., Cullen, D., Goodwin, S. B., Hibbett, D., Jeffries, T. W., Kubicek, C. P., et al. (2011). Fueling the future with fungal genomics. *Mycology* 2, 192–209. doi:10.1080/21501203.2011.584577.

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704. doi:10.1093/nar/gkt1183.

Hartley, B. S. (1960). Proteolytic enzymes. *Annu. Rev. Biochem.* 29, 45–72. doi:10.1146/annurev.bi.29.070160.000401.

Hu, G., and Leger, R. J. S. (2004). A phylogenomic approach to reconstructing the diversification of serine proteases in fungi. *J. Evol. Biol.* 17, 1204–1214. doi:10.1111/j.1420-9101.2004.00786.x.

Jombart, T., Balloux, F., and Dray, S. (2010). adephylo: new tools for investigating the phylogenetic signal in biological traits. *Bioinformatics* 26, 1907–1909.

doi:10.1093/bioinformatics/btq292.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi:10.1093/nar/gkw290.

Martin, F., and Nehls, U. (2009). Harnessing ectomycorrhizal genomics for ecological insights. *Curr. Opin. Plant Biol.* 12, 508–515. doi:10.1016/j.pbi.2009.05.007.

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

Marzluf, G. A. (1997). Genetic regulation of nitrogen metabolism in the fungi. *Microbiol Mol Biol Rev* 61, 17–32.

Monod, M., Capoccia, S., Léchenne, B., Zaugg, C., Holdom, M., and Jousson, O. (2002). Secreted proteases from pathogenic fungi. *Int. J. Med. Microbiol.* 292, 405–419. doi:10.1078/1438-4221-00223.

Mori, H., and Ito, K. (2001). The Sec protein-translocation pathway. *Trends Microbiol.* 9, 494–500.

Muszewska, A., Stepniewska-Dziubinska, M. M., Steczkiewicz, K., Pawlowska, J., Dziedzic, A., and Ginalski, K. (2017). Fungal lifestyle reflected in serine protease repertoire. *Sci. Rep.* 7. doi:10.1038/s41598-017-09644-w.

Nagy, L. G., Tóth, R., Kiss, E., Slot, J., Gácser, A., and Kovács, G. M. (2017). Six Key Traits of Fungi: Their Evolutionary Origins and Genetic Bases. *Microbiol. Spectr.* 5. doi:10.1128/microbiolspec.FUNK-0036-2016.

Nannipieri, P., and Paul, E. (2009). The chemical and functional characterization of

soil N and its biotic components. *Soil Biol. Biochem.* 41, 2357–2369. doi:10.1016/j.soilbio.2009.07.013.

Nehls, U. (2008). Mastering ectomycorrhizal symbiosis: the impact of carbohydrates. *J. Exp. Bot.* 59, 1097–1108. doi:10.1093/jxb/erm334.

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 20, 241–248. doi:10.1016/j.funeco.2015.06.006.

Nguyen, T. T. H., Myrold, D. D., and Mueller, R. S. (2019). Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny and Habitat. *Front. Microbiol.* 10. doi:10.3389/fmicb.2019.00413.

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47, D259–D264. doi:10.1093/nar/gky1022.

Northup, R. R., Yu, Z., Dahlgren, R. A., and Vogt, K. A. (1995). Polyphenol control of nitrogen release from pine litter. *Nature* 377, 227. doi:10.1038/377227a0.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2018). *vegan: Community Ecology Package*. Available at: https://CRAN.R-project.org/package=vegan.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Issac, N., et al. (2018). caper: Comparative Analyses of Phylogenetics and Evolution in R. Available at: ftp://mirror.ac.za/cran/web/packages/caper/vignettes/caper.pdf.

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi:10.1093/bioinformatics/bty633.

Pellitier, P. T., and Zak, D. R. (2018). Ectomycorrhizal fungi and the enzymatic liberation of nitrogen from soil organic matter: why evolutionary history matters. *New Phytol.* 217, 68–73. doi:10.1111/nph.14598.

Petersen, T. N., Brunak, S., Heijne, G. von, and Nielsen, H. (2011). SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi:10.1038/nmeth.1701.

R. Core Team (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015, Vienna, Austria*. ISBN 3-900051-07-0. Available: http://www. R-project. org/(1.12. 2015).

Rao, M. B., Tanksale, A. M., Ghatge, M. S., and Deshpande, V. V. (1998). Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* 62, 597–635.

Rawlings, N. D., and Barrett, A. J. (1993). Evolutionary families of peptidases. *Biochem. J.* 290 (Pt 1), 205–218.

Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632. doi:10.1093/nar/gkx1134.

Rineau, F., Stas, J., Nguyen, N. H., Kuyper, T. W., Carleer, R., Vangronsveld, J., et al. (2016). Ectomycorrhizal Fungal Protein Degradation Ability Predicted by Soil

Organic Nitrogen Availability. *Appl. Environ. Microbiol.* 82, 1391–1400. doi:10.1128/AEM.03191-15.

Root, R. B. (1967). The Niche Exploitation Pattern of the Blue-Gray Gnatcatcher. *Ecol. Monogr.* 37, 317–350. doi:10.2307/1942327.

Semenova, T. A., Dunaevsky, Y. E., Beljakova, G. A., Borisov, B. A., Shamraichuk, I. L., and Belozersky, M. A. (2017). Extracellular peptidases as possible markers of fungal ecology. *Appl. Soil Ecol.* 113, 1–10. doi:10.1016/j.apsoil.2017.01.002.

Shah, F., Rineau, F., Canbäck, B., Johansson, T., and Tunlid, A. (2013). The molecular components of the extracellular protein-degradation pathways of the ectomycorrhizal fungus Paxillus involutus. *New Phytol.* 200, 875–887. doi:10.1111/nph.12425.

Silva, B. A. da, Santos, A. L. S. dos, Barreto-Bergter, E., and Pinto, M. R. (2006). Extracellular Peptidase in the Fungal Pathogen Pseudallescheria boydii. *Curr. Microbiol.* 53, 18–22. doi:10.1007/s00284-005-0156-1.

Soanes, D. M., Alam, I., Cornell, M., Wong, H. M., Hedeler, C., Paton, N. W., et al. (2008). Comparative Genome Analysis of Filamentous Fungi Reveals Gene Family Expansions Associated with Fungal Pathogenesis. *PLOS ONE* 3, e2300. doi:10.1371/journal.pone.0002300.

Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., et al. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 108, 1028–1046. doi:10.3852/16-042.

Talbot, J. M., Bruns, T. D., Smith, D. P., Branco, S., Glassman, S. I., Erlandson, S., et al. (2013). Independent roles of ectomycorrhizal and saprotrophic communities in

soil organic matter decomposition. *Soil Biol. Biochem.* 57, 282–291. doi:10.1016/j.soilbio.2012.10.004.

Tedersoo, L., May, T. W., and Smith, M. E. (2009). Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* 20, 217–263. doi:10.1007/s00572-009-0274-x.

Theron, L. W., and Divol, B. (2014). Microbial aspartic proteases: Current and potential applications in industry. *Appl. Microbiol. Biotechnol.* 98, 8853–8868. doi:10.1007/s00253-014-6035-6.

Treseder, K. K., and Lennon, J. T. (2015). Fungal Traits That Drive Ecosystem Dynamics on Land. *Microbiol. Mol. Biol. Rev.* 79, 243–262. doi:10.1128/MMBR.00001-15.

Wu, J.-W., and Chen, X.-L. (2011). Extracellular metalloproteases from *Bacteria*. *Appl. Microbiol. Biotechnol.* 92, 253. doi:10.1007/s00253-011-3532-8.

Zimmerman, A. E., Martiny, A. C., and Allison, S. D. (2013). Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J.* 7, 1187–1199. doi:10.1038/ismej.2012.176.

Chapter 4: Contribution of Different Catalytic Types of Peptidases to Soil Proteolytic Activity

Trang T. H. Nguyen*

Markus Kleber, David D. Myrold

To be submitted in

Soil Biology and Biochemistry

xx March 2019

4.1. Abstract

Soil organic nitrogen is largely composed of proteinaceous material, hence, the extracellular peptidases that are widely produced by microorganisms play a critical role in the recycling of soil organic nitrogen. But why do microbes produce such a variety of functionally different peptidases? In theory, this could be an adaptation to substrate heterogeneity, but it may also be an adaptation to specific soil conditions. Here we characterized the contribution of different catalytic types, or classes of peptidases, present in soil with the intent to determine if their relative contributions would vary as a function of soil properties. Potential peptidase inhibitors were screened and their concentrations were optimized to work effectively in soil. Total proteolytic activity was partitioned among several peptidase classes by adding class-specific inhibitors to the peptidase assay. Using Pepstatin A, EDTA, PMSF, and E64, we were able to discriminate between aspartic, metallo-, serine, and cysteine peptidases, respectively. We found that diverse peptidases were active and contributed to the proteolytic activity in soil. Extracellular peptidase profiles varied among different soils and were associated with soil chemical and microbial properties. Metallopeptidases contributed 30-50% of the soil proteolytic activity in all soils. Higher relative contribution of metallopeptidase activity was found in soils with higher pHs, reflecting metallopeptidase neutral pH preference. Serine peptidases were only detected in Douglas-fir associated soils (10-20% of total proteolytic activity) but not red alder soils. Aspartic peptidase relative activity correlated with the fungal:bacterial ratios. Our results lend support to the view that microorganisms modify their activities to optimize resource utilization in response to soil and other environmental factors.

4.2. Introduction

Proteinaceous material, including proteins, peptides, and amino acids, is the most abundant form of soil organic nitrogen (N) (Nannipieri and Paul, 2009; Schulten and Schnitzer, 1997). In order to obtain carbon (C) and N for cell growth, microbes produce different kinds of proteolytic enzymes to break down various proteinaceous compounds (Nguyen et al., 2019). Each microbe possesses a unique set of extracellular proteolytic enzymes to perform this function. How and when to produce these peptidases and release them to the environment depends not only on the nutrient demand for growth but also on the energy costs associated with peptidase synthesis and secretion, which must be weighed against the risk that other organisms may benefit from the activity of the peptidase once it is released into the environment (Allison, 2005). To be economic with this energy investment, each microbe has evolved to produce the peptidases that can function well in the soil environments where the microbe lives (Nguyen et al., 2019). For example, selection may favor peptidases with broad protein substrate specificity, or that optimally function under given soil conditions (pH, temperature, etc.). The activity of these extracellular peptidases would be expected to peak at the corresponding native soil conditions, and this performance should reflect the adaptation of the microbes living in that soil. More broadly, such selectivity would allow the microorganisms optimize the cycling of organic N in soils.

Consequently, there should be an inherent correlation between soil properties as the driving force on one hand and microbial proteolytic genetic potential on the other hand, but the existence of such a correlation has not yet been conclusively

demonstrated. For the purpose of this paper, we define microbial proteolytic potential as the cumulative activity from different classes of proteolytic enzymes present under the native soil environment. Here we hypothesize that environment shapes the microbial enzymatic potential and determines which enzymes are produced and most active under different circumstances.

In order to test this supposition, measurements are needed to characterize the presence and activities of different classes of peptidases. Unfortunately, there are limitations to the methods that have been used to measure the microbial proteolytic potential of a soil under native environmental conditions. In the most common peptidase assay, the sample is incubated with protein substrate in pH 8.0 buffer at 50-60^{o}C for 1-2 h to measure the total potential proteolytic enzyme activity (Geisseler and Horwath, 2008; Ladd and Butler, 1972). At this temperature, we argue that some soil enzymes might be inactivated. Secondly, the traditional way to distinguish the four major peptidase classes in environmental samples was to extract proteolytic enzyme fractions and then use peptidase inhibitors that specifically deactivate each class (Hayano, 1993; Kamimura and Hayano, 2000; Watanabe and Hayano, 1995). Soil enzymes are generally extracted with neutral phosphate buffer, concentrated by salting out with ammonium sulfate, and inhibitors added to the enzyme extract to determine the types of peptidases present (Hayano, 1993; Kamimura and Hayano, 2000; Watanabe and Hayano, 1995). However, the efficiency of this enzyme extraction is not known. Another approach to study the potential microbial contribution to soil protein degradability was to isolate and culture specific groups of dominant soil bacteria and fungi (Bach and Munch, 2000; Rineau et al., 2016; Semenova et al., 2017; Shah et al.,

2013), and then apply peptidase inhibitors to classify extracellular proteolytic enzymes produced by specific organisms (Morihara, 1974). By contrast, only a few attempts have been made to use peptidase inhibitors directly in the analysis of soils or aquatic samples: Hoppe et al. (1988) noted that individual inhibitors decreased L-aminopeptidase activity in brackish water of the Baltic Sea from 10 to 68%. Renella et al. (2002) found that a cocktail of peptidase inhibitors decreased total peptidase activity in a calcareous soil by 50%, and an extended abstract by Kumar et al. (2004) reported that peptidase inhibitors added to soil could suppress N mineralization during incubations of soil in the laboratory. All these different approaches have merits and shortcomings as discussed above that have so far not been resolved.

Because of these methodological limitations, we developed an optimized methodological approach to characterize the activities of different classes of extracellular peptidases in soils. Our strategy was to use several classes of peptidase inhibitors to determine the proportion if total activity of different catalytic classes of peptidases in soils, while incubating soils near their normal pH and temperature. Peptidases are classified by their catalytic mechanisms into seven main classes: asparagine, aspartic, cysteine, glutamic, metallo-, serine, and threonine peptidases (Rawlings et al., 2018). We recently demonstrated that aspartic, cysteine, metallo-, and serine peptidases, comprise >75% of all extracellular peptidases encoded in archaea, bacteria and fungi (Nguyen et al., 2019, Chapter 3). Therefore, we focused on using inhibitors that specifically inhibited these four dominant classes of proteolytic enzymes. This method was modified and improved from previous peptidase inhibitor studies in order to be applied directly to soils under ambient conditions (longer

incubation period of 24-48 hours at 20^{o}C). The specificity and effective concentrations of inhibitors were tested with pure peptidases and peptidase-supplemented soils to confirm the inhibitory efficiency before being applied to soils alone.

After developing the method, we addressed our original research question by measuring activities of the different classes the proteolytic enzymes from four soils in Oregon representing a gradient of biochemical properties and evaluating correlations between these enzymatic potentials and different soil properties. The underlying reasoning is as follow: If there is a predictable change in the activity of a proteolytic enzyme in response to the gradient of a certain soil property, we conclude that the activity of the enzyme is a function of the soil property. If the activity of a proteolytic enzyme does not correlate with any soil property, or the combination of soil properties, we conclude that the enzyme activity is independent from the soil characteristics. In other words, the enzyme might be produced constitutively regardless of the soil conditions as an energy opportunity cost to obtain C and N.

4.3. Materials and Methods

4.3.1. Site characteristics and soil sampling

Soil was collected from four forests in Oregon: Cascade Head, H.J. Andrews, McDonald-Dunn, and Black Butte (Table 4.1). Most of the method development used the soils from Cascade Head and H.J. Andrews. Cascade Head Experimental Forest is located 1.6 km from the Pacific Ocean in the Oregon Coast Range at 300 m elevation. It is a highly productive site that receives about 2450 mm annual precipitation. The

silty clay loam soil is an isomesic Typic Fulvudand, high in N content with a low pH (Table 1). H.J. Andrews Experiment Forest is located on the western slopes of the Cascade Mountains at 700 m elevation. It is lower in productivity than Cascade Head and receives about 2290 mm annual precipitation, with some snow in the winter. The loam soil is a mesic Andic Humudept with much lower C and N content but higher pH than Cascade Head soils (Table 4.1). At each study location, plots of Douglas-fir (*Pseudotsuga menziesii*) and red alder (*Alnus rubra*) were established about 30 years ago (Radosevich et al., 2006). Soils under these trees differ in their bacterial and fungal communities, N cycling characteristics, and soil properties (Table 4.1; Boyle et al., 2008; Lu et al., 2015; Yarwood et al., 2010). The McDonald-Dunn forest is located on the eastern foothills of the Oregon Coast Range at 380 m elevation and receives 1080 mm of rainfall annually. The silty clay loam soil is a mesic Humic Haploxerept. Soils were collected under 35-year-old Douglas-fir. Black Butte is an ancient cinder cone located on the east side of the Cascade Mountains and receives 400 mm annual precipitation, mostly as snow. The loamy sand soil is a frigid Humic Vitrixerand and was collected at 1240 m elevation under Douglas-fir.

After removing the organic layer, the top 15 cm of mineral soil was collected. Five to eight samples were taken from each plot, composited into a single sample, sieved at 4 mm to remove rock fragments and plant debris, and stored at 4^{o}C until used.

4.3.2. Peptidase assays

The total, potential proteolytic activity of the soils was measured using the assay developed by (Ladd and Butler, 1972) with some modifications. In the original

protocol, the assay was done at 50ºC for 30 to 60 min; however, we were interested in measuring proteolytic activity at temperatures closer to those found in the field. We did the assays at room temperature (~20ºC) and extended the incubation to 24 or 48 h. Conventionally, the enzyme activity is calculated based on the difference in tyrosine accumulation between samples with and without sodium caseinate at the end of the incubation period. As noted in the original protocol, sometimes negative values of enzyme activity are calculated because of high phenolic background or other unknown factors from the soil matrix (Ladd and Butler, 1972). To overcome this potential problem when using multiple soils that varied greatly in their properties, we calculated proteolytic activity as the difference in the absorbance of 700 nm at the end (i.e., 24 or 48 h) and immediately after sodium caseinate was added (0 h). As in the original protocol, proteolytic activity was measured using the Folin-Ciocalteau reagent, which reacts with aromatic moieties (e.g., tyrosine, tryptophan, and phenylalanine). Proteolytic activity was expressed as μmol tyrosine produced per g of dry soil per day. Details of our protocol can be found in the Appendix 4.6.1.

4.3.3. Effectiveness of peptidase inhibitors and inhibitor concentrations using pure peptidases

We used pure proteolytic enzymes to determine the specificity and effective concentrations of each inhibitor. Pure proteolytic enzymes were chosen from the four most abundant peptidase super-families based on the extracellular peptidase-coding genes of archaeal, bacterial, and fungal genomes: aspartic, cysteine, metallo-, and serine peptidases (Nguyen et al., 2019, Chapter 3). Three of these four pure peptidases were derived from microorganisms commonly found in soils: aspergillopepsin I

aspartic peptidase from *Aspergillus saitoi*, thermolysin metallopeptidase from *Geobacillus stearothermophilus*, and subtilisin A serine peptidase from *Bacillus licheniformis* (Table S4.1). Papain, extracted from *Carica papaya*, was used as the model cysteine peptidase. These enzymes were prepared as concentrated stock solutions using water and stored at -20^{o}C until used. The working solution of each enzyme was diluted from the stock solution using sodium acetate buffer (0.1 M, pH 5) to best resemble the pH range of our soils (3.6-5.1). The concentration for the working solution of each enzyme was determined so that their proteolytic activity *in vitro* was within the range of the proteolytic activity found in our soils (30-100 μmol tyrosine g^{-1} dry soil d^{-1}). Detailed information about these commercial products and their working concentrations is included in Table S4.1.

Based on product information (Sigma-Aldrich, St. Louis, MO), we chose epoxysuccinyl-L-leucylamido (4-guanidino) butane (E64), α-toluenesulfonyl fluoride (PMSF), Pepstatin A, and ethylenediaminetetraacetic acid (EDTA) to inhibit cysteine, serine, aspartic, and metallopeptidases, respectively. Inhibition of serine peptidases by PMSF and cysteine peptidases by E64 is accomplished by the irreversible, covalent binding of the inhibitor to the nucleophilic active site of a peptidase (Hoppe et al., 1988; Rineau et al., 2016; Shah et al., 2013). Metallopeptidases are inhibited with EDTA, which chelates the metal bound in the peptidase's active site (Bach and Munch, 2000; Wu and Chen, 2011). Aspartic peptidase can be inhibited by pepstatin A, which has a statin group that binds to the two aspartyl residues of the active site with very high affinity (Hoppe et al., 1988; Marciniszyn et al., 1976; Rich et al., 1985; Theron and

Divol, 2014). We found some discrepancies in the expected behavior and efficacy of these inhibitors, however.

First, EDTA unexpectedly stimulated cysteine peptidase activity (data not shown). This meant that if EDTA was added to inhibit metallopeptidase activity, the net change would reflect both inhibited metallopeptidase activity and some unknown induced cysteine peptidase activity. Because E64, the cysteine peptidase inhibitor, had no effect on metallopeptidase activity, we used EDTA and E64 together to effectively inhibit both cysteine and metallopeptidases, and E64 to inhibit only cysteine peptidases. The effective final concentrations were 1.82 μM for E64 and 0.45 mM for EDTA (Table S4.2).

Second, PMSF and Pepstatin A were originally prepared in isopropanol and ethanol as suggested by the supplier (Sigma-Aldrich, St. Louis, MO); however, we found that these alcohols unexpectedly interfered by partially inhibiting the activity of some pure enzymes, especially cysteine peptidase (data not shown). Dimethyl sulfoxide (DMSO) was chosen as an alternative carrier for PMSF and Pepstatin A because it did not affect the proteolytic activity of the pure enzymes. Highly concentrated stock solutions for these two inhibitors were prepared in isopropanol due to solubility constraints, but working solutions were prepared using DMSO to minimize interference from the small amount of remaining alcohols. The final concentrations that were shown to be effective were 0.09 mM for PMSF and 0.91 μM for Pepstatin A (Table S4.2). Details about the effective peptidase inhibitor concentrations are provided in Table S4.2.

To test the effectiveness and specificity of each peptidase inhibitor on the model peptidases, we designed an experiment to cover all possible combinations of four peptidases (aspartic, cysteine, metallo-, and serine peptidase) and six peptidase inhibitors (control, E64, E64+EDTA, Pepstatin A, PMSF, and an inhibitor cocktail). The inhibitor cocktail was the combination of all four peptidase inhibitors at the same final concentration, as with the single inhibitor treatment (Table S4.2). We added 25 μL of the model peptidases at their specific working concentrations (Table S4.1) to 500 μL of sodium acetate 0.5M buffer solution at pH 5 in a 2-mL Eppendorf tube. Peptidase inhibitors (25 μL) at their specific working concentrations were added to the same tube to meet their equivalent final concentration (Table S4.2). After addition of inhibitors, the tubes were shaken for 1 h at room temperature (20^{o}C) to allow the inhibitors to interact with the peptidases before adding sodium caseinate and initiating the assay (Appendix 4.6.1). The proteolytic activity was measured after 24 h of incubation at room temperature.

4.3.4. Confirmation of the effectiveness and duration of inhibition in peptidase-supplemented soils

To validate the efficacy of the inhibitors in soils, exogenous enzymes were added, hypothetically to double the soil proteolytic activity. If a specific inhibitor blocked this augmented activity in the presence of soil, we considered that the inhibitor would be effective in soil alone.

We used soil from the Douglas-fir treatment at the H.J. Andrews as our tested soil. There were six peptidase treatments (soil control without added peptidase, soil + aspartic peptidase, soil + cysteine peptidase, soil + metallopeptidase, soil + serine

peptidase, and soil + four-peptidase mixture) and seven peptidase inhibitor treatments (water control, DMSO control, Pepstatin A, E64, E64+EDTA, PMSF, and an inhibitor cocktail of all four inhibitors). The peptidase inhibitor controls (water and DMSO) provided a measure of uninhibited proteolytic activity. The DMSO control accounted for DMSO as a carrier for the Pepstatin A, PMSF, and the inhibitor cocktail treatments. Tested combinations are listed in Table 2. There were three independent replicates of each combination of enzyme and inhibitor for each sampling time.

We added moist soil (10 g of dry weight equivalent) to 100 mL deionized water in a glass bowl, stirred for 20 min to prepare the homogenized soil slurry, and added 500 μL of the soil slurry to a 2-mL Eppendorf tube. Eppendorf tubes were weighed before and after the soil slurry was added to precisely calculate the amount of soil in each tube. Water or peptidase solutions (25 μL at working concentration, Table S4.1) were added to the Eppendorf tube. Tubes were shaken for 1 h to allow the added enzymes to mix well and interact with the soil in the slurry. Next, 25 μL of specific inhibitors at working concentration (Table S4.2) were added and tubes were shaken for 1 h to allow the inhibitors to react before sodium caseinate was added. We measured the samples at 0, 24, and 48 h after the sodium caseinate addition following the modified proteolytic assay protocol in the Appendix 4.6.2.

Because each enzyme treatment was independent, we only compared the data within the same enzyme treatment, not across enzyme treatments. For instance, in the experiments with soil + metallopeptidase, we compared the proteolytic activity between the soil + metallopeptidase control activity with the soil + metallopeptidase activity under different the peptidase inhibitor treatments to determine whether a single

inhibitor was effective against a specific group of peptidases. This experiment allowed us to answer several questions: Was there a difference in proteolytic activity between the water and DMSO controls? Was proteolytic activity linear throughout the 48-h incubation, and did the proteolytic activity differ in the presences of peptidase inhibitors? Were inhibitors both specific and effective when used in soil?

4.3.5. Potential interference of other organic compounds or abiotic cleavage on soil proteolysis

When adding the inhibitor cocktail to the soils, we observed some residual proteolytic activity. This activity could be from: (1) peptidase super-families with different catalytic residues from those of the four-major peptidase super-families (e.g., threonine, glutamic, or asparagine peptidases), (2) the production of other organic compounds that react with the Folin-Ciocalteau reagent, or (3) abiotic peptide cleavage by the soil mineral matrix. Due to the lack of competitive peptidase inhibitors for threonine, glutamic, and asparagine peptidases, the first hypothesis was not tested. We tested the potential interference from the latter two mechanisms.

To determine whether organic compounds other than aromatic amino acids contributed to the measured activity in our peptidase assay, we used an alternative method with OPAME (o-phthaldialdehyde and β-mercaptoethanol) (Darrouzet-Nardi et al., 2013; Jones et al., 2002) to quantify total free amino acids (adapted protocol is described in the Appendix 4.6.2). We correlated the accumulation of total free amino acids (expressed as μmol leucine produced per g dry soil) with the increase in absorbance measured with the Folin-Ciocalteau assay (expressed as μmol tyrosine per g dry soil). We compared the OPAME and Folin-Ciocalteau assays using soils collected

under Douglas-fir from all four sites, which differed markedly in organic matter content (Table 1). Similar slopes for all four soils would indicate that aromatic organic compounds in soils did not interfere with our modified proteolytic activity assay.

To evaluate the possible contribution of the abiotic proteolysis, we used the fine clay (<2 μm) fraction separated from the soils sampled under Douglas-fir and red alder at Cascade Head and H.J. Andrews. The <2 μm fraction was obtained after repeated sedimentation in Atterberg cylinders and treated with hydrogen peroxide to remove all oxidizable organic matter as well as any biotic activity (Moore and Reynolds, 1997). We focused on the fine clay fraction because clay minerals can potentially fragment proteins (Reardon et al., 2018). A slurry was made with 0.5 g of the isolated clay fraction from each soil in 20 mL deionized water and otherwise followed by our modified peptidase assay. Abiotic proteolytic activity was monitored for 48 h and converted to the equivalent activity unit (μmol tyrosine per gram of dry soil per day) for comparison with whole soil assays.

4.3.6. Determination the relative pools of proteolytic enzymes in soils

The contribution of different classes of extracellular peptidase enzymes to soil proteolytic activity was measured for the soils collected under Douglas-fir and red alder at Cascade Head and H.J. Andrews. Red alder soils usually have higher nitrogen content than the Douglas-fir soils, which might result in differences in their proteolytic activity (Table 4.1). We used the modified proteolytic activity protocol with the suite of peptidase inhibitors at effective concentrations (Table S4.2). Soil slurries (10 g dry weight equivalent in 100 mL deionized water) were transferred to Eppendorf tubes, 50

μL of peptidase inhibitors were added to 500 μL of soil slurry, shaken for 1 h before sodium caseinate addition, and incubated at room temperature, with samples taken at 0, 24, and 48 h. To calculate the contribution of each peptidase class, we used the subsets of all the enzymes inhibited by each inhibitor, formulated a specific matrix based on the relationships between the enzymes and the inhibitors (Table 4.3), and estimated the contribution of each enzyme pool using least-squares regression.

4.3.7. Statistical analyses

In general, analysis of variance (ANOVA) was used to determine statistical differences among treatments. Proteolytic activity of the water and DMSO control treatments of the same soil-enzyme treatment were evaluated by ANOVA to determine whether the presence of DMSO in soils had any effect on proteolytic activity. To evaluate the changes of soil proteolytic activity between 0-24 h and 24-48 h, we calculated the activity for each time period and used Student's t-test with Hommel's adjustment for multiple comparisons between treatments (Hommel, 1988). We used ANOVA to assess whether each peptidase inhibitor efficiently inhibits their target enzyme(s) in soil. To compare all inhibitors to a single control (water or DMSO), we did multiple testing using Dunnett's test. We used Pearson correlation to explore relationships between soil properties and peptidase activities. All of statistical analyses and calculations were written and implemented in R Studio (R. Core Team, 2016).

4.4. Results

4.4.1. Effective peptidase inhibitor concentrations

To determine the efficacy of the peptidase inhibitors on pure enzymes, we compared the difference in proteolytic activity of peptidase solutions with and without inhibitors after 24 h. Pepstatin A selectively suppressed aspartic peptidase by about 85% ($p < 0.001$, Table 4.4) and did not interfere with the activity of other non-aspartic peptidases (p-values > 0.01, Table S4.3). Activity of the pure cysteine peptidase was inhibited >95% by E64 ($p < 0.001$, Table 4.4) without affecting the other three enzyme classes (p-values > 0.01, Table S4.3). The combination of E64 and EDTA inhibited both cysteine peptidase (>95%) and metallopeptidase (>90%) activity (p-values < 0.001, Table 4.4) and did not interfere with the aspartic and serine peptidase activities (p-values > 0.01, Table S4.3). Activity of the model serine peptidase was inhibited 85% by PMSF ($p < 0.001$, Table 4.4), whereas PMSF did not interfere with the activity of aspartic, cysteine, and metallopeptidases (p-values > 0.01, Table S4.3). Lastly, the inhibitor cocktail—the combination of the four peptidase inhibitors—successfully stopped 80-96% the proteolytic activity of the four model peptidases ($p < 0.0001$, Tables 4.4 and S4.3). In summary, each of the peptidase inhibitors when used at their proposed final concentration had the desired specific inhibitory effect and could therefore be further evaluated for use in soils to separate the proteolytic activity from different classes of peptidases.

4.4.2. Inhibition confirmed in peptidase-supplemented soils

Pure peptidases were added to soils in quantities to at least double the soil proteolytic activity. We evaluated if the rate of proteolysis of each individual treatment

differed between the first (0-24 h) and second day (24-48 h) of the incubation (Fig. 4.1). With only one exception (PMSF in the cysteine peptidase amended soil), there was no evidence that the rate of proteolysis changed through time (Table 4.5; Hommel adjusted p-values > 0.01). Linearity through 48 h increases the sensitivity of the assay enabling it to be used in soils with low proteolytic activity.

To evaluate the effectiveness of peptidase inhibitors on proteolytic activity, we compared the differences between each inhibitor treatment and their peptidase-supplemented control using Dunnett's test. Across all the enzyme-supplemented treatments, there was no evidence that DMSO caused any difference in the soil proteolytic activity in comparison to the water controls (p-values > 0.01, Fig. 4.1 and Table S4.4). Thus, either DMSO (used to prepare Pepstatin A and PMSF solutions) or water can serve as the control for the peptidase inhibitor treatments. Both E64 and E64 + EDTA deactivated >90% of the added-cysteine activity, reducing it nearly to the soil-alone level ($p < 0.001$, Fig. 4.1 and Table S4.4). There were no non-target effects of E64 on proteolytic activity of soils supplemented with serine or metallopeptidases (p-values > 0.01, Fig. 4.1 and Table S4.4). The combination of E64 and EDTA suppressed the metallopeptidase and cysteine peptidase supplemented treatment by >85% (p-values < 0.001, Fig. 4.1 and Table S4.4). The activity of serine peptidases was inhibited 75% by PMSF ($p < 0.001$, Fig. 4.1 and Table S4.4), and PMSF had no non-target effects on soil supplemented with cysteine peptidase ($p > 0.01$, Fig. 4.1 and Table S4.4). Pepstatin A displayed statistically significant inhibition, but was less effective that the other peptidase inhibitors, inhibiting just 35% of the activity in soils with aspartate peptidase added ($p = 0.006$, Fig. 4.1 and Table S4.4).

As expected, the inhibitor cocktail was equally, or slightly more effective, than each specific inhibitor alone for all enzyme-added treatments (p-values < 0.001, Fig. 4.1 and Table S4.4). In soil supplemented with all four peptidases, we observed that about 25% of the total soil proteolytic activity remained after adding the inhibitor cocktail (Fig. 4.1). This residual activity that could not be inhibited by any of the four peptidase inhibitors could derive from other peptidases that were not deactivated by the inhibitors used, abiotic proteolysis, or the production of other compounds that react in the Folin-Ciocalteau assay (e.g., phenols).

4.4.3. Specificity of Folin-Ciocalteau assay confirmed for proteolytic activity

To evaluate the hypothesis that the Folin-Ciocalteau assay may measure compounds other than aromatic amino acids, we used an alternate approach to quantify soil proteolytic activity, the OPAME assay, which measures total free amino acids. The proteolytic activity of four soils was measured at 0, 24, and 48 h after sodium caseinate addition using both methods. For each soil, the increase during the 48-h incubation of fluorescence measured using the OPAME reagent was linearly related to the absorbance measured using the Folin-Ciocalteau reagent (Fig. 4.2). The slopes of the relationship for the four soils did not differ significantly ($p > 0.05$, Table S4.5), but the y-intercepts did ($p < 0.05$, Table S4.5). Positive y-intercepts indicate that the Folin-Ciocalteau reagent does react with some aromatics that are not amino acids, and this background is higher at Cascade Head and H.J. Andrews. Despite this difference in the background, the similar slopes of all four soils suggest that the increased absorbance

measured by the Folin-Ciocalteau assay reflects only amino acids produced by proteolytic activity and not the production of other aromatic compounds.

4.4.4. Abiotic proteolysis is negligible

We found little evidence for abiotic proteolysis in the fine clay fractions of the two soils we tested (Fig. 4.3). The rates of proteolysis calculated by the rate of tyrosine accumulated between 0 and 48 hour were not significantly different from zero (Fig. 4.4).

4.4.5. Extracellular peptidase profiles are related to soil properties

Peptidase inhibitors were applied to the soils in order to characterize their extracellular peptidase profiles. There was a higher background for Cascade Head than H.J. Andrews soils at the initial measurement, particularly for the soil from red alder site at Cascade Head (Fig. 4.3). Proteolytic activity increased linearly through time in all treatments (Fig. 4.3, Table S4.6). The H.J. Andrews red alder had the highest total peptidase activity (12.8±0.3 μmol tyrosine g^{-1} dry soil d^{-1}), followed by Cascade Head Douglas-fir (7.8±0.4 μmol tyrosine g^{-1} dry soil d^{-1}), Cascade Head red alder (6.3±0.5 μmol tyrosine g^{-1} dry soil d^{-1}), and H.J. Andrews Douglas-fir (5.6±0.3 μmol tyrosine g^{-1} dry soil d^{-1}).

The proteolytic activity of the control, inhibitor treatments, and abiotic proteolysis were analyzed with a least squares regression model to calculate the activities of different peptidases in soils. The relative contribution of each peptidase class varied among soils, often with a large fraction being uncharacterized (Fig. 4.4). Because we determined that accumulation of non-amino acid aromatics did not occur

(Fig. 4.2) and that abiotic proteolysis was negligible (Fig. 4.3), we assume that the uncharacterized activity is from peptidases that were not affected by the inhibitors we used. Metallopeptidases were the major contributors to proteolytic activity in all soils (Fig. 4.4). In all soils, the activity of cysteine peptidases was not significantly different from zero. Interestingly, the contribution of serine peptidases varied by tree type: serine peptidase activity was not significantly different from zero in red alder soils but contributed to 7-14% activity in the Douglas-fir soils (Fig. 4.4). A high proportion of proteolytic activity remained uncharacterized among all soils but was much higher in soils from Cascade Head (61-66% compared to 20-31%) (Fig. 4.4).

4.5. Discussion

4.5.1. Substrate-induced proteolytic assay: initial control, length of incubation, and linearity

In this study, we evaluated modifications to the classic enzyme assay that is commonly used to measure proteolytic activity in soils. According to Ladd and Butler (1972), the potential proteolytic activity of a soil was defined by the difference in tyrosine concentrations of a soil slurry incubated with and without sodium caseinate at pH 8 for 1 h at 50^{o}C. Tyrosine concentrations were quantified with the Folin-Ciocalteau reagent. To better reflect native soil conditions, we modified the potential proteolytic assay by using water rather than a buffer to keep the pH close to that of the soil and incubated at a temperature closer to that of the environment, as others have done (Reiskind et al., 2011).

Ladd and Butler (1972) noted that soil organic matter sometimes interfered with the Folin-Ciocalteau reagent resulting in negative activities, presumably because the

added caseinate complexed some of the interfering organic compounds. To avoid such supposedly negative activities, we calculated proteolytic activity as the tyrosine accumulation between the time when substrate was added and the end of the assay. We further explored the non-specificity of the Folin-Ciocalteau assay, which interacts not only with aromatic amino acids, such as the tyrosine used as the standard, but also with other organic aromatic compounds. Although the Folin-Ciocalteau reagent did react with other compounds in most of the soils we tested (positive y-intercept, Fig. 4.2, Table S4.5), the similar slopes obtained when comparing the Folin-Ciocalteau and OPAME assays (Fig. 4.2, Table S4.5) suggested little production of interfering compounds during the incubation. This finding indicates that the modified Folin-Ciocalteau assay works well to quantify substrate-induced proteolytic activity in soils.

Non-linearity has sometimes been encountered in soils assayed with the caseinate-supplemented measure of potential proteolytic activity (Ladd and Butler, 1972; Reiskind et al., 2011; Vranova et al., 2013). We sometimes observed a lag in activity during the first 4-10 h with the H.J. Andrews Douglas-fir soil (data not shown), potentially due to the complexity of sodium caseinate as a large protein composed of many amino acid units and the lower temperature we used. Increasing the incubation time to 48 h usually resulted in linear accumulation of products, using either the Folin-Ciocalteau or OPAME assays (Fig. 4.2). The longer incubation also increased the sensitivity of the assay.

4.5.2. Modifications to measure activities of different peptidase classes

The interference of alcohols to some peptidase activities (data not shown) led us to the use of DMSO as a substitute carrier solution for dissolving PMSF and Pepstatin A because it is a good solvent for both polar and nonpolar compounds. We found that soils amended with DMSO at the concentration that we used (4.7%) had the same proteolytic activity as those amended with water (Fig. 4.1, Table S4.4). This agrees with Obayashi et al. (2017), who suggested that DMSO at final concentration of < 5% would not cause a significant influence on proteolytic enzyme activity in seawater.

Because the degree of peptidase inhibition is dependent on the inhibitor concentration, incubation time, temperature, and pH (Bach and Munch, 2000; Pérez-Lloréns et al., 2003), we tested the effect of peptidase inhibitors on solutions of pure peptidases under the pH and temperature conditions relevant to the soils we used and in peptidase-supplemented soils. To our knowledge, these validating tests have not been done in prior studies that have used peptidase inhibitors (Hayano, 1993; Hoppe et al., 1988; Kamimura and Hayano, 2000; Kumar et al., 2004; Renella et al., 2002; Rineau et al., 2016; Shah et al., 2013; Watanabe and Hayano, 1995). The linearity of the remaining activity in peptidase-supplemented soils during the course of incubation suggested that inhibition was effective and constant during the 48-h incubation (Fig. 4.1, Table 4.5 and S4.4). This indicated that the inhibitors worked well and continuously deactivated the specific enzymes in the soils.

We noted that the Pepstatin A did not seem to perform very well at the concentration that was used to suppress the supplemented aspartic peptidase activity.

It is possible that Pepstatin A interacted more strongly with the soil matrix than did the other inhibitors, decreasing its effective concentration, which might be overcome by using Pepstatin A at higher concentrations. It should be noted, however, that the absolute inhibition by Peptstatin A of about 4 μmol tyrosine g^{-1} dry soil d^{-1} in soil amended with aspartate peptidase exceeded the 1.5 μmol tyrosine g^{-1} dry soil d^{-1} attributable to the aspartate peptidases in the non-amended soil. Thus, the Pepstatin A concentrations we used may have been sufficient for use in non-amended soils. Interestingly, in pure aspartic peptidase experiment, the inhibition efficacy of Pepstatin A was about 80-86% of the aspartic peptidase activity. However, with the presence of soils, this inhibition efficacy was only 60-65%. This observation was more likely due to the interaction between the soil matrix (especially phenolic groups) and the Pepstatin A protein molecules, which potentially deactivated part of the activity of Pepstatin A. We suggested future studies to implement this reaction at higher Pepstatin A concentrations to overcome this complex interaction.

We also observed some remaining activity in the presence of the inhibitor cocktail (Figs. 4.1 and 4.2), as did Renella et al. (2002). This residual activity might arise at least partially from abiotic proteolysis by soil minerals (Reardon et al., 2018), but we did not measure any proteolytic activity associated with the minerals in the purified fine clay fractions isolated from our soils (Figs. 4.2 and 4.4). The uncharacterized activity might be mainly associated with biotic proteolysis, either from the underperformance of Pepstatin A as aspartic inhibitor, or from the inhibitor-inaccessible enzymes bound to clays or organic matter in soils (Renella et al., 2002)

and/or less-dominant extracellular peptidase groups (such as threonine, glutamic, asparagine) that we did not characterize.

4.5.3. Biotic proteolysis: Correlation with soil biochemical properties

First, we found a strong positive correlation (r=0.983) between the total peptidase activities and the gross ammonification rates measured by Yarwood et al. (2010) for these same four soils (Fig. 4.5A). Proteinaceous compounds are broken down into oligomeric peptides, which are then taken up by plants or microorganisms, or further mineralized into ammonium and nitrate. The rate of protein decomposition therefore is considered to be the rate-limiting step of nitrogen cycling in terrestrial ecosystems (Schimel and Bennett, 2004). The positive correlation between proteolytic activity and gross ammonification rate confirms that biotic proteolysis is the bottleneck that regulates the N cycling in soils.

The peptidase inhibitor method we developed provides a way to identify and quantify the contribution of four different peptidase classes to a soil's potential proteolytic activity. This method allowed us to discover several distribution patterns of different catalytic groups of peptidases across our four soils that have a range of soil biochemical properties (Fig. 4.4).

4.5.4. Metallopeptidase activity dominant and reflective of soil pH

Metallopeptidases were the most active class of identified peptidases in each of our soils, and their activity correlated with gross ammonification (r=0.834); activities of the other classes of peptidases were not correlated. The relative contribution of metallopeptidases was strongly correlated with organic C content (r=-0.962), clay

content (r=-0.958), and pH (Fig. 4.5B, r=0.937), all of which were highly correlated (r>0.997) with each other. Interactions with clay minerals or organic matter may inactivate metallopeptidase activity, accounting for the negative correlation, but it is more likely that the positive correlation between relative metallopeptidase activity and pH reflects the neutral pH optimum associated with this class of peptidases (Wu and Chen, 2011). Prior studies of peptidases extracted from soil or produced by bacterial isolates have also found metallopeptidases abundant (Bach and Munch, 2000; Hayano, 1993; Kamimura and Hayano, 2000; Watanabe and Hayano, 1995). The majority of extracellular metallopeptidases are endopeptidases, which cleave the internal peptide linkages of proteins (Wu and Chen, 2011). Therefore, metallopeptidases likely interact with other classes of extracellular peptidases to fully degrade proteins in soil.

4.5.5. Extracellular aspartic peptidase activity correlates to fungal:bacteria ratio

Aspartic peptidases contributed the second highest activity across our four soils (Fig. 4.4). Their relative contribution to total peptidase activity was correlated most highly (Fig 4.5C, r=0.834) with the ratio of fungal to bacterial marker phospholipid fatty acids measured by (Boyle et al., 2008). Aspartic peptidases were consistently found to be the prevalent peptidases in the extracellular peptidase repertoires of most fungal species according to several fungal pure culture and genomic studies (Caldwell, 2005; Rao et al., 1998; Rineau et al., 2016; Shah et al., 2013; Theron and Divol, 2014; Vranova et al., 2013, Chapter 3). Aspartic peptidase activity typically peaks between pH 3-5 (Theron and Divol, 2014), which spans the pH range of our soils.

4.5.6. Serine peptidases associated with soils under different tree species

We did not detect a significant contribution of serine peptidases to the biotic proteolytic activity in red alder soils, but they were active in Douglas-fir soils (Fig. 4.4). Serine peptidases usually display neutral and alkaline pH optima (Rao et al., 1998), but their activity was not correlated to pH (or other soil physical and chemical properties) of our soils ($r<0.4$). We found a weak correlation ($r=0.734$) between the serine peptidase activity and the fungal:bacterial ratios in these soils (Boyle et al., 2008), but at the kingdom level serine peptidases are highly abundant in archaea, bacteria, and fungi (Nguyen et al., 2019, Chapter 3). Thus, we hypothesize the distinction in serine peptidase activity between soils under the two tree species may arise from the composition of their microbial communities, most likely the fungal communities because it is well known that Douglas-fir and red alder establish different ectomycorrhizal symbioses (Horton et al., 2005; Kennedy et al., 2014).

4.5.7. Uncharacterized proteolytic activity

Cysteine peptidase activity was below the detection limit in all of our soils, whereas the biotic activity from uncharacterized peptidases contributed a large proportion to the soil proteolytic activity (Fig. 4.4). More than 50% of the coastal soil peptidase activity was uncharacterized. This bias between the coastal soils and H.J. Andrews soils might be partly from the fact that some inhibitors did not perform as well in coastal soils compared to our H.J. Andrews soils, which were used as the model soils in our study to validate the inhibitor protocol. Thus, it may be necessary to determine the effective inhibitor concentrations for a given soil.

This uncharacterized activity correlated positively with clay content ($r=0.913$) and archaeal abundance ($r=0.985$). We found no abiotic proteolytic activity associated

with the isolated clay fraction; however, Renella et al. (2002) suggested that uncharacterized peptidase activity might come from peptidases trapped on the mineral surfaces and inaccessible to inhibitors. It is difficult to imagine that the relatively small peptidase inhibitors (MW from 174 to 686) would be excluded when the added protein substrate (sodium caseinate, MW>10,000) is not. The possible connection between the uncharacterized biotic peptidase activity and archaeal population is intriguing. Archaea in marine sediments have been shown to encode extracellular cysteine, serine, and metallopeptidases, where they are thought to play a major role in protein turnover (Lloyd et al., 2013). These peptidase classes should have been inhibited, although extracellular glutamic and threonine peptidase-coding genes have been found in a small number of complete archaeal genomes in Nguyen et al. (2019).

First, this research confirmed the importance of microbial peptidases in regulating N mineralization. Second, we showed that the absolute and relative contributions the peptidases with four different catalytic mechanisms vary among soils of different physical, chemical, and microbial characteristics. This observation raises the hypothesis that soil microorganisms, individually or collectively, respond to their surrounding environments by investing in the peptidases that optimize their activity. Third, we have developed a functional protocol for the use of specific peptidase inhibitors in the investigation of protease activities in soil that can be used to further evaluate this hypothesis. Measuring peptidase activity profiles would complement research that uses genomic, transcriptomic, and proteomic approaches to study the regulation of peptidases in the environment.

Table 4.1. Soil chemical properties (Boyle et al., 2008; Lu et al., 2015) and soil texture determined by hydrometer method. Data are mean ± standard error (n=3).

Site	Tree species	Total C (mg/g)	Total N (mg/g)	C:N	pH	Clay (%)	Silt (%)	Sand (%)
Black Butte	Douglas-fir	16±2.5	0.8±0.0	20.7±3.9	7.0	3.8±1.2	16.6±3.1	79.6±1.9
Cascade Head	Red alder	144±10.4	9.2±0.9	15.7±0.5	3.6	31.1±3.7	45.3±5.0	23.6±7.7
	Douglas-fir	128±2.3	6.7±0.1	19.1±0.0	4.1	27.5±3.6	50.8±0.4	21.7±3.9
H.J. Andrews	Red alder	82±12.1	3.4±0.2	23.8±2.4	5.1	17.3±6.4	40.4±2.2	42.3±6.0
	Douglas-fir	90±7.5	2.7±0.1	33.2±2.1	5.0	18.8±2.6	40.5±4.8	40.7±6.7
McDonald-Dunn	Douglas-fir	54±0.8	3.7±0.0	14.6±0.0	6.0	40.0±4.9	42.8±2.9	17.3±3.0

Table 4.2. Peptidases and peptidase inhibitors used in peptidase-supplemented soil experiments. Tested combinations have "x". Generally, inhibitors were used against only their target peptidase unless non-target interactions had been observed in pure enzyme studies.

Peptidase Addition	**Peptidase Inhibitor Treatment**						
	Water	DMSO	Pepstatin A	E64	E64+EDTA	PMSF	Inhibitor Cocktail
None added	x	x	x	x	x	x	x
Aspartic	x	x	x				x
Cysteine	x	x		x	x	x	x
Metallo-	x	x		x	x		x
Serine	x	x		x		x	x
Enzyme mixture	x						x

Table 4.3. The matrix used for least squares regression to calculate for the relative activity of each peptidase group in soil, based on the relationships between the peptidases and the peptidase inhibitors. One indicates that proteolytic activity in not inhibited by the treatment; zero indicates the treatment completely inhibits proteolytic activity.

Peptidase		**Peptidase Inhibitor Treatment**						
		No inhibitor	Pepstatin A	E64	E64+EDTA	PMSF	Inhibitor Cocktail	Clay fraction (<2 μm)
Peptidase Super-family (Biotic proteolysis)	Aspartic	1	0	1	1	1	0	0
	Cysteine	1	1	0	0	1	0	0
	Metallo-	1	1	1	0	1	0	0
	Serine	1	1	1	1	0	0	0
	Uncharacterized	1	1	1	1	1	1	0
Abiotic proteolysis	Abiotic activity	1	1	1	1	1	1	1

Table 4.4. The inhibitory efficiency of peptidase inhibitors on model peptidases. The numbers represent the percentage of remaining activity after the peptidase inhibitor additions. Bold represent significant inhibition ($p \leq 0.001$; Supplemental Table S4.3).

Model peptidase	Peptidase Inhibitor Treatment				
	Pepstatin A	E64	E64+EDTA	PMSF	Inhibitor Cocktail
Aspartic	**16±3%**	83±9%	103±5%	96±0%	**20±5%**
Cysteine	95±5%	**3±1%**	**4±2%**	86±1%	**4±1%**
Metallo-	77±16%	73±22%	**9±3%**	71±15%	**16±2%**
Serine	96±4%	93±3%	102±12%	**15±4%**	**16±4%**

Table 4.5. Significance of proteolytic activity during the first and second 24-h periods of the incubation for each individual peptidase-supplemented soil combination. The p-values for these multiple Welch's two sample t-test analyses were adjusted using the Hommel procedure. Data are p-values for comparisons between the two time periods, with p-values of $\alpha \leq 0.01$ shown in bold. When the p-values are large (> 0.01), it indicates that proteolytic activity was constant during the entire 48-h incubation.

Peptidase Addition	**Peptidase Inhibitor**						
	Control	DMSO	Pepstatin A	E64	E64+EDTA	PMSF	Inhibitor Cocktail
Aspartic peptidase	0.531	0.642	0.792	x	x	x	0.616
Cysteine peptidase	0.792	0.313	x	0.776	0.776	**0.009**	0.090
Metallopeptidase	0.792	0.792	x	0.792	0.792	x	0.792
Serine peptidase	0.615	0.232	x	0.792	x	0.792	0.792
Enzyme mixture	0.792	x	x	x	x	x	0.792

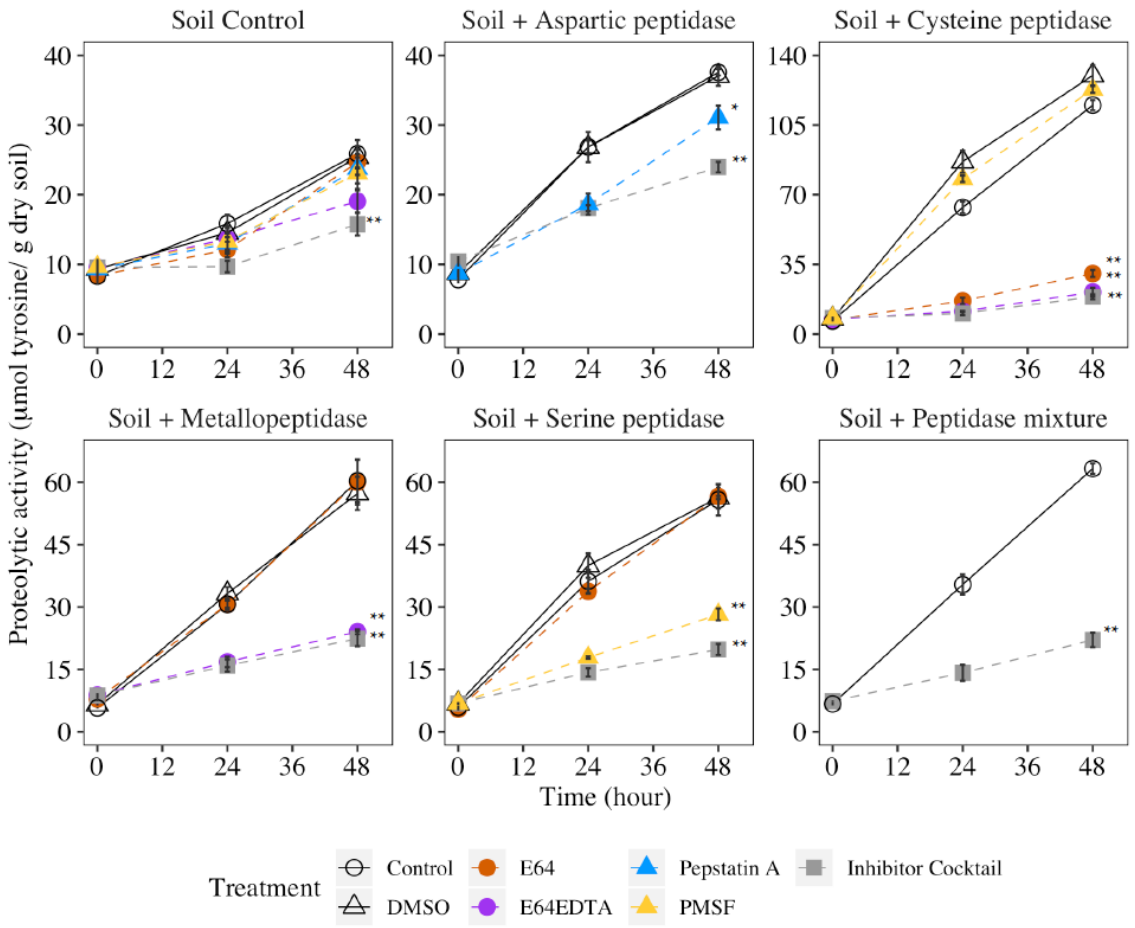


Figure 4.1. Proteolytic activities over 48 h in peptidase-supplemented soil in response to different peptidase inhibitor treatments (Table 2). Data points are means with standard error bars (n=3). Symbol shapes, colors, and line types represent different peptidase inhibitor treatments. Asterisks denote significant difference from control at 48 h (*) for $0.001 < p < 0.01$; (**) for $p < 0.001$).

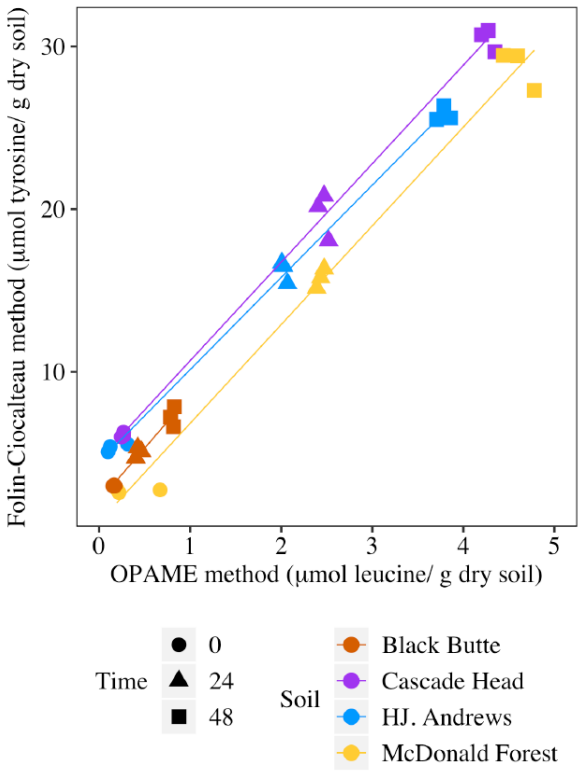


Figure 4.2. Correlation between proteolytic activity measured by the OPAME assay (total free amino acids, leucine as standard) and by the Folin-Ciocalteau assay (using tyrosine as standard). Soil proteolytic activities from four different locations were monitored over 48 hours (n=3). Slopes for individual sites were not statistically different ($p \leq 0.05$).

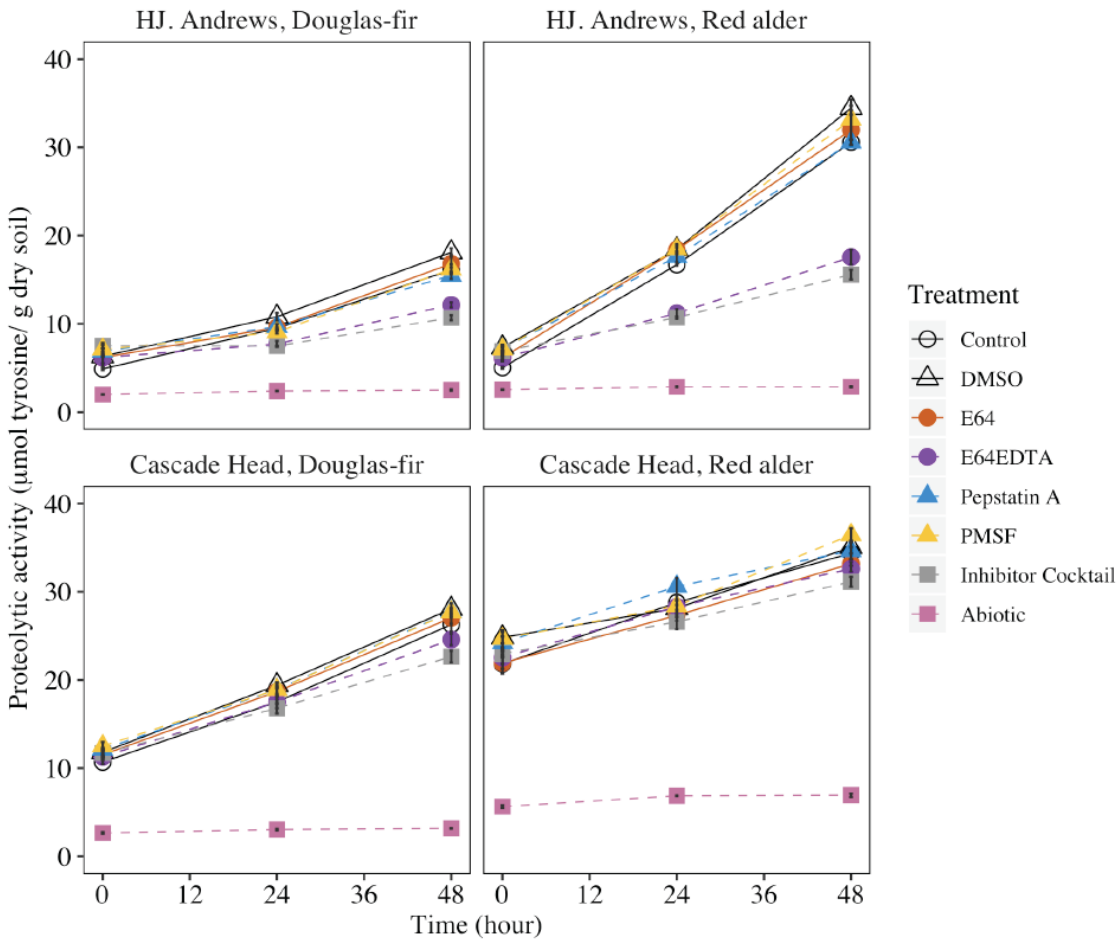
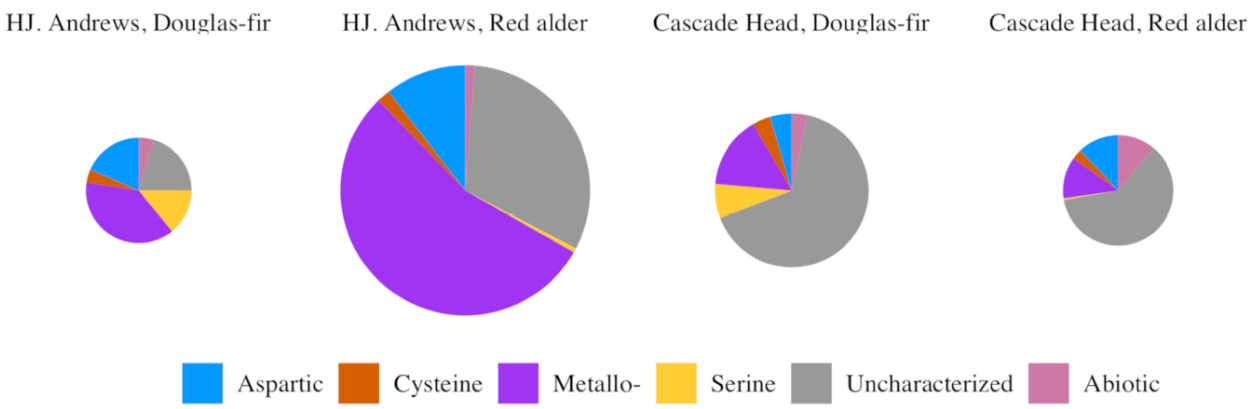


Figure 4.3. Proteolytic activities over 48 h in soils from two different ecoregions (Cascade Head and H.J. Andrews Experimental Forests) and two tree species (Douglas-fir and red alder). Data points are means with standard error bars (n=3). Symbol shapes, colors, and line types represent different peptidase inhibitor treatments.



(A)

(B)

Proteolytic Activity	HJDF	HJRA	CHDF	CHRA
Aspartic	**1.03±0.25**	**1.38±0.49**	0.37±0.26	**0.70±0.33**
Cysteine	0.23±0.32	0.23±0.61	0.31±0.33	0.18±0.41
Metallo-	**2.12±0.37**	**7.15±0.72**	**1.23±0.39**	0.71±0.48
Serine	**0.79±0.25**	0.08±0.49	**0.59±0.26**	0.03±0.33
Uncharacterized	**1.13±0.37**	**4.13±0.72**	**5.30±0.39**	**3.55±0.48**
Abiotic	0.26±0.28	0.17±0.53	0.27±0.29	0.60±0.36

Figure 4.4. Biotic and abiotic sources of proteolytic activity from four different soils were calculated using least square regression. The pie charts (A) represent the relative abundance of each peptidase family in each of the four soils, with the size of each chart proportional to the total proteolytic activity. The table (B) shows proteolytic activity (mean ± standard error, μmol tyrosine g^{-1} dry soil d^{-1}) for different peptidase pools in each soil. The statistical significance of each pool size to zero (n=3) are shown in bold (p-values of $\alpha \leq 0.05$).

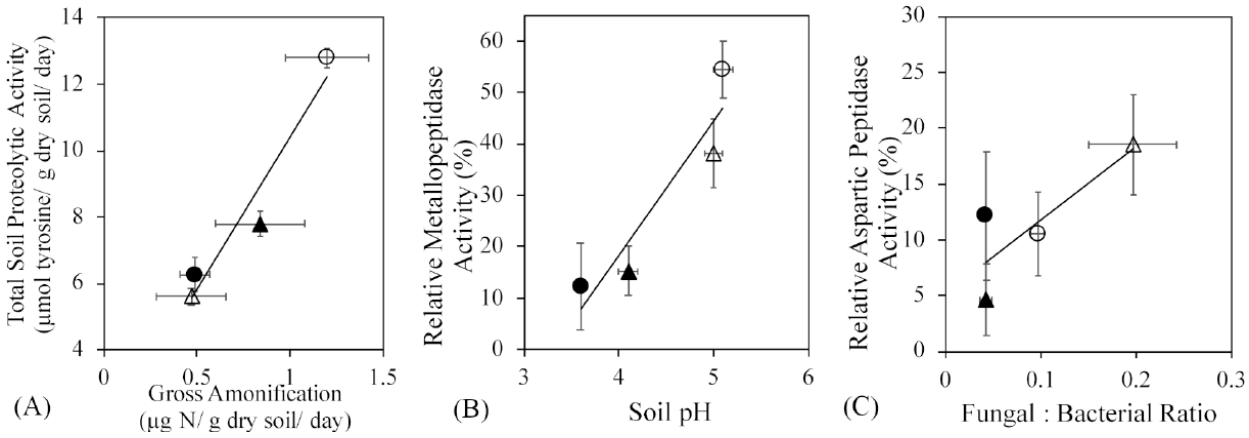


Figure 4.5. Correlation between some soil properties and different components of soil proteolytic activity of four different soils: (A) total soil peptidase activity (µmol tyrosine g^{-1} dry soil d^{-1}) vs. gross ammonification (µg N g^{-1} dry soil d^{-1}), (B) relative metallopeptidase activity vs. soil pH, and (C) relative aspartic peptidase activity vs. fungal:bacterial PLFA ratio. Different points represent the means of different soils (solid = Cascade Head soils, open = H.J. Andrews, Circle = red alder, Triangle = Douglas-fir) with standard error bars.

4.6. Appendix for Chapter 4

4.6.1. Folin-Ciocalteau Total Proteolytic Activity Assay

This assay is modified from Ladd and Butler (1972)

Chemical preparation

2% Casein Solution

2 g Casein Na salt

1 g Na azide

100 ml DI water

15% Tricholoracetic acid (TCA)

7.5 g TCA

50 ml DI Water

3-Fold diluted Folin-Ciolcalteu's

Phenol Reagent

8 ml Folin-Ciolcalteu's Phenol Reagent

16 ml DI water

1.4 M Na_2CO_3

37.1 g Na_2CO_3

250 ml DI water

2000 µM Tyrosine

36.238 mg

100 ml DI water

Tube preparation

Label two identical sets of Eppendorf tubes (2-mL) then weigh and record the weight of the tubes. The first set of tubes is for storing the sample during the incubation, the second set of tubes is for transferring the supernatant sample after the protein precipitation step using TCA. More details are below.

Soil slurry preparation

1. Weigh 3 g soil (dry weight equivalent) into a glass bowl and add 30 mL water into the glass bowl with a stirring bar. Then stir the bowl on top of the stirring plate for about 15 minutes.

2. Aliquot 500 µL soil slurry into Eppendorf tubes with three technical replicates at least per sample. Cut the tip of the pipet (1000 µL pipet, cut about a few millimeter). Use the same tip for each soil slurry. Use the permanent marker to mark how deep the pipet tip should go down in the slurry for consistency in drawing samples.

3. Weigh the tubes after adding soil slurry then calculate the exact amount of soil for each assay using this equation: [(Tube+Slurry) – (Tube)]/11, assuming that the slurry is homogenous with the soil:water ratio is 3:30

Add casein substrate and start the incubation

Add substrate into each Eppendorf tube (500 µL casein 2%). Incubate all of the tubes at room temperature (20^{o}C) in the dark room. Samples will be collected at 0, 24 and 48 hours.

Stop the proteolysis and extract the supernatant for tyrosine measurement

Stop the enzyme activity 500 µL TCA 15% into each tube to stop the reaction. Shake the tubes for about 15 minutes to allow TCA precipitate all the leftover proteins in the solution. Centrifuge the tubes using a mini-centrifuge machine for 5 minutes at 13,200 rpm. Transfer 500 µL of the supernatant in each tube into new set of Eppendorf tubes (same label).

Colorimetric assay to measure tyrosine concentration

Standard solutions: Using 2000 μM Tyrosine stock solution then serially dilute two-fold to low concentration (usually making standard curve from 1000 μM to ~7 μM Tyrosine and 0 μM—just water)

Color reagent: Make fresh 3-fold Folin-Ciolcalteu's Phenol Reagent (by diluting the original solution that came as in the bottle by three times with miliQ water). When ready, add 250 μL FC phenol solution into each tube. Continue adding 750 μL 1.4M Na_2CO_3 (sodium carbonate) into each tube to change the pH and start developing the color. Close the tube, Shake well, then incubate in dark for 20 minutes at room temperature (20^{o}C)

Colorimetric reading: Aliquot 200 μL sample from each tube to a 96-well microtiter plate (make sure that it follows the plate layout that you set up), including the standard solutions. Use the colorimetric reader at the wavelength of 700 nm. The enzyme activity is calculated as the rate of tyrosine accumulation between 0 and 24/48 hour incubation based on the estimated tyrosine concentration.

Reference

Ladd, J.N., Butler, J.H.A., 1972. Short-term assays of soil proteolytic enzyme activities using proteins and dipeptide derivatives as substrates. Soil Biol. Biochem. 4, 19–30. https://doi.org/10.1016/0038-0717(72)90038-7

4.6.2. OPAME Total Free Amino Acid Analysis

This protocol is adapted to measure soil proteolytic activity

Reagents

0.02 M pH 9.5 Borate Buffer:

Dissolve 3.81 g sodium tetraborate decahydrate (MW=381.37, B-3545, Sigma-Aldrich) in 500 mL milliQ H_2O; adjust to pH 9.5 with 10 M NaOH (40 g in 100 ml) or KOH (it will take ~ 2 mL). Stored at 4°C.

OPAME Concentrate:

- Dissolve 10 mg of o-phthaldialdehyde (OPA; P-0657, Sigma-Aldrich) in 1 ml HPLC-grade methanol. Typically, the OPA dissolves in about a minute with moderate shaking.
- Once the OPA is dissolved, add 20 µl of β-mercaptoethanol (ME; M6250, Sigma-Aldrich) to the OPA-methanol in a fume hood and shake to mix. This reagent is light sensitive – store in an amber bottle. Long-term storage of the reagent is possible in the dark at 4°C with further addition of β-mercaptoethanol every 48-72 hours.

Color Reagent: Add OPAME concentrate to 40 mL Borate buffer in an amber bottle. Let stand 2 hours to overnight to reduce background fluorescence (caused by intermediate species produced during reaction of OPA and ME)

Soil sample preparation

Preparation: Label all set of tubes and weigh tube before adding soil slurry. Record the numbers.

1. Make soil slurry from 3 g soil (dry weight equivalent) and 30 mL milliQ water and stir for 15 minutes.

2. Draw 500 µL soil slurry into each Eppendorf tube (3 technical replicates for each treatment). Weigh each tube before and after adding soil slurry and record this number to estimate exactly the amount of soil that goes into each tube.

5. Add 500 µL Casein into the tube and start the incubation at room temperature (20^{o}C) in the dark room. Data will be collected for 0, 25 and 48 hour (separate set of tubes).

6. Add TCA 15% to each of the tube to stop the proteolysis. Shake the tubes for 15 minutes to allow the protein precipitation to happen.

7. Centrifuge the tubes to separate the precipitated protein pellets and the supernatant (to be use as sample for the amino acid analysis).

Sample Analysis

Soil+OPAME: 20 µL of sample or standard (Leucine/NH_4^+, Glycine would also work) and 200 µL of Color reagent are combined in the wells of a 96-well plate and read after 1 hour (Darrouzet-Nardi et al., 2013) on a fluorometric microplate reader with an excitation wavelength of 340 nm and an emission wavelength of 450 nm.

Soil+Buffer: Blank readings can be made by adding 20 µL of sample to 200 µL of Borate buffer containing no OPAME reagent and measuring the fluorescence as described above.

Standards: Standards of Leucine and NH_4Cl (or NH_4NO_3) are recommended, with concentrations bracketing those in your samples – this will have to be determined with preliminary analyses of your samples and standards. The easiest way to do it is to begin with a wide range of standard concentrations. Suggested: Leucine 0-500 µM; NH_4^+ 0-50 ppm

L-Leucine (L8000, MW=131.17, Sigma). Stock 2000 µM Leucine = 26.2 mg L-Leucine in 100 mL milliQ water

Equation: As the OPAME procedure also detects NH_4^+, this can be accounted for using the equation:

$$\textbf{Amino acids (µM)} = (O_o - B_o - A_o) / (S_o / 10)$$

$\mathbf{O_o}$ is the OPAME fluorescence reading of the sample = Soil + OPAME + Buffer (FU ~ Fluorescence Unit)

$\mathbf{B_o}$ is the blank fluorescence reading of the sample with no OPAME reagent present = Soil + Buffer (FU)

$\mathbf{S_o}$ is the fluorescence reading of a 10 µM amino acid standard, or $S_o/10$ = Slope of Leucine curve (FU/uM)

A_o accounts for the interference of NH_4^+ in the OPAME procedure. Unless the samples show a high degree of brown coloration B_o is close to zero and can largely be ignored.

$\mathbf{A_o}$ is defined as $\mathbf{A_o = (AC_o / AS_o) * AR_o}$.

$\mathbf{AC_o}$ is the NH_4^+ concentration of the sample determined separately (usually by Ammonium colorimetric assay)

$\mathbf{AS_o}$ is the NH_4^+ standard concentration (µM)

$\mathbf{AR_o}$ is the fluorescence reading of the NH_4^+ standard using the OPAME procedure.

Reference

Darrouzet-Nardi, Anthony, Mallory P. Ladd, and Michael N. Weintraub. 2013. "Fluorescent Microplate Analysis of Amino Acids and Other Primary Amines in Soils." *Soil Biology and Biochemistry* 57 (February): 78–82. https://doi.org/10.1016/j.soilbio.2012.07.017.

Jones, David L, Andrew G Owen, and John F Farrar. 2002. "Simple Method to Enable the High Resolution Determination of Total Free Amino Acids in Soil Solutions and Soil Extracts." *Soil Biology and Biochemistry* 34 (12): 1893–1902. https://doi.org/10.1016/S0038-0717(02)00203-1.

Table S4.1. Model peptidases used in this study.

Peptidase Super-family	**Sigma-Aldrich product code**	**Model peptidase**	**Source**	**Molecular Weight**	**Working concentration**
Aspartic	P2143	Aspergillopepsin I	*Aspergillus saitoi*	41.3	1.0 mg/mL
Cysteine	10108014001	Papain	*Carica papaya*	23.0	0.28 mg/mL
Metallo-	T7902	Thermolysin	*Geobacillus stearothermophilus*	34.6	0.10 mg/mL
Serine	P4860	Subtilisin A	*Bacillus licheniformis*	27.0	4.3 μg/mL

Table S4.2. Peptidase inhibitors and targeted peptidases.

Inhibitors	**Sigma-Aldrich product code**	**Targeted peptidase(s)**	**Working concentration**	**Equivalent final concentration**
Pepstatin A	P5381	Aspartic	10 μM	0.91 μM
E64	E3132	Cysteine	20 μM	1.82 μM
E64+EDTA	E3132 CAS 6381-92-6	Cysteine & Metallo-	40 μM E64 10 mM EDTA	1.82 μM E64 0.45 mM EDTA
PMSF	P7626	Serine	1.0 mM	0.09 mM
Inhibitor cocktail	All the above	Aspartic, Cysteine, Metallo-, Serine	4.0 mM PMSF 80 μM E64 40 μM Pepstatin A 20 mM EDTA	0.09 mM PMSF 1.82 μM E64 0.91 μM Pepstatin A 0.45 mM EDTA

Table S4.3. Dunnett's p-values for comparisons of differences between the peptidase inhibitor treatments and the control for each individual model peptidase. Between-treatment comparison p-values of $\alpha \leq 0.01$ in bold.

Model peptidase	Peptidase Inhibitor				
	Pepstatin A	E64	E64+EDTA	PMSF	Inhibitor Cocktail
Aspartic	**< 0.001**	0.675	1.000	0.731	**< 0.001**
Cysteine	0.043	**< 0.001**	**< 0.001**	0.035	**< 0.001**
Metallo-	0.070	0.906	**< 0.001**	0.017	**< 0.001**
Serine	0.909	0.572	0.998	**< 0.001**	**< 0.001**

Table S4.4. Dunnett's p-values for comparisons of differences between the peptidase inhibitor treatments and the control for each individual peptidase-supplemented soil combination. Between treatment comparison p-values of $\alpha \leq 0.01$ in bold.

Peptidase Addition	**Peptidase Inhibitor**					
	DMSO	Pepstatin A	E64	E64+EDTA	PMSF	Inhibitor Cocktail
Soil Control	0.950	0.465	0.963	0.012	0.072	**< 0.001**
Soil + Aspartic peptidase	0.729	**0.006**	x	x	x	**< 0.001**
Soil + Cysteine peptidase	0.015	x	**< 0.001**	**< 0.001**	0.373	**< 0.001**
Soil + Metallopeptidase	0.887	x	0.972	**< 0.001**	x	**< 0.001**
Soil + Serine peptidase	1.000	x	0.986	x	**< 0.001**	**< 0.001**
Soil + Enzyme mixture	x	x	x	x	x	**< 0.001**

Table S4.5. Summary for regression analyses for the correlation between proteolytic activity measured by the OPAME assay (total free amino acids, leucine as standard) and by the Folin-Ciocalteau assay (using tyrosine as standard). Soil proteolytic activities from four different locations were monitored over 48 hours (n=3).

Soil	Regression			Slope			Intercept		
	Adjusted R^2	**F-value**	**p**	**Value**	**SE**	**p**	**Value**	**SE**	**p**
Black Butte	0.955	171.148	< 0.001	6.473	0.495	< 0.001	2.084	0.266	< 0.001
Cascade Head	0.990	774.242	< 0.001	6.051	0.217	< 0.001	4.642	0.620	< 0.001
H.J. Andrews	0.995	1531.838	< 0.001	5.667	0.145	< 0.001	4.478	0.359	< 0.001
McDonald Forest	0.983	454.525	< 0.001	6.066	0.285	< 0.001	0.779	0.858	0.394

Table S4.6. Comparison of rates of proteolytic activity (μmol tyrosine g^{-1} dry soil h^{-1}) during the first and second 24-h periods of the incubation for each individual soil-inhibitor combination. The p-values for these multiple Welch's two sample t-test analyses were adjusted using the Hommel procedure. Data are p-values for t-test comparisons between the two time periods, with p-values of $\alpha \leq 0.05$ shown in bold. When the p-values are large (> 0.05), it indicates that proteolytic activity was constant during the entire 48-h incubation.

Soil	**Peptidase Inhibitor**						
	Control	DMSO	Pepstatin A	E64	E64EDTA	PMSF	Inhibitor Cocktail
H.J. Douglas-fir	0.206	0.545	0.493	0.588	0.545	0.105	0.063
H.J. Red alder	0.276	**0.040**	0.245	0.637	0.637	0.637	0.588
CH Douglas-fir	0.560	0.637	0.491	0.552	0.637	0.572	0.637
CH Red alder	0.637	0.637	0.637	0.637	0.545	0.553	0.637

4.7. Reference

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

Bach, H.-J., Munch, J.C., 2000. Identification of bacterial sources of soil peptidases. Biology and Fertility of Soils 31, 219–224. doi:10.1007/s003740050648

Boyle, S.A., Yarwood, R.R., Bottomley, P.J., Myrold, D.D., 2008. Bacterial and fungal contributions to soil nitrogen cycling under Douglas fir and red alder at two sites in Oregon. Soil Biology and Biochemistry 40, 443–451. doi:10.1016/j.soilbio.2007.09.007

Caldwell, B.A., 2005. Enzyme activities as a component of soil biodiversity: A review. Pedobiologia 49, 637–644. doi:10.1016/j.pedobi.2005.06.003

Darrouzet-Nardi, A., Ladd, M.P., Weintraub, M.N., 2013. Fluorescent microplate analysis of amino acids and other primary amines in soils. Soil Biology and Biochemistry 57, 78–82. doi:10.1016/j.soilbio.2012.07.017

Geisseler, D., Horwath, W.R., 2008. Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon. Soil Biology and Biochemistry 40, 3040–3048. doi:10.1016/j.soilbio.2008.09.001

Hartley, B.S., 1960. Proteolytic enzymes. Annual Review of Biochemistry 29, 45–72. doi:10.1146/annurev.bi.29.070160.000401

Häse, C.C., Finkelstein, R.A., 1993. Bacterial extracellular zinc-containing metalloproteases. Microbiological Reviews 57, 823–837.

Hayano, K., 1993. Protease activity in a paddy field soil: Origin and some properties.

Soil Science and Plant Nutrition 39, 539–546. doi:10.1080/00380768.1993.10419794

Hommel, G., 1988. A Stagewise Rejective Multiple Test Procedure Based on a Modified Bonferroni Test. Biometrika 75, 383–386. doi:10.2307/2336190

Hoppe, H.-G., Kim, S.-J., Gocke, K., 1988. Microbial Decomposition in Aquatic Environments: Combined Process of Extracellular Enzyme Activity and Substrate Uptake. Appl. Environ. Microbiol. 54, 784–790.

Horton, T.R., Molina, R., Hood, K., 2005. Douglas-fir ectomycorrhizae in 40- and 400-year-old stands: mycobiont availability to late successional western hemlock. Mycorrhiza 15, 393–403. doi:10.1007/s00572-004-0339-9

Jones, D.L., Owen, A.G., Farrar, J.F., 2002. Simple method to enable the high resolution determination of total free amino acids in soil solutions and soil extracts. Soil Biology and Biochemistry 34, 1893–1902.

Kamimura, Y., Hayano, K., 2000. Properties of protease extracted from tea-field soil. Biology and Fertility of Soils 30, 351–355. doi:10.1007/s003740050015

Kennedy, P., Nguyen, N., Cohen, H., Peay, K., 2014. Missing checkerboards? An absence of competitive signal in Alnus-associated ectomycorrhizal fungal communities. PeerJ 2. doi:10.7717/peerj.686

Kumar, K., Rosen, C., Russelle, M., 2004. A novel approach to regulate nitrogen mineralization in soil. In: Controlling Nitrogen Flows and Losses., in: Controlling Nitrogen Flows and Losses. Academic Publishers, The Netherlands.

Ladd, J.N., Butler, J.H.A., 1972. Short-term assays of soil proteolytic enzyme activities using proteins and dipeptide derivatives as substrates. Soil Biology and

Biochemistry 4, 19–30. doi:10.1016/0038-0717(72)90038-7

Lloyd, K.G., Schreiber, L., Petersen, D.G., Kjeldsen, K.U., Lever, M.A., Steen, A.D., Stepanauskas, R., Richter, M., Kleindienst, S., Lenk, S., Schramm, A., Jørgensen, B.B., 2013. Predominant archaea in marine sediments degrade detrital proteins. Nature 496, 215–218. doi:10.1038/nature12033

Lu, X., Bottomley, P.J., Myrold, D.D., 2015. Contributions of ammonia-oxidizing archaea and bacteria to nitrification in Oregon forest soils. Soil Biology and Biochemistry 85, 54–62. doi:10.1016/j.soilbio.2015.02.034

Marciniszyn, J., Hartsuck, J.A., Tang, J., 1976. Mode of inhibition of acid proteases by pepstatin. Journal of Biological Chemistry 251, 7088–7094.

Moore, D.M., Reynolds, R.C., 1997. X-ray Diffraction and the Identification and Analysis of Clay Minerals. Oxford University Press.

Mooshammer, M., Wanek, W., Hämmerle, I., Fuchslueger, L., Hofhansl, F., Knoltsch, A., Schnecker, J., Takriti, M., Watzka, M., Wild, B., Keiblinger, K.M., Zechmeister-Boltenstern, S., Richter, A., 2014. Adjustment of microbial nitrogen use efficiency to carbon:nitrogen imbalances regulates soil nitrogen cycling. Nature Communications 5, 3694. doi:10.1038/ncomms4694

Morihara, K., 1974. Comparative Specificity of Microbial Proteinases, in: Advances in Enzymology and Related Areas of Molecular Biology. John Wiley & Sons, Ltd, pp. 179–243. doi:10.1002/9780470122860.ch5

Nannipieri, P., Paul, E., 2009. The chemical and functional characterization of soil N and its biotic components. Soil Biology and Biochemistry 41, 2357–2369. doi:10.1016/j.soilbio.2009.07.013

Nguyen, T.T.H., Myrold, D.D., Mueller, R.S., 2019. Distributions of Extracellular Peptidases Across Prokaryotic Genomes Reflect Phylogeny and Habitat. Frontiers in Microbiology 10. doi:10.3389/fmicb.2019.00413

Obayashi, Y., Wei Bong, C., Suzuki, S., 2017. Methodological Considerations and Comparisons of Measurement Results for Extracellular Proteolytic Enzyme Activities in Seawater. Frontiers in Microbiology 8. doi:10.3389/fmicb.2017.01952

Pérez-Lloréns, J.L., Benítez, E., Vergara, J.J., Berges, J.A., 2003. Characterization of proteolytic enzyme activities in macroalgae. European Journal of Phycology 38, 55–64. doi:10.1080/0967026031000096254

R. Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015, Vienna, Austria. ISBN 3-900051-07-0. Available: http://www. R-project. org/(1.12. 2015).

Radosevich, S.R., Hibbs, D.E., Ghersa, C.M., 2006. Effects of species mixtures on growth and stand development of Douglas-fir and red alder. Canadian Journal of Forest Research 36, 768–782. doi:10.1139/x05-280

Rao, M.B., Tanksale, A.M., Ghatge, M.S., Deshpande, V.V., 1998. Molecular and biotechnological aspects of microbial proteases. Microbiology and Molecular Biology Reviews 62, 597–635.

Rawlings, N.D., 2016. Peptidase specificity from the substrate cleavage collection in the MEROPS database and a tool to measure cleavage site conservation. Biochimie, A potpourri of proteases and inhibitors: from molecular toolboxes to signaling scissors 122, 5–30. doi:10.1016/j.biochi.2015.10.003

Rawlings, N.D., Barrett, A.J., 1993. Evolutionary families of peptidases. The Biochemical Journal 290 (Pt 1), 205–218.

Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A., Finn, R.D., 2018. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Research 46, D624–D632. doi:10.1093/nar/gkx1134

Reardon, P.N., Walter, E.D., Marean-Reardon, C.L., Lawrence, C.W., Kleber, M., Washton, N.M., 2018. Carbohydrates protect protein against abiotic fragmentation by soil minerals. Scientific Reports 8, 813–813. doi:10.1038/s41598-017-19119-7

Reiskind, J.B., Lavoie, M., Mack, M.C., 2011. Kinetic studies of proteolytic enzyme activity of arctic soils under varying toluene concentrations. Soil Biology and Biochemistry 43, 70–77. doi:10.1016/j.soilbio.2010.09.014

Renella, G., Landi, L., Nannipieri, P., 2002. Hydrolase activities during and after the chloroform fumigation of soil as affected by protease activity. Soil Biology and Biochemistry 34, 51–60. doi:10.1016/S0038-0717(01)00152-3

Rich, D.H., Bernatowicz, M.S., Agarwal, N.S., Kawai, M., Salituro, F.G., Schmidt, P.G., 1985. Inhibition of aspartic proteases by pepstatin and 3-methylstatine derivatives of pepstatin. Evidence for collected-substrate enzyme inhibition. Biochemistry 24, 3165–3173. doi:10.1021/bi00334a014

Rineau, F., Stas, J., Nguyen, N.H., Kuyper, T.W., Carleer, R., Vangronsveld, J., Colpaert, J.V., Kennedy, P.G., 2016. Ectomycorrhizal Fungal Protein Degradation Ability Predicted by Soil Organic Nitrogen Availability. Applied and

Environmental Microbiology 82, 1391–1400. doi:10.1128/AEM.03191-15

Schimel, J.P., Bennett, J., 2004. Nitrogen mineralization: Challenges of a changing paradigm. Ecology 85, 591–602. doi:10.1890/03-8002

Schulten, H.-R., Schnitzer, M., 1997. The chemistry of soil organic nitrogen: A review. Biology and Fertility of Soils 26, 1–15. doi:10.1007/s003740050335

Semenova, T.A., Dunaevsky, Y.E., Beljakova, G.A., Borisov, B.A., Shamraichuk, I.L., Belozersky, M.A., 2017. Extracellular peptidases as possible markers of fungal ecology. Applied Soil Ecology 113, 1–10. doi:10.1016/j.apsoil.2017.01.002

Shah, F., Rineau, F., Canbäck, B., Johansson, T., Tunlid, A., 2013. The molecular components of the extracellular protein-degradation pathways of the ectomycorrhizal fungus Paxillus involutus. New Phytologist 200, 875–887. doi:10.1111/nph.12425

Theron, L.W., Divol, B., 2014. Microbial aspartic proteases: Current and potential applications in industry. Applied Microbiology and Biotechnology 98, 8853–8868. doi:10.1007/s00253-014-6035-6

Vranova, V., Rejsek, K., Formanek, P., 2013. Proteolytic activity in soil: A review. Applied Soil Ecology 70, 23–32. doi:10.1016/j.apsoil.2013.04.003

Watanabe, K., Hayano, K., 1995. Seasonal variation of soil protease activities and their relation to proteolytic bacteria and Bacillus spp in paddy field soil. Soil Biology and Biochemistry 27, 197–203. doi:10.1016/0038-0717(94)00153-R

Wu, J.-W., Chen, X.-L., 2011. Extracellular metalloproteases from *Bacteria*. Applied Microbiology and Biotechnology 92, 253. doi:10.1007/s00253-011-3532-8

Yarwood, S.A., Bottomley, P.J., Myrold, D.D., 2010. Soil Microbial Communities Associated with Douglas-fir and Red Alder Stands at High- and Low-Productivity Forest Sites in Oregon, USA. Microbial Ecology 60, 606–617. doi:10.1007/s00248-010-9675-9

Chapter 5: Conclusion

Trang T. H. Nguyen

This dissertation focused on determining the microbial contribution to organic nitrogen cycling in soils by characterizing the diversity and proteolytic activity of extracellular peptidases.

The first research chapter recognized the wide spectrum of extracellular peptidases across *Archaea* and *Bacteria* that follow the phylogenetic relationships and environmental habitats. The two prokaryotic kingdoms differ not only in the total numbers of secreted peptidase coding genes but also in their enzymatic complements. Generally, bacteria have more secreted peptidases per genome and possess a more diverse set of secreted peptidases than archaea, suggesting that bacteria might be more competitive in organic nitrogen acquisition compared to archaea. Serine, metallo-, and cysteine peptidases contribute to 80-90% to the peptidase coding genes across these two kingdoms. Aspartic and threonine peptidases are more enriched in archaeal species. I found the evidence that archaeal and bacterial species from different microhabitats (pH and temperature) have unique sets of secreted peptidases, indicating that the distinction in secreted peptidase complements among microorganisms might be partially driven by the optimization of enzyme catalytic reaction to specific environmental conditions. Additionally, microbial lifestyles (such as free-living vs. host-associated) have influence on the variation of secreted peptidase coding genes. Free-living bacteria commonly possess more secreted peptidases in their genomes compared to host-associated species.

The second research chapter revealed the diversity and distribution of secreted peptidases in fungal species. Serine, metallo-, and aspartic peptidases are among the most abundant catalytic peptidases across fungal kingdom, whereas, cysteine peptidases, which is more common in prokaryotic taxa, contributes only 2% to the total eukaryotic secreted peptidases. The majority of the secreted peptidase families that we analyzed follow the evolutionary model, meaning that these functional genes are shared among closely related taxa. Consequently, fungal species from *Ascomycota*, *Basidiomycota*, and *Mucoromycota* had different complements of extracellular peptidase coding genes. Fungal species from different ecological groups also varied in their peptidase repertoires. Saprotrophic fungi, well-known for high lignocellulolytic ability, possessed surprisingly few secreted peptidases compared to symbiotrophic and pathotrophic fungi. Most symbiotrophs and pathotrophs have enriched numbers of secreted peptidases in their genomes, which mostly encode for the host-associated functions, either in a mutualistic or antagonistic way. Not only fungal lifestyles but also fungal evolutionary history plays a big role in shaping their extracellular proteolytic functions. In short, I used extensive data mining to analyze the phylogenetic conservation of secreted peptidases among soil dominant microbial groups and to discover the influence of microhabitats on this diversity. This provides foundational knowledge about the complements of secreted peptidases in microorganisms within different environments, which could be very informative to incorporate with other soil metagenomic and biochemical data to elucidate the microbial functions in organic nitrogen turnover.

In the third research chapter, I found that the profile of extracellular peptidase activities belonging to different catalytic types vary among soils and correlate with both soil chemical and microbial properties. I developed the protocol that uses peptidase inhibitors to classify different peptidase types and tested it against pure enzymes and peptidase-supplemented soils before applying to natural soils. I found some evidence that metallopeptidase activity correlates with soil pH, aspartic peptidase activity correlates with fungal: bacterial ratio, and total proteolytic activity correlates with gross ammonification rate. This is in line with our assumption that soil microorganisms respond to the environments by investing in peptidases that can optimize the energy cost of the extracellular enzyme secretion. The method successfully provided a profile of soil peptidase activity, but it may be necessary to optimize the concentrations of peptidase inhibitors before applying to soils with different characteristics.

Overall, the research in this dissertation demonstrated the correlative relationship between the environmental conditions and proteolytic functions in soils. Each microbial species has a unique set of extracellular proteolytic enzymes and depending on the nature of the environments, different enzymes will be expressed in order to maximize the proteolytic reaction. In the context of soils, this will determine the rate of organic nitrogen cycling.