# Ecological and Evolutionary Dynamics of the Symbiotic Microbial Communities of Fungus-Growing Insects

by

Frank O'Neill Aylward

A dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
(Microbiology)

at the
UNIVERSITY OF WISCONSIN – MADISON
2013

Date of final oral examination: 08/01/2013
The dissertation is approved by the following members of the Final Oral Committee:

**Cameron R. Currie**, Professor, Bacteriology
**Heidi Goodrich-Blair**, Professor, Bacteriology
**Katherine D. McMahon**, Professor, Civil and Environmental Engineering, Bacteriology
**Nicole T. Perna**, Professor, Genetics
**Paul J. Weimer**, Bacteriology, USDA Diary Forage Research Center

# Dissertation Abstract

## Ecological and Evolutionary Dynamics of the Symbiotic Microbial Communities of Fungus-Growing Insects

By Frank O'Neill Aylward

Under the supervision of Professor Cameron R. Currie

at the University of Wisconsin-Madison

Microbes have fundamentally shaped the evolution of metazoans on Earth by influencing the chemistry of the planet and altering the ecology of multicellular life-forms through symbioses. Herbivores present an example in which both of these influences are prominent; often, not only do symbiotic communities of microbes convert plant biomass into usable nutrients for their host, they also play a critical role in carbon cycling that influences global nutrient conditions and the climate. In this dissertation I explore symbioses between herbivores and their symbiotic microbial communities using fungus-growing insects, especially leaf-cutter ants, as a model. In Chapter 1, I review recent research into the composition and function of the fungus gardens of leaf-cutter ants to establish the current breadth of knowledge into these symbiotic ecosystems. In Chapter 2, I present a metagenomic and metaproteomic characterization of leaf-cutter ant fungus gardens that explores the identity and potential symbiotic role of bacteria that reside in these environments. I show that similar bacterial

communities are present in the fungus gardens of different leaf-cutter ant species, and that these microbes likely play a role in nutrient biosynthesis and the provisioning of nitrogenous compounds. In Chapter 3, I present a draft genome of *Leucoagaricus gongylophorus*, the fungus cultivated by leaf-cutter ants, and use metaproteomic techniques to verify that this microbe produces a variety of enzymes for the degradation of plant biomass in leaf-cutter ant fungus gardens. Lastly, in Chapter 4, I use metagenomic techniques to compare the bacterial communities associated with leaf-cutter ant fungus gardens to those of other fungus-growing insects, including fungus-growing termites, Ambrosia beetles, mountain pine beetles, and southern pine beetles. Here I present evidence consistent with evolutionary convergence of these fungus-growing insects towards highly similar bacterial communities. The work presented in these chapters suggests that the fungus-bacteria communities cultivated by fungus-growing insects are simple, highly specialized, and have drastically affected the evolution of their host insect lineages. Moreover, due to the role of some fungus-growing insects as dominant herbivores and invasive pests of trees, this work underscores the importance of symbiotic microbial communities in shaping ecosystems around the planet.

**Acknowledgements**

The work presented here was supported by many wonderful individuals who provided guidance, mentorship, technical assistance, and helpful comments along the way. Many of these individuals were extremely patient with me over the years, and I hope they realize how much I appreciate their help. I must ask your forgiveness if you are reading this and I have forgotten to list you here, although I dearly hope that is not the case.

Firstly I must thank the professors, graduate students, post-docs, and advisors at the University of Arizona that supported me during my undergraduate education and helped me get into the wonderful microbiology program at the UW-Madison in the first place. I am indebted to Rick Michod for letting me work in his laboratory when I was just a sophomore, and to Matthew Herron for spending the time to teach me basic techniques in molecular biology and to get me started on phylogenetics. I am grateful to Howard Ochman for letting me work in his laboratory and for all of the long conversations – which were probably quite painful for him – in which I vaguely grasped at an explanation of what I was interested in studying so that he could help me find the best-fit graduate program. John McCutcheon, and John Stavrinides also graciously spent a great deal of time mentoring me in the lab, advising me on graduate schools, and encouraging me to study bioinformatics. Nancy Moran, Gaelen Burke, Kevin Vogel, and Hema Narra were also very supportive throughout my undergraduate work in the laboratory, and Olivia Mendoza provided great assistance in guiding me throughout my undergraduate coursework.

I am very grateful to have had the privilege to work with Cameron Currie over the last five years. His willingness to support the wide variety of projects on which I worked is greatly appreciated, as are his insightful ideas for experiments, his helpful comments on my writing, and

the exciting laboratory atmosphere that he created. I also greatly appreciate the support and helpful comments from my wonderful thesis committee: Heidi Goodrich-Blair, Trina McMahon, Nicole Perna, and Paul Weimer.

I have been lucky enough to work with many excellent scientists at the UW-Madison as well. I am deeply grateful to Garret Suen for helping me hone my bioinformatic skills early in my graduate education. The help Garret gave me during my first few years in the Currie lab set the tone for the rest of my research, and he has been invaluable in helping me improve my writing and teaching skills. I am also indebted to a great many past and present members of the Currie lab for their helpful comments, fruitful collaborations, and friendly advice; in particular I would like to thank Michael Poulsen, Bradon McDonald, Kirk Grubbs, Lily Khadempour, Heidi Horn, Gina Lewin, Jonathan Klassen, Sandye Adams, Alissa Hanshew, Dan Tremmel, Gabe Starrett, Joe Moeller, Becca Schmidt, and Alejandra Valenzuela. I would especially like to thank Kirk for keeping all of the computer hardware and laboratory equipment in tip-top shape while he was in the lab; he saved me from many frustrated hours taking apart my computer to find out why it wouldn't start, etc. For the field work I conducted in Panama in April-May 2010 I am grateful for the help I received from Jarrod Scott. Jarrod's field expertise was also extraordinarily helpful in various other aspects of the work presented here; without his help it would have taken considerably longer to retrieve samples similar to those analyzed in this dissertation, and some may not have been obtained at all. With regards to the field work I conducted in Panama, I am also grateful for many individuals associated with the Smithsonian Tropical Research Institute, in particular Marc Seid, Laurie Seid, Orelis Arosemena, Lil Camacho, Yara Clemons, and Marcela Paz.

Other scientists not in the Currie lab have also been very helpful. I thank Kristin Burnum-Johnson, Mary Lipton, Dick Smith, Susannah Tringe, Stephanie Malfatti, Tijana Glavina del Rio, Kerry Barry, and the entire staff of Pacific Northwest National Laboratories and the DOE Joint Genome Institute for all of our fruitful and exciting collaborations over the years. I also thank Aaron Adams and Ken Raffa for introducing me to their mountain pine beetle project, and Erin Gontang and Roberto Kolter for their wonderful collaboration on their cockroach microbiome project.

The professors and staff of the Microbial Sciences Building, the Microbiology Doctoral Training Program, the Great Lakes Bioenergy Research Center, and the Department of Bacteriology were also very helpful in ensuring that the equipment in the building and the program(s) in general continued to operate smoothly over the years. I especially appreciate all of the hard work from Cathy Davis Gray, Kathy Holtgraver, Katy France, Leah Schmid, Kari Strauss, Jenni Hart, Saragale Tucker, Gary Pine, Dean Maly, Janet Newlands, Lauren Meyers, Catherine Carter, Daniel Walker, Tim Donohue, Jo Handelsman, and Rick Gourse.

I would also like to thank the various funding agencies that provided support for my projects. The National Science Foundation, Great Lakes Bioenergy Research Center, and Department of Energy all provided funds, while Pacific Northwest National Laboratories, the DOE Joint Genome Institute, and Roche Biosciences all provided proteomic analyses or genome/metagenome sequencing that was essential for my work. I am also grateful for the funding from the Wisconsin Distinguished Graduate Fellowship from September 2012-August 2013.

Lastly, I would be remiss if I failed to thank my wonderful family for their support and encouragement over the years. My loving parents did an amazing job homeschooling me and fostering my love of nature and science from a very early age. Considering they were also working full-time managing their own restaurant, I consider this to be nothing less than a Herculean task. I am also grateful for my wonderful older siblings, who have always encouraged me to pursue my goals and continue to support me. Last but not least, I thank my soon-to-be wife Leah for her love and encouragement.

# Table of Contents

**Dissertation Abstract……………………………………………………………...………… i**
**Aknowledgements……………………………………………………………..………….. iii**
**Table of Contents……………...……………………………………..……..………… vii**
**Chapter 1: The Evolutionary Innovation of Nutritional Symbioses in Leaf-Cutter Ants…………………………………………………………………..……………. 1**
1.1 Abstract…………………………………………………...………………... 1
1.2 Introduction……...……………………………………………………………… 1
1.3 The Fungus Garden Ecosystem……………………………….................... 4
1.4 The Ancillary Gut Hypothesis……………………………………………..…… 6
1.5 Evolution of Hygiene in the Attines...……………………………………….…. 7
1.6 Plant Biomass Degradation in Fungus Gardens……………………..…………… 14
1.7 Reciprocal Adaptation of the Ant Genome..…………………………………..…… 18
1.8 Conclusions and Future Outlook...……...………………………………………… 20
1.9 References………………………………..…………………………………... 21
**Chapter 2: Metagenomic and Metaproteomic Insights into Bacterial Communities in Leaf-Cutter Ant Fungus Gardens…………………………………………….………… 32**
2.1 Abstract…………………………………………………..………………... 32
2.2 Introduction……...…………………………………………………..................…… 33
2.3 Materials and Methods……...……………………………………………...... 36
2.4 Results……………………...…………………………………………..…...……… 37
2.5 Discussion……………………......…………….…...……………….…………..... 45
2.6 References………………………………………………………………..……… 56
**Chapter 3: *Leucoagaricus gongylophorus* Produces Diverse Enzymes for Recalcitrant Plant Polymer Degradation in Leaf-Cutter Ant Fungus Gardens……..…………… 61**
3.1 Abstract…………………………………………………..………………... 61
3.2 Introduction……...……………………………………………….................…… 62
3.3 Materials and Methods……...……………………………………………...... 65
3.4 Results……………………...…………………………………....................…… 69
3.5 Discussion……………………......…………….…...…………………..……..... 73
3.6 References………………………………………………………………..……… 82
**Chapter 4: Convergence of Simple and Highly-Specific Microbiomes in Fungus-Growing Insects …………...………………...…………………………………….…… 88**
3.1 Abstract……………………………………………………………………….. 88
3.2 Introduction……...…………………………………………….................…… 88

3.3 Materials and Methods………..…………………………………………………...... 90
3.4 Results……………………...………………………………………...…..……. 97
3.5 Discussion……………………......……..…………...…………………………..... 104
3.6 References…………………………………………………………………….……. 110
**Summary and Future Outlook……………...…...……………………………………115**
**Appendix 1: Supplementary Material for Chapter 2…………………………...………..118**
Supplementary Text………………………………………………………………..…...118
**Appendix 2: Supplementary Material for Chapter 3……………………………...……..134**
Supplementary Text……………………………………………………………………..…134
Supplementary Tables and Figures………………………………………………...…...141
**Appendix 3: Mountain Pine Beetles Colonizing Historical and Naïve Host Trees are Associated with a Bacterial Community Highly Enriched in Genes Contributing to Terpene Metabolism……………………………………………………………..151**
A3.1 Abstract…………………………………………………………...……..…….151
A3.2 Introduction……...………………………………………...…………...……152
A3.3 Materials and Methods………..………………………………...…………..... 155
A3.4 Results……………………...…………………………………………..…...……158
A3.5 Discussion………………………......……..…………...…………………..…....163
A3.6 References…………………………………………………………………..…… 166
**Appendix 4: Complete Genome of *Enterobacteriaceae* sp. Strain FGI 57, a Strain Associated with Leaf-Cutter Ant Fungus Gardens……………......………………….………..172**
A4.1 Abstract…………………………………………………………………….……..172
A4.2 Genome Announcement…………………………………………………….…….172
A4.3 Nucleic Acid Accession Numbers……………………………………………..174
A4.4 References……………………………………………………………………...174
**Appendix 5: Complete Genome of *Serratia* sp. Strain FGI 94, a Strain Associated with Leaf-Cutter Ant Fungus Gardens…………………………………….…………...176**
A5.1 Abstract………………………………………………………………………... 176
A5.2 Genome Announcement………………………………………………………..176
A5.3 Nucleic Acid Accession Numbers……………………..……………………178
A5.4 References………………………………………………………...…………178

# Chapter 1: The Evolutionary Innovation of Nutritional Symbioses in Leaf-Cutter Ants

**Previously published:** Frank O. Aylward, Cameron R. Currie, Garret Suen (2012). The Evolutionary Innovation of Nutritional Symbioses in Leaf-Cutter Ants. *Insects*, 3: 41-61.

**Co-authors contributed in the following ways**: Conceived and designed this work: FOA, CRC, GS. Wrote the paper: FOA, CRC, GS.

## 1.1 Abstract

Fungus-growing ants gain access to nutrients stored in plant biomass through their association with a mutualistic fungus they grow for food. This 50 million-year-old obligate mutualism likely facilitated some of these species becoming dominant Neotropical herbivores that can achieve immense colony sizes. Recent culture-independent investigations have shed light on the conversion of plant biomass into nutrients within ant fungus gardens, revealing that this process involves both the fungal cultivar and a symbiotic community of bacteria including *Enterobacter*, *Klebsiella*, and *Pantoea* species. Moreover, the genome sequences of the leaf-cutter ants *Atta cephalotes* and *Acromyrmex echinatior* have provided key insights into how this symbiosis has shaped the evolution of these ants at a genetic level. Here we summarize the findings of recent research on the microbial community dynamics within fungus-growing ant fungus gardens and discuss their implications for this ancient symbiosis.

## 1.2 Introduction

Symbioses between microbes and metazoans are widespread in nature (1-3). Although these associations form for a variety of reasons, often the diverse metabolic capabilities of symbiotic microbes allow host organisms to occupy ecological niches that would otherwise be unavailable. These symbioses can involve one or a few symbionts, but many associations in

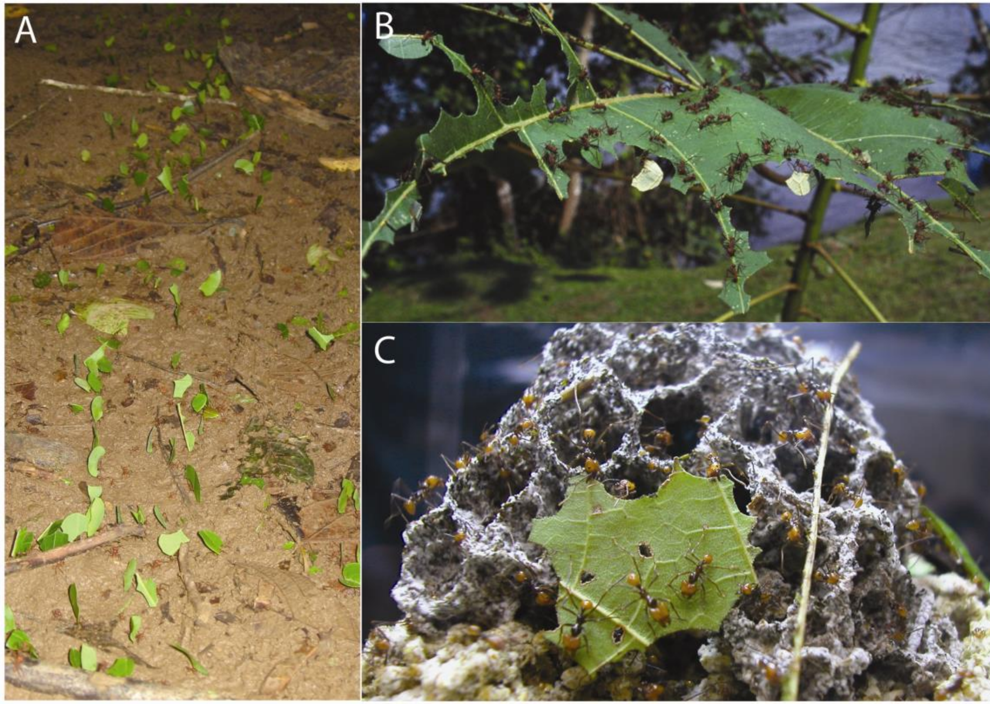



**Figure 1.1** (**a, b**) Leaf-cutter ants forage on large quantities of fresh foliar biomass. (**c**) They bring this material into their subterranean nests, where it is integrated into symbiotic fungus gardens they cultivate for food. [Photo credits: A; Jarrod J. Scott, B; Christian R. Linder, used under the GNU Free Documentation License, Version 1.2, C; Austin D. Lynch.].

nature involve complex communities of microbes. The taxonomic and physiological diversity of these communities can be massive, and research has only recently shed light on the extent to which they have shaped the evolution and ecology of metazoans (2, 4-6). Some of the best studied examples of associations between metazoans and complex microbial communities are in herbivores, where communities of microbes have been shown to be largely responsible for the deconstruction and conversion of recalcitrant plant material into nutrients for their hosts. Symbiotic microbial communities that provide this service have been shown to be associated with a vast array of hosts, including insects, mammals, and even mollusks (6-9).

The association between attine ants and their fungus gardens is a paradigmatic example of symbiosis between herbivores and microbial communities. Thought to have originated 45–50 million years ago in the Amazon basin (10), the symbiosis between these ants and their symbiotic fungus allowed for the subsequent diversification into >230 species ranging from Argentina to the shores of New Jersey in the USA (11-13). Although many species form small colonies of only a few dozen ant workers, the most derived species, the leaf-cutter ants, have evolved to become dominant Neotropical herbivores capable of foraging on up to 17% of the foliar biomass in some ecosystems (14). Moreover, the size of attine colonies varies dramatically, with the largest containing upwards of 8 million ants, 7 orders of magnitude more than colonies of the smallest species (13, 15). The diversity of these ants and their symbiosis with fungus garden communities have made them a model system for the study of the ecological and evolutionary implications of symbiosis.

In this review, we focus on recent research on the association between fungus-growing ants and their fungus gardens, with an emphasis on how the evolution and ecology of the organisms in this system have been shaped through symbiosis. We pay particular attention to how the microbial communities associated with these ants mediate the deconstruction and conversion of plant biomass into usable energy. Furthermore, we discuss how the recently sequenced genomes of two leaf-cutter ant species provide insight into how this ancient symbiosis has impacted the ants on a genetic and physiological level (16, 17). Finally, we discuss how our view of this system has changed with the recent discovery of additional microbial symbionts, and suggest future avenues of research that will yield novel insights into this complex symbiotic system.

**1.3 The Fungus Garden Ecosystem**

The most conspicuous symbiont of fungus-growing ants is the basidiomycetous fungus they grow for food. Although in the majority of species this is a lepiotaceious fungus of the genus *Leucoagaricus*, a small number of fungus-growing ant species culture a distantly-related pterulaceous group (18, 19). In most cases, the fungus is cultured by the ants on plant forage and subsequently consumed for food (Figure 1.1) (20, 21). The most derived group of fungus-farmers, referred to as the “higher attines”, culture a specific clade of the fungus, among which is the well-studied species *Leucoagaricus gongylophorus*. This fungus produces nutrient-rich hyphal swellings, called gongylidia, which nourish the queen and brood of a colony (18, 22). The less-derived groups, or “lower attines”, culture a broader range of fungal symbionts, and appear to have re-acquired cultivars from the environment multiple times in the course of their evolutionary history (18, 23).

Only recently have efforts focused on characterizing microbes in this ecosystem other than the fungal cultivar. Fungus gardens of leaf-cutter ants have so far been found to contain numerous microbial groups in addition to the dominant fungal mutualist (Table 1.1). The most well-described symbionts include a specialized parasite of the fungal cultivar, *Escovopsis*, as well as an antibiotic-producing Actinobacterium (genus *Pseudonocardia*) found to defend against it (24-26). Research on these two symbionts has previously been reviewed (26, 27), and will be discussed here only briefly. Numerous other microfungi and yeasts have also been found associated with the fungus gardens of many ant species (28-36). While filamentous fungi likely represent garden “weeds”, it is unclear if the yeasts have a deleterious effect on the fungus garden ecosystem. One study even found evidence that yeasts may antagonize microfungal pests

(33). Of the bacteria cultured from fungus gardens, many have been proposed to have important roles in the fungus garden ecosystem ranging from antibiotic-mediated exclusion of pathogens to nutrient biosynthesis (37-39). However, the consistent presence of many of these microbes has yet to be demonstrated, and it remains a possibility that they represent allochthonous groups introduced from the incoming foliar material or surrounding soil.

Recently, culture-independent techniques have begun to shed light on bacterial groups thought to be common constituents of fungus gardens (Table 1.1). Membrane lipid profiles have shown that different leaf-cutter ant fungus gardens are highly similar, and that Gram-negative bacteria likely dominate the prokaryotic component of these ecosystems (40). Subsequent 16S libraries have confirmed this, and further indicated that γ-proteobacteria are particularly diverse in these environments, although sequences matching to α-, β- and δ-Proteobacteria, Bacteriodetes, Firmicutes, Actinobacteria, Acidobacteria, and several other phyla were also recovered (41). The most recent metagenomic investigation of leaf-cutter ant fungus gardens has indicated that γ-proteobacteria in the genera *Enterobacter*, *Pantoea*, *Klebsiella*, *Citrobacter*, and *Escherichia* may constitute a core prokaryotic community (42). In addition to culture-independent work, one study successfully cultured *Klebsiella* and *Pantoea* isolates from a variety of leaf-cutter ant nests and demonstrated their capacity to fix nitrogen (43). Moreover, this study traced nitrogen in fungus gardens to the biomass of the ants themselves, indicating that bacteria could be playing an important nutritional role in these ecosystems. Nitrogen fixation was proposed to be a critical process in fungus gardens given the relatively high nitrogen content of ant biomass compared to the incoming plant forage and the surrounding ecosystem.

**1.4 The Ancillary Gut Hypothesis**

Some authors have postulated that the fungus garden ecosystem serves as an external gut for the entire ant colony (41-43). Fungus-growing ants have previously been described as a "superorganism" in that the worker castes are not reproductively viable, and survival of the colony hinges completely on the queen (13). Given that colonies of fungus growing ants are superorganisms rather than conglomerations of individuals, it follows that fungus gardens can be viewed as "organs" responsible for nutrient conversion and assimilation, similar to the digestive gut of other herbivores.

The external gut hypothesis goes deeper than mere analogy. For example, both true guts and fungus gardens are specialized structures that harbor populations of bacteria that assist in the conversion of dietary material into nutrients for the host (41, 44, 45). Prokaryotic populations in both true guts and fungus gardens have been implicated in the biosynthesis of nutrients for their hosts (42, 43, 45, 46). Perhaps the largest difference between these ecosystems is that the structural integrity of fungus gardens is provided by the fungal symbiont, whereas the prokaryotic component of true guts is harbored directly within the host. Moreover, the fungus garden ecosystem is dominated by the fungal cultivar, whereas true guts have a relatively smaller and more diverse fungal component (47-49).

Non-taxonomic similarities in prokaryotic diversity are also evident when comparing fungus gardens and mammalian guts. Low-phylum level and high strain-level diversity have been observed in both of these ecosystems when 16S ribosomal genes have been sequenced (4, 6, 41, 42, 50). This pattern is independent of microbial taxonomy, as many mammalian guts are composed of primarily Bacteriodetes and Firmicutes (6, 9, 50-53), whereas γ-proteobacteria

appear to dominate the prokaryotic component of fungus gardens. The causes of this distinct diversity profile are unclear, although in mammals it has been hypothesized that it may be the result of an adaptive radiation of a few initial prokaryotic "colonists" (4). The relatively recent origin of these symbiotic niches compared to free living environments on the planet has also been implicated in limiting the number of microbial phyla that have evolved to live in them (4).

### 1.5 Evolution of Hygiene in the Attines

The finding that fungus gardens are composed predominantly of one fungal symbiont and 5–6 bacterial groups has led to the question of how this community composition is maintained. The partially-degraded plant material present in fungus gardens could potentially be used as a substrate for countless microorganisms, yet somehow only a small number predominate. Furthermore, the composition of the fungus garden community has been shown to be relatively consistent between fungus garden strata irrespective of the extent of biomass degradation (41, 42), suggesting the existence of selective pressures to maintain a consistent community.

A number of factors likely contribute to the low diversity in fungus gardens (Table 1.2). The meticulous cleaning of fungus gardens by the ants is likely paramount, both for maintaining healthy cultures of the fungal symbiont and a consistent prokaryotic assemblage. Three main hygienic behaviors have been documented in fungus-growing ants. The first, termed weeding, is characterized by the specific removal of whole fragments of fungus garden material (54). The ants weed their gardens to remove dead fungal debris as well as areas infected with pests, especially the specialized parasite *Escovopsis*. Some species of leaf-cutter ants have been shown to maintain specialized waste dumps for their agricultural waste, and the separation between these refuse heaps and fungus gardens is likely key to the maintenance of overall nest hygiene.

The second behavior, termed ‘fungus grooming’, is characterized by the licking of fungus garden material by the ants and selective filtering of foreign spores into infrabuccal pockets located in their oral cavities (54). This behavior appears to be critical for removing spores of foreign fungi that could lead to future infection. Both weeding and grooming have been stimulated in experimental ant nests by the addition of foreign fungal spores, suggesting that the ants have acute mechanisms for assessing the composition and health of their gardens (54).

The third behavior involves the application of fecal droplets to the fungus garden matrix. Some species of attines have been shown to concentrate fungal chitinases and lignocellulases in these droplets, and their integration into fungus gardens has been hypothesized to contribute both to plant biomass degradation and the removal of fungal pests (32, 55, 56). These drastic changes in the behavior of attines highlight how the long-term maintenance of symbiosis can have profound impacts on the life history of host organisms. Glandular secretions of the ants themselves have also been shown to play a role in maintaining the hygiene of fungus gardens. Although known to ward off infection in all ants (57-59), metapleural glands may also be used by fungus-growing ants to remove unwanted microbial groups from their fungus gardens. For example, studies have found that ants consistently rub their legs against their metapleural glands while weeding and grooming so as to apply glandular secretions to the fungus garden (60, 61). Moreover, the application of metapleural gland secretions increases with weeding and grooming behavior when foreign fungal spores are experimentally introduced into a nest (60, 61).

**Table 1.1** Recent research on microbial diversity and plant biomass degradation in attine fungus gardens.

| Reference | Ant Genera | Microbes Analyzed | Plant Polymers Analyzed | Methods | Principle Findings |
|---|---|---|---|---|---|
| (62) | *Atta* | *L. gongylophorus* | Cellulose | Growth assays, enzymatic assays. | Evidence that *L. gongylophorus* can grow on cellulose in pure culture and hydrolyze this polymer efficiently. |
| (63) | *Atta* | Bacteria | Gelatin, cellulose, cellobiose, casein | Directed culturing. | Isolation of plant polymer-degrading bacteria from fungus gardens. |
| (25, 28-30, 33-36, 64, 65) | *Acromyrmex* *Atta* *Myrmicocrypta* *Trachymyrmex* *Cyphomyrmex* | Microfungi, yeasts | NA | Culturing, bioassays. | Isolation and characterization of microfungi and yeasts from fungus gardens*;* evidence that yeasts may antagonize potential garden pathogens; Identification of a novel *Trichosporon* species in fungus gardens. |
| (66, 67) | *Acromyrmex*, *Atta* | *L. gongylophorus* | Cellulose | Pure culture growth assays, estimation of lignin and cellulose content. | Indication that *L. gongylophorus* does not degrade cellulose in pure culture. |
| (68, 69) | *Acromyrmex*, *Atta* | Whole fungus garden, ants, larvae | Polysaccharides, heterosides, oligosaccharides | Enzymatic activity assays on workers, larvae, and fungus gardens. | Enzymatic activity profiles for the fungus garden and host ants were largely non-overlapping; xylanase, amylase, laminarinase, cellulase, and lichenanase activities identified in fungus garden samples; evidence for high variability in the enzymatic activities of fungus gardens between different nests and ant species. |
| (56) | *Acromyrmex*, *Atta* | *L. gongylophorus* | Pectin, CMC, protein | Isoelectric focusing, enzymatic assays. | Identification of fungal pectinases, CMCases, proteinases, and laccases present in leaf-cutter ant fecal droplets. |

**Table 1.1** *Cont.*

| Reference | Ant Genera | Microbes Analyzed | Plant Polymers Analyzed | Methods | Principle Findings |
|---|---|---|---|---|---|
| (70-72) | *Atta* | *L. gongylophorus* | Starch, pectin, xylan, cellulose, CMC | Growth and enzymatic assays. | Rapid growth of *L. gongylophorus* on xylan and starch, but poor growth on cellulose; Production of pectinases, xylanases, cellulases, and amylases by *L. gongylophorus* when grown in pure culture on different carbon sources; Identification of potential mechanisms for regulation of *L. gongylophorus* starch metabolism by the ant hosts. |
| (73) | *Atta* | Whole fungus garden | All biomass | Estimation of cellulose and lignin content. | Evidence that the lignin:cellulose ratio is higher in fungus garden waste than leaf material. |
| (74) | *Acromyrmex* | *L. gongylophorus* | Xylan | Activity measurements of a xylanase, AZCL-based colorimetric assays. | Identification and characterization of an *L. gongylophorus* xylanase. |
| (75) | *Acromyrmex* | Whole fungus garden, *L. gongylophorus* pure cultures | Numerous polysaccharides | Enzymatic activity assays on *L. gongylophorus* and whole fungus gardens, SEM. | Demonstration of broad lignocellulolytic capabilities of *L. gongylophorus* and whole fungus gardens, including activity against, laminarin, chitin, pectin, and CMC. |
| (76) | Cross phylogeny | Whole fungus gardens | Numerous polysaccharides | AZCL-based colorimetric assays. | Evidence for an evolutionary transition towards more efficient proteinase and amylase activity in leaf-cutter ant fungus gardens; evidence for broad lignocellulolytic capacity in lower attine fungus gardens. |

**Table 1.1** *Cont.*

| Reference | Ant Genera | Microbes Analyzed | Plant Polymers Analyzed | Methods | Principle Findings |
|---|---|---|---|---|---|
| (22) | *Atta* | *L. gongylophorus* | NA | Microsattelite profiling of *L. gongylophorus* from different ant nests and time points. | Confirms that a single strain of *L. gongylophorus* is cultured in leaf-cutter ant fungus gardens. |
| (43) | *Atta*, *Acromyrmex* | *Pantoea*, *Klebsiella* | NA | Directed culturing, stable isotope analysis, acetylene reduction analysis, phylogenetic comparisons. | Identification of nitrogen-fixing *Klebsiella* and *Pantoea* isolates in fungus gardens; Evidence that nitrogen fixed in fungus gardens is integrated into ant biomass. |
| (77) | *Acromyrmex* | *L. gongylophorus* | Pectin | Proteomics, RT-qPCR, enzymatic assays. | Identification of diverse fungal pectinases concentrated in the fecal droplets of the ants; evidence that *L. gongylophorus* produces enzymes specifically in the hyphal swellings fed to the ants. |
| (40) | *Atta*, *Acromyrmex* | Primarily Gram-negative bacteria | NA | Lipid profiling using PLFA and FAME. | Evidence that ant fungus gardens and refuse dumps contain distinct microbial communities; evidence that the prokaryotic fraction of fungus gardens is dominated by Gram-negative bacteria. |
| (41) | *Atta* | γ-proteobacteria, primarily *Klebsiella* and *Pantoea* | Cellulose, hemicelluloses | Community metagenomics, 16S surveys, genome sequencing, enzymatic assays, sugar composition analysis. | Survey of bacterial diversity in fungus gardens; Identification of abundant *Klebsiella* and *Pantoea* populations; characterization of bacterial glycoside hydrolases; Evidence for significant amounts of cellulose degradation in fungus gardens. |

**Table 1.1** *Cont.*

| Reference | Ant Genera | Microbes Analyzed | Plant Polymers Analyzed | Methods | Principle Findings |
|---|---|---|---|---|---|
| (78) | *Atta* | Whole fungus garden | Numerous polysaccharides | AZCL-based colorimetric assays | Evidence that enzyme profiles in fungus gardens shift rapidly when integrated foliar biomass changes. |
| (79) | *Acromyrmex* | Whole fungus garden | Pectin, xylose | Antibody and CBM-based polysaccharide microarray profiling, AZCL-based colorimetric assays. | Evidence for the degradation of xylan and pectin, but not cellulose, in fungus gardens; Indication that plant material is only partially degraded in these ecosystems. |
| (80) | *Atta* | *L. gongylophorus* | All biomass | Dye and photomicrography of plant biomass. | Evidence for substantial degradation of all non-lignified plant tissues in fungus gardens; indication that *L. gongylophorus* may degrade a large quantity of cellulose in fungus gardens |
| (81) | Cross phylogeny | *L. gongylophorus* | Protein | Enzymatic assays, pH and buffering analysis. | Indication that the fungal cultivars of higher attines have evolved proteinases with activity optima at pH ~5, closer to the pH of fungus gardens; characterization of different proteinase classes and buffing capacities in different fungus gardens |
| (42) | *Atta* | Bacteria | Cellulose, hemicelluloses | Community metagenomics, 16S surveys, metaproteomics. | Identification of abundant *Enterobacter* population in fungus gardens; Further characterization of *Klebsiella* and *Pantoea* populations; Proteomic identification of bacterial glycoside hydrolases; Identification of bacteriophage in fungus gardens |

**Table 1.2** Factors contributing to microbial assemblage composition in attine ant fungus gardens.

| | |
|---|---|
| **Factors limiting diversity** | Weeding and grooming of fungus gardens, application of glandular secretions, application of antimicrobials from *Pseudonocardia*, antibiotics produced by the fungal cultivar, fecal droplets of the ants |
| **Factors promoting diversity** | Complex, nutrient rich substrate |
| **Potential sources of microbial groups** | Maternal transmission from parent colony, phyllosphere microbes on foliar biomass, surrounding soil, the ants themselves |

Chemical analysis of leaf-cutter ant metapleural gland secretions identified phenylacetic acid and number of short-chain fatty acids known to have antimicrobial properties (82, 83). Bioassays have confirmed that these glandular secretions have broad antifungal and antibacterial activity (84, 85). Secretions of the mandibular gland have also been implicated in potentially inhibiting the germination of alien fungi (86). Because glandular secretions could potentially inhibit the growth of the fungal cultivar if consistently introduced to fungus gardens, the ants may rely only on selective application of these secretions to areas thought to be infected.

The fungus itself may also produce compounds that selectively inhibit or promote the growth of other microbes in its environment. Basidiomycetes in general are a rich source of secondary metabolites (87), and novel antimicrobial compounds have previously been identified from *Leucoagaricus* species (88). Moreover, the fungus cultured by the ants has been implicated in the production of organohalogens (89), which may be involved in lignocellulose degradation or antibiosis in basidiomycetes (90). Antibiosis of cultivated *Leucoagaricus* isolates against *Escovopsis* species has also been shown, suggesting these fungi have at least some capacity for the production of secondary metabolites (91).

Lastly, fungus-growing ants have been shown to constrain microbial diversity in their fungus gardens through association with antibiotic-producing Actinobacteria (24, 26, 39, 92-96). Bacteria of the genus *Pseudonocardia* have been shown to produce compounds that inhibit the specialized garden parasite *Escovopsis* (24, 97), and experimental evidence suggests these microbes play a role in maintaining garden hygiene (93, 98). Other Actinobacteria isolated from ant colonies have also been proposed to play a role in the defense against garden pathogens, but the consistent presence of these microbes in attine nests has yet to be determined (39, 99, 100). Regardless, the combination of compounds produced by Actinobacteria and the fungal cultivar, together with the glandular secretions of the ants, likely produces a potent antimicrobial cocktail that could be critical for shaping microbial diversity in fungus gardens.

### 1.6 Plant Biomass Degradation in Fungus Gardens

The conversion of plant biomass into nutrients usable by the ants is the central role of fungus gardens. Despite the central importance of this process, the mechanisms through which plant biomass is degraded in these ecosystems are only beginning to be elucidated (Table 1.1). In general, the plant biomass integrated into fungus gardens is a rich source of cellulose, hemicelluloses, protein, lignin, simple sugars, and various other compounds. In the fungus gardens of higher attines, this is converted into hyphal swellings called gongylidia, which are rich in lipids, carbohydrates, and other nutrients produced by the fungal cultivar (101, 102). Importantly, gongylidia serve as a primary food source for the entire ant colony, and are the exclusive nutrient source for the developing larvae and brood (13, 20). Identification of the mechanisms through which biomass is degraded, how it is converted into energy for the ants, and

which microbes are taking part in these processes is critical to a fundamental understanding of the ant-fungus garden symbiosis.

Most work has focused on the lignocellulolytic capacity of *L. gongylophorus*, the fungus cultured by leaf-cutter ants. Numerous growth- and enzyme-based assays have indicated that pure cultures of this organism can degrade and grow rapidly on both starch and xylan (62, 70-72). Furthermore, while pectin is degraded rapidly by this organism, this polymer supports only intermediate growth (70, 72). Numerous pectinases and one xylanase have been identified in *L. gongylophorus* (74, 77), providing evidence that it possesses coding potential typical of other saprotrophic basidiomycetes.

Analysis of whole fungus garden extract has indicated that a wide variety of plant polymers are degraded in fungus gardens, including xylan, pectin, starch, laminarin, cellulose, lichenan, and chitin (68, 69, 75, 76, 78). One study sought to compare enzymatic profiles of fungus gardens across the phylogeny of the attines (76). This analysis found that the fungus gardens of leaf-cutter ants had higher amylase activity than fungus gardens of the lower attines (76). Moreover, the overall proteinase activity was significantly higher in all higher attine fungus gardens compared to the fungus gardens of lower attines (76). Another study found that the pH optima of proteinases in leaf-cutter ant fungus gardens was lower than those in lower attine fungus gardens, and that the buffering capacity in leaf-cutter ant fungus gardens was also higher (81). Together these results suggest that multiple evolutionary transitions throughout the history of the ant-fungus garden association have led to a specialized form of biomass degradation in leaf-cutter ant fungus gardens (76, 81).

The most controversial aspect of plant biomass degradation in fungus gardens is the deconstruction of cellulose. Early work found support for the hypothesis that cellulose was the primary polymer supporting fungal growth in these environments, and it was estimated that up to 45% of the cellulose in foliar biomass was degraded in fungus gardens (103). More recently, however, studies have challenged this hypothesis, reporting that *L. gongylophorus* cannot grow in pure culture with cellulose as the sole carbon source (62, 66, 67), suggesting that the secreted proteins of this fungus have only limited cellulolytic activity (70, 72). Measurements of cellulose degradation in fungus gardens have also varied, with sugar composition analyses indicating that ~30% of the cellulose in foliar biomass is degraded in fungus gardens (41), while a polysaccharide microarray approach documented only limited degradation of this polymer (79). One microscopy-based study examined plant biomass in fungus gardens throughout different stages of decomposition and found strong support for the degradation of parenchyma, endodermis, and vascular bundle cells, indicating that all plant polymers in the plant cell wall, with the exception of lignin, were degraded extensively (80).

High variability in the lignocellulolytic activity in fungus gardens may explain some of these conflicting results. Measurements of the enzymatic activity of whole fungus garden protein extracts typically vary over a wide range, even when fungus gardens of the same species of ant are compared (69, 76). One study even documented a rapid shift in enzymatic activity in this environment when laboratory ant colonies were switched from a diet of foliar biomass to one of starch-rich rice (78). It is thus possible that large quantities of cellulose are degraded in fungus gardens, but that this amount varies depending on the foliar biomass provided by the ants.

Indeed, because such a variety of plants are harvested by leaf-cutter ants (104-106), it may be crucial that *L. gongylophorus* is capable of both degrading a variety of plant polymers and changing its lignocellulolytic activity to match the plant substrate available. The selection of plant polymers degraded by *L. gongylophorus* may also depend on the nutritional status of the host colony, as more recalcitrant polymers present a vast food supply but require the input of additional resources for effective degradation.

A potential explanation to the limited lignocellulolytic capacity of *L. gongylophorus* pure cultures is that the host ants enhance the biodegradative capacity of this organism *in situ*. Attines are known to concentrate lignocellulolytic enzymes in fecal droplets that they deposit on freshly integrated plant biomass in fungus gardens, presumably to assist in the first stages of biomass degradation, or, as mentioned above, to inhibit potential garden pests (55, 107-109). Recent work has confirmed these enzymes originate from the fungal gongylidia consumed by the ants, and that a variety of pectinases, carboxymethylcellulases, amylases, and even laccases are present in this cocktail (56, 77). The relative importance of this pre-treatment step to overall biomass degradation in fungus gardens is not known, nor is the full extent of the enzymes that may be present in fecal droplets.

It is also possible that microbes other than the fungal cultivar are essential for plant biomass degradation, either by deconstructing plant polymers directly or by enhancing the lignocellulolytic activity of the fungal cultivar. There is precedent for bacterial-fungal interactions promoting fungal lignocellulose degradation in other systems, but the mechanisms for these phenomena remain unclear (110). Lignocellulolytic bacteria have also been cultured

directly from fungus gardens (63), indicating that these microbes could participate directly in biomass processing. Recent metagenomic analyses have confirmed that bacteria in fungus gardens possess a variety of glycoside hydrolase genes that could potentially deconstruct plant polysaccharides directly (41, 42), and proteomic work has confirmed that at least some of these genes are expressed in fungus gardens (42). A number of studies have investigated microfungi and yeasts in fungus gardens (Table 1.1), but the lignocellulolytic capacities of these organisms have yet to be investigated.

Regardless if specific microbe-microbe or ant-microbe interactions are significantly influencing biomass degradation, it is clear that the physiology of fungus garden microbes in pure culture do not fully reflect their *in situ* physiology. The diverse lignocellulolytic activities repeatedly measured from both *L. gongylophorus* and directly from fungus gardens indicate that a wide variety of plant polymers, including cellulose, are likely degraded in this ecosystem (41, 62, 69, 74, 76, 80). However, the amount of degradation may vary greatly depending on the foliar biomass harvested by the ants and other community-level processes that have yet to be elucidated (69, 76, 78). Confirming the extent to which the fungal cultivar, resident bacteria, and other microbes in fungus gardens contribute to lignocellulose degradation, and more thoroughly characterizing the *in situ* physiology of this ecosystem, remains an important future direction in this field.

### 1.7 Reciprocal Adaptation of the Ant Genome

The application of sequenced-based approaches to investigate the fungus-growing ant system has advanced to include the ants themselves. The recently sequenced genomes of the

leaf-cutter ant *Atta cephalotes* and *Acromyrmex echinatior* have provided a wealth of information for studying the symbiosis between these ants and their fungus gardens (16, 17). The obligate dependence of these ants on their cultivated fungi has been thought to lead to reductions at the genomic level, as has previously been observed in other model nutritional symbiosis such as between the pea aphid and its nutrient-producing bacterial endosymbionts (111). In the pea aphid system, the host genome was found to lack genes for the biosynthesis of specific amino acids known to be produced by its endosymbiont *Buchnera* (111, 112). Similarly, both genomes of the leaf-cutter ants were found to lack genes required for arginine biosynthesis, in contrast to other ant genomes that contain the entire pathway (16, 17). One hypothesis is that the fungus may provide arginine to the ants, thereby limiting the need for this particular pathway. Previous work documenting the compounds in an *Atta colombica* cultivar has shown the presence of free arginine, providing some support for this hypothesis (101). The *A. cephalotes* genome was also found to be missing a hexamerin gene thought to be involved in amino acid sequestration during larval development (16), potentially indicating that these ants have a reduced need to store amino acids since larvae may get these nutrients from the fungal cultivar.

Serine proteases, which are potentially important in the degradation of proteins in the diet, were also found to be reduced in the *A. cephalotes* genome compared to other insects (16). As with the loss of arginine biosynthesis and hexamerins, this may reflect a decreased capacity of leaf-cutter ants to acquire nutrients from their environment coincident with their dependence on *L. gongylophorus* for food. Another possible hypothesis is that this reduction in proteases is related to the ants' ability to concentrate fungal enzymes in their fecal droplets. Because

extensive degradation of proteins in the diet would preclude the concentration and application of fungal enzymes to the fungus gardens, it is possible that the lack of these genes is a result of millions of years of this peculiar behavior.

Overall these data are consistent with the hypothesis that fungus-growing ants have lost some capacity to acquire nutrients over the course of their 50 million year co-evolutionary history with their obligate fungal mutualist. It is interesting to note that co-evolved symbioses, characterized by the evolutionary innovation provided by both partners, appear to be accompanied by physiological restrictions in the hosts. Perhaps most striking is the possibility that a behavioral innovation of the ants—the concentration of lignocellulolytic and potentially antimicrobial fungal enzymes in fecal droplets—may have also resulted in profound genomic changes. These results, together with what is known for other symbioses such as the pea aphid-*Buchnera* system, suggest a common theme in nutritional symbioses: prolonged reliance by the host on a single symbiont for nutrition resulting in extensive and elaborate evolutionary transitions towards obligate association.

### 1.8 Conclusions and Future Outlook

Attine ants have long been a model system for the study of symbiosis, co-evolution, and evolutionary innovation. The intriguing symbiosis between these ants and their fungus-bacteria consortia is remarkable both because of its stability (~50 million years) and the drastic effects it has had on both partners (production of hyphal swellings and shifts in enzymatic profiles by the fungal cultivar, and genomic and behavioral changes in the host ants). Here we have reviewed how advances in genome sequencing and culture-independent investigations of microbial

communities, together with more traditional approaches, have transformed our understanding of co-evolution and evolutionary innovation in the fungus-growing ants. Future work answering key questions of plant biomass degradation and nutrient conversion in fungus gardens, reciprocal ant-fungus co-evolution, and fungus garden microbial diversity will allow for a more fundamental understanding of this remarkable system. For example, advances in proteomics and *in situ* analyses of microbial communities will shed light on the degradation of cellulose in fungus gardens, and the role different microbial groups play in biomass processing. Moreover, by leveraging the large amount of sequence information now available for the organisms in this symbiosis, it will soon be possible to understand the degree of metabolic integration between the ants and their fungus, and the extent to which the behavior of the host has been altered by genetic and physiological constraints of the symbiosis. Given the prevalence of symbiotic microbial communities in nature, gaining a fundamental understanding of the ant-fungus garden system will likely have far-reaching implications for understanding broader trends in the ecology and evolution of metazoans.

**1.9 References**

1. **Moran, NA.** 2006. Symbiosis. *Curr. Biol.* **16:**R866.

2. **Zilber-Rosenberg, I, and Rosenberg, E.** 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32:**723-735.

3. **Douglas, AE.** 2010. The symbiotic habit. Princeton University Press, Princeton, New Jersey, USA.

4. **Ley, RE, Peterson, DA, and Gordon, JI.** 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124:**837-848.

5. **Woyke, T, Teeling, H, Ivanova, NN, Huntemann, M, Richter, M, Gloeckner, FO, Boffelli, D, Anderson, IJ, Barry, KW, Shapiro, HJ, Szeto, E, Kyrpides, NC, Mussmann, M, Amann, R, Bergin, C, Ruehland, C, Rubin, EM, and Dubilier, N.** 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443:**950-955.

6. **Ley, RE, Hamady, M, Lozupone, C, Turnbaugh, PJ, Ramey, RR, Bircher, JS, Schlegel, ML, Tucker, TA, Schrenzel, MD, Knight, R, and Gordon, JI.** 2008. Evolution of mammals and their gut microbes. *Science* **320:**1647-51.

7. **Distel, DL, and Roberts, SJ.** 1997. Bacterial endosymbionts in the gills of the deep-sea wood-boring bivalves *Xylophaga atlantica* and *Xylophaga washingtona*. *Biol. Bull.* **192:**253-261.

8. **Yang, JC, Madupu, R, Durkin, AS, Ekborg, NA, Pedamallu, CS, Hostetler, JB, Radune, D, Toms, BS, Henrissat, B, Coutinho, PM, Schwarz, S, Field, L, Trindade-Silva, AE, Soares, CA, Elshahawi, S, Hanora, A, Schmidt, EW, Haygood, MG, Posfai, J, Benner, J, Madinger, C, Nove, J, Anton, B, Chaudhary, K, Foster, J, Holman, A, Kumar, S, Lessard, PA, Luyten, YA, Slatko, B, Wood, N, Wu, B, Teplitski, M, Mougous, JD, Ward, N, Eisen, JA, Badger, JH, and Distel, DL.** 2009. The complete genome of *Teredinibacter turnerae* T7901: an intracellular endosymbiont of marine wood-boring bivalves (shipworms). *PLoS One* **4:**e6085.

9. **Pope, PB, Denman, SE, Jones, M, Tringe, SG, Barry, K, Malfatti, SA, McHardy, AC, Cheng, JF, Hugenholtz, P, McSweeney, CS, and Morrison, M.** 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc. Natl. Acad. Sci. USA* **107:**14793-8.

10. **Schultz, TR, and Brady, SG.** 2008. Major evolutionary transitions in ant agriculture. *Proc. Natl. Acad. Sci. USA* **105:**5435-5440.

11. **Mayhe-Nunes, AJ, and Jaffe, K.** 1998. On the biogeography of Attini (Hymenoptera: Formicidae). *Ecotropicos* **11:**45-54.

12. **Hölldobler, B, and Wilson, EO.** 1990. The ants. Harvard University Press, Cambridge, MA.

13. **Hölldobler, B, and Wilson, EO.** 2008. The superorganism: the beauty, elegance, and strangeness of insect societies. W. W. Norton & Company, New York, NY.

14. **Costa, AN, Vasconcelos, HL, Vieira-Neto, EHM, and Bruna, EM.** 2009. Do herbivores exert top-down effects in Neotropical savannas? Estimates of biomass consumption by leaf-cutter ants. *J. Veg. Sci.* **19:**849-854.

15. **Hölldobler, B, and Wilson, EO.** 2010. The leafcutter ants: civilization by instinct. W. W. Norton and Company, Inc., New York, NY.

16. **Suen, G, Teiling, C, Li, L, Holt, C, Abouheif, E, Bornberg-Bauer, E, Bouffard, P, Caldera, EJ, Cash, E, Cavanaugh, A, Denas, O, Elhaik, E, Fave, MJ, Gadau, J, Gibson, JD, Graur, D, Grubbs, KJ, Hagen, DE, Harkins, TT, Helmkampf, M, Hu, H, Johnson, BR, Kim, J, Marsh, SE, Moeller, JA, Munoz-Torres, MC, Murphy, MC, Naughton, MC, Nigam, S, Overson, R, Rajakumar, R, Reese, JT, Scott, JJ, Smith, CR, Tao, S, Tsutsui, ND, Viljakainen, L, Wissler, L, Yandell, MD, Zimmer, F, Taylor, J, Slater, SC, Clifton, SW, Warren, WC, Elsik, CG, Smith, CD, Weinstock, GM, Gerardo, NM, and Currie, CR.** 2011. The genome sequence of the leaf-cutter ant *Atta cephalotes* reveals insights into its obligate symbiotic lifestyle. *PLoS Genet.* **7:**e1002007.

17. **Nygaard, S, Zhang, G, Schiott, M, Li, C, Wurm, Y, Hu, H, Zhou, J, Ji, L, Qiu, F, Rasmussen, M, Pan, H, Hauser, F, Krogh, A, Grimmelikhuijzen, CJ, Wang, J, and Boomsma, JJ.** 2011. The genome of the leaf-cutting ant *Acromyrmex echinatior* suggests key adaptations to advanced social life and fungus farming. *Genome Res.* **21:**1339-1348.

18. **Chapela, IH, Rehner, SA, Schultz, TR, and Mueller, UG.** 1994. Evolutionary history of the symbiosis between fungus-growing ants and their fungi. *Science* **266:**1691-1694.

19. **Villesen, P, Mueller, UG, Schultz, TR, Adams, RM, and Bouck, AC.** 2004. Evolution of ant-cultivar specialization and cultivar switching in *Apterostigma* fungus-growing ants. *Evolution* **58:**2252-2265.

20. **Weber, NA.** 1966. Fungus-growing ants. *Science* **153:**587-604.

21. **Weber, NA.** 1972. The fungus-culturing behavior of ants. *Am. Zool.* **12:**577-587.

22. **Mueller, UG, Scott, JJ, Ishak, HD, Cooper, M, and Rodrigues, A.** 2010. Monoculture of leafcutter ant gardens. *PLoS One* **5:**e12668.

23. **Mikheyev, AS, Mueller, UG, and Abbot, P.** 2010. Comparative dating of attine ant and lepiotaceous cultivar phylogenies reveals coevolutionary synchrony and discord. *Am. Nat.* **175:**E126-133.

24. **Currie, CR, Scott, JA, Summerbell, RC, and Malloch, D.** 1999. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398:**701-704.

25. **Currie, CR, Mueller, UG, and Malloch, D.** 1999. The agricultural pathology of ant fungus gardens. *Proc. Natl. Acad. Sci. USA* **96:**7998-8002.

26. **Currie, CR.** 2001. A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. *Annu. Rev. Microbiol.* **55:**357-380.

27. **Caldera, EJ, Poulsen, M, Suen, G, and Currie, CR.** 2009. Insect symbioses: a case study of past, present, and future fungus-growing ant research. *Environ. Entomol.* **38:**78-92.

28. **Fisher, PJ, Stradling, DJ, Sutton, BC, and Petrini, LE.** 1996. Microfungi in the fungus gardens of the leaf-cutting ant *Atta cephalotes*: a preliminary study. *Mycol. Res.* **100:**541-546.

29. **Carreiro, SC, Pagnocca, FC, Bacci, M, Jr., Bueno, OC, Hebling, MJ, and Middelhoven, WJ.** 2002. Occurrence of killer yeasts in leaf-cutting ant nests. *Folia Microbiol. (Praha)* **47:**259-262.

30. **Carreiro, SC, Pagnocca, FC, Bacci, M, Jr., Lachance, MA, Bueno, OC, Hebling, MJ, Ruivo, CC, and Rosa, CA.** 2004. *Sympodiomyces attinorum* sp. nov., a yeast species associated with nests of the leaf-cutting ant *Atta sexdens*. *Int. J. Syst. Evol. Microbiol.* **54:**1891-1894.

31. **Carreiro, SC, Pagnocca, FC, Bueno, OC, Bacci, MJ, Hebling, MJ, and da Silva, OA.** 1997. Yeasts associated with nests of the leaf-cutting ant *Atta sexdens rubropilosa* Forel, 1908. *A. Van Leeuw. J. Microb.* **71:**243-248.

32. **Rodrigues, A, Bacci Jr, M, Mueller, UG, Ortiz, A, and Pagnocca, FC.** 2008. Microfungal "Weeds" in the Leafcutter Ant Symbiosis. *Microb. Ecol.* **56:**604-614.

33. **Rodrigues, A, Cable, RN, Mueller, UG, Bacci, M, Jr., and Pagnocca, FC.** 2009. Antagonistic interactions between garden yeasts and microfungal garden pathogens of leaf-cutting ants. *A. Van Leeuw. J. Microb.* **96:**331-342.

34. **Rodrigues, A, Mueller, UG, Ishak, HD, Bacci Jr, M, and Pagnocca, FC.** 2011. Ecology of microfungal communities in gardens of fungus-growing ants (Hymenoptera: Formicidae): a year-long survey of three species of attine ants in central Texas. *FEMS Microbiol. Ecol.* **78:**244-255.

35. **Rodrigues, A, Pagnocca, FC, Bacci, MJ, Hebling, MJ, Bueno, OC, and Pfenning, LH.** 2005. Variability of non-mutualistic filamentous fungi associated with *Atta sexdens rubropilosa* nests. *Folia Microbiol. (Praha)* **50:**421-425.

36. **Pagnocca, FC, Legaspe, MF, Rodrigues, A, Ruivo, CC, Nagamoto, NS, Bacci, M, Jr., and Forti, LC.** 2010. Yeasts isolated from a fungus-growing ant nest, including the description of *Trichosporon chiarellii* sp. nov., an anamorphic basidiomycetous yeast. *Int. J. Syst. Evol. Microbiol.* **60:**1454-1459.

37. **Santos, AV, Dillon, RJ, Dillon, VM, Reynolds, SE, and Samuels, RI.** 2004. Ocurrence of the antibiotic producing bacterium *Burkholderia* sp. in colonies of the leaf-cutting ant *Atta sexdens rubropilosa*. *FEMS Microbiol. Lett.* **239:**319-323.

38. **Mueller, U, Gerardo, N, Aanen, D, Six, D, and Schultz, T.** 2005. The evolution of agriculture in insects. *Annu. Rev. Ecol. Evol. Syst.* **36:**563-595.

39. **Haeder, S, Wirth, R, Herz, H, and Spiteller, D.** 2009. Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *Proc. Natl. Acad. Sci. USA* **106:**4742-4746.

40. **Scott, JJ, Budsberg, KJ, Suen, G, Wixon, DL, Balser, TC, and Currie, CR.** 2010. Microbial community structure of leaf-cutter ant fungus gardens and refuse dumps. *PLoS One* **5:**e9922.

41. **Suen, G, Scott, JJ, Aylward, FO, Adams, SM, Tringe, SG, Pinto-Tomas, AA, Foster, CE, Pauly, M, Weimer, PJ, Barry, KW, Goodwin, LA, Bouffard, P, Li, L, Osterberger, J, Harkins, TT, Slater, SC, Donohue, TJ, and Currie, CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6:**e1001129.

42. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **6:**1688–1701.

43. **Pinto-Tomas, AA, Anderson, MA, Suen, G, Stevenson, DM, Chu, FS, Cleland, WW, Weimer, PJ, and Currie, CR.** 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* **326:**1120-1123.

44. **Backhed, F, Ding, H, Wang, T, Hooper, LV, Koh, GY, Nagy, A, Semenkovich, CF, and Gordon, JI.** 2004. The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl. Acad. Sci. USA* **101:**15718-23.

45. **Backhed, F, Ley, RE, Sonnenburg, JL, Peterson, DA, and Gordon, JI.** 2005. Host-bacterial mutualism in the human intestine. *Science* **307:**1915-1920.

46. **Gill, SR, Pop, M, Deboy, RT, Eckburg, PB, Turnbaugh, PJ, Samuel, BS, Gordon, JI, Relman, DA, Fraser-Liggett, CM, and Nelson, KE.** 2006. Metagenomic analysis of the human distal gut microbiome. *Science* **312:**1355-1359.

47. **Teunissen, MJ, and Op den Camp, HJ.** 1993. Anaerobic fungi and their cellulolytic and xylanolytic enzymes. *A. Van Leeuw. J. Microb.* **63:**63-76.

48. **Brookman, JL, Mennim, G, Trinci, AP, Theodorou, MK, and Tuckwell, DS.** 2000. Identification and characterization of anaerobic gut fungi using molecular methodologies based on ribosomal ITS1 and 185 rRNA. *Microbiology* **146 (Pt 2):**393-403.

49. **Liggenstoffer, AS, Youssef, NH, Couger, MB, and Elshahed, MS.** 2010. Phylogenetic diversity and community structure of anaerobic gut fungi (phylum Neocallimastigomycota) in ruminant and non-ruminant herbivores. *ISME J.* **4:**1225-1235.

50. **Ley, RE, Lozupone, CA, Hamady, M, Knight, R, and Gordon, JI.** 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6:**776-88.

51. **Eckburg, PB, Bik, EM, Bernstein, CN, Purdom, E, Dethlefsen, L, Sargent, M, Gill, SR, Nelson, KE, and Relman, DA.** 2005. Diversity of the human intestinal microbial flora. *Science* **308:**1635-1638.

52. **Brulc, JM, Antonopoulos, DA, Miller, ME, Wilson, MK, Yannarell, AC, Dinsdale, EA, Edwards, RE, Frank, ED, Emerson, JB, Wacklin, P, Coutinho, PM, Henrissat, B, Nelson, KE, and White, BA.** 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. USA* **106:**1948-1953.

53. **Hess, M, Sczyrba, A, Egan, R, Kim, TW, Chokhawala, H, Schroth, G, Luo, S, Clark, DS, Chen, F, Zhang, T, Mackie, RI, Pennacchio, LA, Tringe, SG, Visel, A, Woyke, T, Wang, Z, and Rubin, EM.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331:**463-467.

54. **Currie, CR, and Stuart, AE.** 2001. Weeding and grooming of pathogens in agriculture by ants. *P. Roy. Soc. B-Biol. Sci.* **268:**1033-1039.

55. **Martina, MM, Gieselmanna, MJ, and Martin, JS.** 1973. Rectal enzymes of attine ants. α-Amylase and chitinase. *J. Insect Physiol.* **19:**1409-1416.

56. **Ronhede, S, Boomsma, JJ, and Rosendahl, S.** 2004. Fungal enzymes transferred by leaf-cutting ants in their fungus gardens. *Mycol. Res.* **108:**101-106.

57. **Hölldobler, B, and Engel-Siegel, H.** 1984. On the metapleural gland of ants. *Psyche* **91:**201-224.

58. **Beattie, AJ, Turnbull, CL, Hough, T, and Knox, RB.** 1986. Antibiotic production: a possible function for the metapleural glands of ants (Hymenoptera: Formicidae). *Ann. Entomol. Soc. Am.* **79:**448-450.

59. **Veal, DA, Trimble, JE, and Beattie, AJ.** 1992. Antimicrobial properties of secretions from the metapleural glands of *Myrmecia gulosa* (the Australian bull ant). *J. Appl. Bacteriol.* **72:**188-194.

60. **Fernandez-Marin, H, Zimmerman, JK, Rehner, SA, and Wcislo, WT.** 2006. Active use of the metapleural glands by ants in controlling fungal infection. *P. Roy. Soc. B-Biol. Sci.* **273:**1689-1695.

61. **Fernandez-Marin, H, Zimmerman, JK, Nash, DR, Boomsma, JJ, and Wcislo, WT.** 2009. Reduced biological control and enhanced chemical pest management in the evolution of fungus farming in ants. *P. Roy. Soc. B-Biol. Sci.* **276:**2263-2269.

62. **Bacci, M, Anversa, MM, and Pagnocca, FC.** 1995. Cellulose degradation by *Leucocoprinus gongylophorus*, the fungus cultured by the leaf-cutting ant *Atta sexdens rubropilosa*. *A. Van Leeuw. J. Microb.* **67:**385-386.

63. **Bacci M, Ribeiro SB, Casarotto MEF, and Pagnocca FC.** 1995. Biopolymer-degrading bacteria from nests of the leaf-cutting ant *Atta sexdens Rubropilosa*. *Braz. J. Med. Biol. Res.* **28:**79-82.

64. **Currie, CR.** 2001. Prevalence and impact of a virulent parasite on a tripartite mutualism. *Oecologia* **128:**99-106.

65. **Silva, A, Rodrigues, A, Bacci, M, Jr., Pagnocca, FC, and Bueno, OC.** 2006. Susceptibility of the ant-cultivated fungus *Leucoagaricus gongylophorus* (Agaricales: Basidiomycota) towards microfungi. *Mycopathologia* **162:**115-9.

66. **Abril, AB, and Bucher, EH.** 2002. Evidence that the fungus cultured by leaf-cutting ants does not metabolize cellulose. *Ecol. Lett.* **5:**325-328.

67. **Abril, AB, and Bucher, EH.** 2004. Nutritional sources of the fungus cultured by leaf-cutting ants. *App. Soil Ecol.* **26:**243-247.

68. **D'Ettorre, P, Mora, P, Dibangou, V, Rouland, C, and Errard, C.** 2002. The role of the symbiotic fungus in the digestive metabolism of two species of fungus-growing ants. *J. Comp. Physiol. B.* **172:**169-176.

69. **Richard, FJ, Mora, P, Errard, C, and Rouland, C.** 2005. Digestive capacities of leaf-cutting ants and the contribution of their fungal cultivar to the degradation of plant material. *J. Comp. Physiol. B* **175:**297-303.

70. **Silva, A, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJ.** 2006. Production of polysaccharidases in different carbon sources by *Leucoagaricus gongylophorus* Moller

(Singer), the symbiotic fungus of the leaf-cutting ant *Atta sexdens* Linnaeus. *Curr. Microbiol.* **53:**68-71.

71. **Silva, A, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJ.** 2006. Starch metabolism in *Leucoagaricus gongylophorus*, the symbiotic fungus of leaf-cutting ants. *Microbiol. Res.* **161:**299-303.

72. **Gomes De Siqueira, C, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJA.** 1998. Metabolism of plant polysaccharides by *Leucoagaricus gongylophorus*, the symbiotic fungus of the leaf-cutting ant *Atta sexdens*. *Appl. Environ. Microbiol.* **64:**4820-4822.

73. **Sousa-Souto, L, Guerra, MBB, Schoereder, JH, Ernesto, C, Schaefer, GR, and Silva, WLd.** 2007. Determination of the conversion factor in colonies of *Atta sexdens rubropilosa* (*Hymenoptera: Formicidae*) and its relationship with the quality of harvested leaf substrate. *Revista Arvore* **31:**163-166.

74. **Schiott, M, De Fine Licht, HH, Lange, L, and Boomsma, JJ.** 2008. Towards a molecular understanding of symbiont function: identification of a fungal gene for the degradation of xylan in the fungus gardens of leaf-cutting ants. *BMC Microbiol.* **8:**40.

75. **Erthal, M, Jr., Silva, CP, Cooper, RM, and Samuels, RI.** 2009. Hydrolytic enzymes of leaf-cutting ant fungi. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **152:**54-59.

76. **De Fine Licht, HH, Schiott, M, Mueller, UG, and Boomsma, JJ.** 2010. Evolutionary transitions in enzyme activity of ant fungus gardens. *Evolution* **64:**2055-2069.

77. **Schiott, M, Rogowska-Wrzesinska, A, Roepstorff, P, and Boomsma, JJ.** 2010. Leaf-cutting ant fungi produce cell wall degrading pectinase complexes reminiscent of phytopathogenic fungi. *BMC Biol.* **8:**156.

78. **Kooij, PW, Schiott, M, Boomsma, JJ, and De Fine Licht, HH.** 2011. Rapid shifts in *Atta cephalotes* fungus-garden enzyme activity after a change in fungal substrate (Attini, Formicidae). *Insectes Soc.* **58:**145-151.

79. **Moller, IE, De Fine Licht, HH, Harholt, J, Willats, WG, and Boomsma, JJ.** 2011. The dynamics of plant cell-wall polysaccharide decomposition in leaf-cutting ant fungus gardens. *PLoS One* **6:**e17506.

80. **Nagamoto, NS, Garcia, MG, Forti, LC, Verza, SS, Noronha, NC, and Rodella, RA.** 2011. Microscopic evidence supports the hypothesis of high cellulose degradation capacity by the symbiotic fungus of leaf-cutting ants. *J. Biol. Res.-Thessaloniki* **16:**308-12.

81. **Semenova, TA, Hughes, DP, Boomsma, JJ, and Schiott, M.** 2011. Evolutionary patterns of proteinase activity in attine ant fungus gardens. *BMC Microbiol.* **11:**15.

82. **do Nascimento, RR, Schoeters, E, Morgan, ED, Billen, J, and Stradling, DJ.** 1996. Chemistry of metapleural gland secretions of three attine ants, *Atta sexdens rubropilosa*, *Atta cephalotes*, and *Acromyrmex octospinosus* (Hymenoptera: Formicidae). *J. Chem. Ecol.* **22:**987-1000.

83. **Ortius-Lechner, D, Maile, R, Morgan, ED, and Boomsma, JJ.** 2000. Metapleural gland secretion of the leaf-cutter ant *Acromyrmex octospinosus*: new compounds and their functional significance. *J. Chem. Ecol.* **26:**1667-1683.

84. **Bot, A, Ortius-Lechner, D, Finster, K, Maile, R, and Boomsma, J.** 2002. Variable sensitivity of fungi and bacteria to compounds produced by the metapleural glands of leaf-cutting ants. *Insectes Soc.* **49:**363-370.

85. **Mendonca Ade, L, da Silva, CE, de Mesquita, FL, Campos Rda, S, Do Nascimento, RR, Ximenes, EC, and Sant'Ana, AE.** 2009. Antimicrobial activities of components of the glandular secretions of leaf cutting ants of the genus Atta. *A. Van Leeuw. J. Microb.* **95:**295-303.

86. **Rodrigues, A, Carletti, CD, BuenoI, OC, and PagnoccaI, FC.** 2008. Leaf-cutting ant faecal fluid and mandibular gland secretion: effects on microfungi spore germination. *Braz. J. Microbiol.* **39:**64-67.

87. **Lorenzen, K, and Anke, T.** 1998. Basidiomycetes as a source for new bioactive natural products. *Curr. Org. Chem.* **2:**329-364.

88. **Huff, T, Kuball, HG, and Anke, T.** 1994. 7-Chloro-4,6-dimethoxy-1(3H)-isobenzofuranone and basidalin: antibiotic secondary metabolites from *Leucoagaricus carneifolia* Gillet (basidiomycetes) [corrected]. *Z Naturforsch C* **49:**407-410.

89. **Mead, MI, Khan, AH, Nickless, G, Greally, BR, Tainton, D, B, TP, and Shallcross, DE.** 2008. Leaf cutter ants: a possible missing source of biogenic halocarbons. *Environ. chem.* **5:**5-10.

90. **de Jong, E, and Field, JA.** 1997. Sulfur tuft and turkey tail: biosynthesis and biodegradation of organohalogens by Basidiomycetes. *Annu. Rev. Microbiol.* **51:**375-414.

91. **Gerardo, NM, Jacobs, SR, Currie, CR, and Mueller, UG.** 2006. Ancient host-pathogen associations maintained by specificity of chemotaxis and antibiosis. *PLoS Biol.* **4:**e235.

92. **Cafaro, MJ, Poulsen, M, Little, AE, Price, SL, Gerardo, NM, Wong, B, Stuart, AE, Larget, B, Abbot, P, and Currie, CR.** 2011. Specificity in the symbiotic association between fungus-growing ants and protective *Pseudonocardia* bacteria. *Proc. Roy. Soc. B-Biol. Sci.* **278:**1814-1822.

93. **Currie, CR, Bot, ANM, and Boomsma, JJ.** 2003. Experimental evidence of a tripartite mutualism: bacteria protect ant fungus gardens from specialized parasites. *Oikos* **101:**91-102.

94. **Mueller, UG, Dash, D, Rabeling, C, and Rodrigues, A.** 2008. Coevolution between Attine Ants and Actinomycete Bacteria: a Reevaluation. *Evolution*.

95. **Sen, R, Ishak, HD, Estrada, D, Dowd, SE, Hong, E, and Mueller, UG.** 2009. Generalized antifungal activity and 454-screening of *Pseudonocardia* and *Amycolatopsis* bacteria in nests of fungus-growing ants. *Proc. Natl. Acad. Sci. USA* **106:**17805-10.

96. **Poulsen, M, Cafaro, M, Boomsma, JJ, and Currie, CR.** 2005. Specificity of the mutualistic association between actinomycete bacteria and two sympatric species of Acromyrmex leaf-cutting ants. *Mol. Ecol.* **14:**3597-3604.

97. **Oh, DC, Poulsen, M, Currie, CR, and Clardy, J.** 2009. Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nat. Chem. Biol.* **5:**391-393.

98. **Poulsen, M, Cafaro, M, Erhardt, D, Little, A, NM, G, Tebbets, B, Klein, B, and Currie, C.** 2010. Variation in Pseudonocardia antibiotic defence helps govern parasite-induced morbidity in Acromyrmex leaf-cutting ants. *Environ. Microbiol. Rep.* **2:**534-540.

99. **Seipke, RF, Barke, J, Brearley, C, Hill, L, Yu, DW, Goss, RJ, and Hutchings, MI.** 2011. A single *Streptomyces* symbiont makes multiple antifungals to support the fungus farming ant *Acromyrmex octospinosus*. *PLoS One* **6:**e22028.

100. **Barke, J, Seipke, RF, Gruschow, S, Heavens, D, Drou, N, Bibb, MJ, Goss, RJ, Yu, DW, and Hutchings, MI.** 2010. A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant *Acromyrmex octospinosus*. *BMC Biol.* **8:**109.

101. **Martin, MM, Carman, RM, and Macconnell, JG.** 1969. Nutrients Derived from the Fungus Cultured by the Fungus-Growing Ant *Atta colombica* tonsipes. *Ann. Entom. Soc. Am.* **62:**11.

102. **Mueller, UG, Schultz, TR, Currie, CR, Adams, RM, and Malloch, D.** 2001. The origin of the attine ant-fungus mutualism. *Q. Rev. Biol.* **76:**169-197.

103. **Martin, MM, and Weber, NA.** 1969. The cellulose-utilizing capability of the fungus cultured by the attine ant *Atta colombica tonsipes*. *Ann. Entomol. Soc. Am.* **62:**1386-1387.

104. **Rockwood, LL.** 1975. The effects of seasonality on foraging in two species of leaf-cutting ants (*Atta*) in Guanacaste province, Costa Rica. *Biotropica* **7:**176-193.

105. **Wirth, R, Herz, H, Ryel, RJ, Beyschlag, W, and Hoelldobler, B.** 2003. Herbivory of leaf-cutting ants: A case study on *Atta colombica* in the tropical rainforest of Panama. Springer, New York, NY.

106. **De Fine Licht, HH, and Boomsma, J.** 2010. Forage collection, substrate preparation, and diet composition in fungus-growing ants. *Ecol. Ent.* **35:**259-269.

107. **Martin, MM, Gieselmann, MJ, and Martin, JS.** 1970. Rectal enzymes of attine ants. α-Amylase and chitinase. *J. Insect Physiol.* **19:**1409-1416.

108. **Martin, MM, and Martin, JS.** 1971. The presence of protease activity in the rectal fluid of primitive attine ants. *J. Insect Physiol.* **17:**1897-1906.

109. **Martin, JS, and Martin, MM.** 1970. The presence of protease activity in the rectal fluid of attine ants. *J. Insect Physiol.* **16:**227-232.

110. **Boer, W, Folman, LB, Summerbell, RC, and Boddy, L.** 2005. Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol. Rev.* **29:**795-811.

111. **International Aphid Genomics Consortium.** 2010. Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol.* **8:**e1000313.

112. **Shigenobu, S, Watanabe, H, Hattori, M, Sakaki, Y, and Ishikawa, H.** 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407:**81-86.

# Chapter 2: Metagenomic and Metaproteomic Insights into Bacterial Communities in Leaf-Cutter Ant Fungus Gardens

**Previously Published**: Frank O Aylward, Kristin E Burnum, Jarrod J Scott, Garret Suen, Susannah G Tringe, Sandra M Adams, Kerrie W Barry, Carrie D Nicora, Paul D Piehowski, Samuel O Purvine, Gabriel J Starrett, Lynne A Goodwin, Richard D Smith, Mary S Lipton, Cameron R Currie (2012). Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME Journal*, 6: 1688-1701.

**Author Contributions**: For this chapter I performed all computational experiments, and I conceived of all analyses with Cameron R. Currie, Kristin Burnum, and Mary Lipton. Proteomic analyses were conducted at Pacific Northwest National Laboratories, and metagenomes were generated by the Joint Genome Institute. Other authors collected the samples, provided helpful advice for analyses, and helped with writing.

## 2.1 Abstract

Herbivores gain access to nutrients stored in plant biomass largely by harnessing the metabolic activities of microbes. Leaf-cutter ants of the genus *Atta* are a hallmark example; these dominant Neotropical herbivores cultivate symbiotic fungus gardens on large quantities of fresh plant forage. As the external digestive system of the ants, fungus gardens facilitate the production and sustenance of millions of workers. Using metagenomic and metaproteomic techniques, we characterize the bacterial diversity and physiological potential of fungus gardens from two species of *Atta*. Our analysis of over 1.2 Gbp of community metagenomic sequence and three 16S pyrotag libraries reveals that in addition to harboring the dominant fungal crop, these ecosystems contain abundant populations of *Enterobacteriaceae*, including the genera *Enterobacter, Pantoea*, *Klebsiella*, *Citrobacter*, and *Escherichia*. We show that these bacterial communities possess genes associated with lignocellulose degradation and diverse biosynthetic pathways, suggesting they may play a role in nutrient cycling by converting the nitrogen-poor forage of the ants into B-vitamins, amino acids, and other cellular components. Our metaproteomic analysis confirms that bacterial glycosyl hydrolases and proteins with putative biosynthetic functions are produced in both field-collected and laboratory-reared colonies. These

results are consistent with the hypothesis that fungus gardens are specialized fungus-bacteria communities that convert plant material into energy for their ant hosts. Together with recent investigations into the microbial symbionts of vertebrates, our work underscores the importance of microbial communities to the ecology and evolution of herbivorous metazoans.

## 2.2 Introduction

Ants are critical components of terrestrial ecosystems around the world (1). Among ants, leaf-cutters in the genus *Atta* (Figure 2.1a) are particularly dominant, with mature colonies achieving immense sizes and housing millions of workers (2, 3). Ranging from the southern United States to Argentina, species of leaf-cutter ants can construct elaborate subterranean nests containing hundreds of chambers and displacing up to 40,000 Kg of soil (2). The ant societies housed within these nests are equally impressive, with an intricate division of labor observed between different castes of workers (2). Associated with this division of labor is substantial worker size polymorphism: the dry weight of individual workers in the same colony can differ by 200-fold (2). The success of leaf-cutter ants is largely attributed to their obligate mutualism with a basidiomycetous fungus (*Leucoagaricus gongylophorus*) that they culture for food in specialized gardens (Figure 2.1b) (2-4). Fresh plant forage collected by these ants serve to nourish the fungus, which in turn produces nutrient-rich hyphal swellings (gongylidia) that feed the colony (4). The symbiosis between leaf-cutter ants and their cultivar is thought to have originated 8–12 million years ago, and numerous adaptations in both the ants and the fungus have occurred over this long history of agriculture (4-6).

The fresh foliar biomass leaf-cutter ants integrate into their fungus gardens is composed largely of recalcitrant lignocellulosic polymers. The ants presumably gain indirect access to the

carbon stored in plant cell walls through the metabolic activities of their fungus gardens, which act as an ancillary digestive system (7). Despite being a critical aspect of leaf-cutter ant biology, the process through which fungus gardens degrade plant forage has only recently been intensely investigated (8-11). Originally it was thought that the fungal cultivar primarily degraded cellulose, and that this was the main polymer converted into nutrients for the ants (12). However, the cellulolytic capacity of this fungus has come into question, as it has been shown that pure cultures cannot grow on cellulose as a sole carbon source (13). This has led to the suggestion that cellulose is not deconstructed in leaf-cutter ant fungus gardens, but rather that the fungal cultivar uses a variety of hemicellulases to deconstruct primarily starch, xylan, and other plant polymers (14-17).

Another model posits that plant cell wall degradation in fungus gardens is partially mediated by lignocellulolytic bacteria. There is some support for this model. Importantly, recent work has found evidence for substantial cellulose deconstruction in the fungus gardens of *Atta colombica* and the presence of lignocellulolytic bacteria in these ecosystems (11). Another study, employing the culture-independent analysis of membrane lipid markers, has supported the hypothesis that a distinct community of predominantly Gram-negative bacteria resides in fungus gardens (18), and the presence of symbiotic nitrogen-fixing bacteria in the genera *Pantoea* and *Klebsiella* has also been established (7). Together with culture-dependant investigations recovering microbial groups with a broad array of metabolic activities (19, 20), these experiments have led to the suggestion that fungus gardens represent specialized fungus-bacteria consortia selected for by the ants, and that bacteria play essential roles including plant biomass degradation, nutrient biosynthesis, and competitive or antibiotic-mediated exclusion of pathogens (7, 11, 21, 22).

Using a combination of metagenomics and metaproteomics, we provide insights into the microbial activities in leaf-cutter ant fungus gardens. Culture-independent investigations have previously been performed on leaf-cutter ant fungus gardens (11, 18), but to date, only a small quantity of bacterial sequences (~6Mb) from the fungus gardens of a single ant species have been characterized. Here, by expanding on previous work, we sought to document the non-eukaryotic component of fungus gardens, describe the similarity of communities from different ant species, and examine potential microbial activities *in situ*. To this end, we generated three 16S pyrotag libraries of over 8,000 sequences each and over 1.2 Gbp of raw 454 Titanium community metagenomic data from the bacterial component of *Atta cephalotes* and *Atta colombica* fungus gardens. To account for potential differences in microbial communities due to the extent of plant biomass degradation, we individually examined the top and bottom strata of *A. colombica* fungus gardens, which correspond to where the ants integrate fresh forage and remove partially-degraded plant substrate, respectively. We then conducted metaproteomic analyses on whole fungus gardens to identify proteins produced in these ecosystems and examine the physiology of resident bacteria in more detail. We find that similar bacterial communities inhabit all fungus garden samples analyzed, and that the metabolic potential of resident bacteria includes nutrient biosynthesis, hemicellulose and oligosaccharide degradation, and other functions that potentially enhance plant biomass processing in these ecosystems. We discuss a novel framework for understanding the complex interplay between leaf-cutter ants and the symbiotic communities residing in their fungus gardens.

## 2.3 Materials and Methods

### 2.3.1 Sample Processing for Community Metagenomes and 16S Pyrotag Libraries

Fungus gardens from healthy *A. cephalotes* and *A. colombica* colonies were collected from nests near Gamboa, Panama, in April 2009. Whole *A. cephalotes* gardens were combined for subsequent analyses, while fungus gardens of *A. colombica* were laterally bisected to separate top and bottom strata. Immediately after collection, the bacterial fraction of the samples was isolated and DNA was extracted as previously described (11). Briefly, plant, ant, and fungal material were removed from all samples through a series of wash/centrifugation steps using 1X PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 2 mM $KH_2PO_4$). DNA was subsequently extracted from the remaining bacterial fraction using a Qiagen DNeasy Plant Maxi Kit (Qiagen Sciences, Germantown, MD, USA). One community metagenome and one 16S library were generated from each of the three samples using 454 Titanium pyrosequencing technology (23). Draft genomes of three bacteria isolated from *Atta* fungus gardens were also generated to supplement the reference databases used for the phylogenetic binning of metagenomic data. Technical details for the sequencing, assembly, and annotation of all data can be found in the Appendix 1.

### 2.3.2 Metaproteomics

Metaproteomic analysis was conducted on fungus garden material collected in Gamboa, Panama, from a nest *A. colombica* distinct from that used for metagenomic analyses. Moreover, we also conducted metaproteomic analyses on a lab-reared colony of *A. sexdens* for comparison. Detailed methods can be found in Appendix 1. Briefly, proteins were extracted from whole fungus garden material, and the resulting protein solution was digested into peptides

subsequently analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS). The resulting peptide tandem mass spectra were compared to predicted protein datasets of the three community metagenomes individually. Peptide matches were filtered using Sequest (24) scores, MS-GF spectral probabilities (25), and false discovery rates. We restricted our functional analyses to peptides mapped to proteins phylogenetically binned as bacterial, and the IMG-ER and KEGG annotations of these proteins were inspected to identify those potentially involved in biomass degradation or nutrient cycling (Table 2.5). Peptides mapping to these select proteins were inspected manually (Figure 2.5).

## 2.4 Results

### 2.4.1 Community Metagenomes and 16S Pyrotag Libraries

Pyrosequencing of the V6–V8 variable region of the bacterial 16S rRNA gene for the same three samples yielded between 8,000–12,000 reads (termed "pyrotags") each (Table 2.2). Previous attempts to recover Archaeal 16S sequences from fungus gardens were unsuccessful (11), and amplification of these genes was not attempted here. Pyrosequencing of community DNA from the three samples, representing both the individual top and bottom strata of *A. colombica* fungus gardens as well as the combined strata of *A. cephalotes* gardens, each yielded 382–441 Mb of raw sequence data (Table 2.1). Reads from each library were assembled into community metagenomes comprising 40–100 Mbp of sequence data.

### 2.4.2 Microbial Diversity in Fungus Gardens

Clustering of sequences in the 16S pyrotag libraries from the *A. colombica* top, *A. colombica* bottom, and *A. cephalotes* fungus garden samples recovered 204, 274, and 25 operational taxonomic units (OTUs, 97% identity cutoff), respectively. The majority of the OTUs

were most similar to sequences of Gammaproteobacteria and Firmicutes (22–217 OTUs, 72–89% of OTUs, and 2–35 OTUs, 5–12% of OTUs, respectively), and only OTUs corresponding to those phyla were represented in all three samples (Figure 2.1c). Phyla represented in lower abundance and more sporadically included the Betaproteobacteria (≤10 OTUs, ≤4.9% of OTUs), Alphaproteobacteria (≤7 OTUs, ≤3.4% of OTUs), Bacteriodetes (≤4 OTUs, ≤2% of OTUs), Acidobacteria (≤1 OTU, ≤1% of OTUs), and Actinobacteria (≤1 OTU, ≤4% of OTUs) (Figure 2.1c). Most pyrotags corresponded to the Gammaproteobacterial family *Enterobacteriaceae* (79–99% of individual pyrotags, Table 2.2). Although taxonomic profiles were similar between all three pyrotag libraries, the bacterial diversity of each of the *A. colombica* samples was greater than that recovered from the *A. cephalotes* fungus garden sample.

**Table 2.1.** Sequencing statistics of the community metagenomes.

| **Ant Species** | ***A. colombica* Top** | ***A. colombica* Bottom** | ***A. cephalotes*** |
|---|---|---|---|
| # or trimmed reads | 998,047 | 862,246 | 1,068,791 |
| Amount of Raw Sequence (Mbp) | 441.2 | 382.1 | 431.0 |
| # of contigs | 28,034 | 21,203 | 17,914 |
| Largest contig (Kbp) | 359.8 | 361.9 | 168.3 |
| N50 Contig size (Kbp) | 1.8 | 1.8 | 4.7 |
| Number of Singleton Reads | 188,523 | 161,267 | 55,949 |
| Protein Coding Genes | 240,966 | 199,019 | 73,881 |
| Size of Assembled Data (Mbp) | 100.9 | 83.2 | 40.6 |

Community metagenomic analyses recovered primarily bacterial sequences (71–80% of total assembled bp) (Table 2.3). Consistent with the 16S pyrotag libraries, the majority of sequences in all 3 datasets matched most closely to Gammaproteobacteria (69–72%), especially *Enterobacteriaceae* (53–70%). To refine taxonomic resolution and infer the relative abundance

of microbial groups, raw reads were phylogenetically classified using the Genome relative Abundance and Average Size (GAAS) tool (26). Estimates based on GAAS analyses indicate that the Gammaproteobacteria were particularly abundant, with the genus *Enterobacter* comprising over 50% of the bacterial population in all 3 metagenomes (Figure 2.1d). The community metagenomes also contained representative sequences from the genera *Klebsiella* (3.8–4.9%), *Pantoea* (1.8–15.6%), *Escherichia* (5.3–6.3%), *Citrobacter* (3–5.8%), *Pseudomonas* (0.04–4.2%), and *Lactococcus* (0.01–2.2%). BLAST-based classification of the assembly indicated that ~1% of the sequences corresponded to bacteriophage in each of the community metagenomes (Table 2.3), whereas the GAAS tool estimated that 15.1%, 16.8%, and 6.8% of the *A. colombica* top, *A. colombica* bottom, and *A. cephalotes* metagenomes could be comprised of bacteriophage, respectively.

Consistent with the GAAS- and BLAST-based analyses, the largest phylogenetic bins created by phymmBL were assigned to the genera *Enterobacter, Pantoea, Klebsiella, Escherichia, Citrobacter,* and *Pseudomonas*. The *Enterobacter* bins were by far the largest, containing 15.3-29.5 Mb of sequence. The majority of these sequences were most similar to the draft genome of *Enterobacter* FGI 35, a strain isolated in this study from an *Atta colombica* fungus garden. The *Pantoea* bins were the next largest, containing between 5-7.2 Mb of sequence each.

### 2.4.3 Metabolic Potential of Bacterial Lineages

To compare the coding potential of different bacterial groups in fungus gardens, we analyzed genus-level phylogenetic bins of sequences constructed from the community metagenomes. Comparison of the coding potential in the bins to the Kyoto Encyclopedia of

Genes and Genomes (KEGG, (27)) recovered well-represented sugar metabolism pathways in most of the *Enterobacteriaceae* bins (Figure 2.2). Moreover, pathways involved in B-vitamin and amino acid metabolism were found to be highly-represented in both the *Pseudomonas* and *Enterobacteriaceae* bins. The *Lactococcus* bins showed relatively low representation in most of these pathways. Clustering of phylogenetic bins from each of the metagenomes by their KEGG pathway representation indicated that bacterial members corresponding to the same genus, with the exception of *Citrobacter*, had similar metabolic profiles.

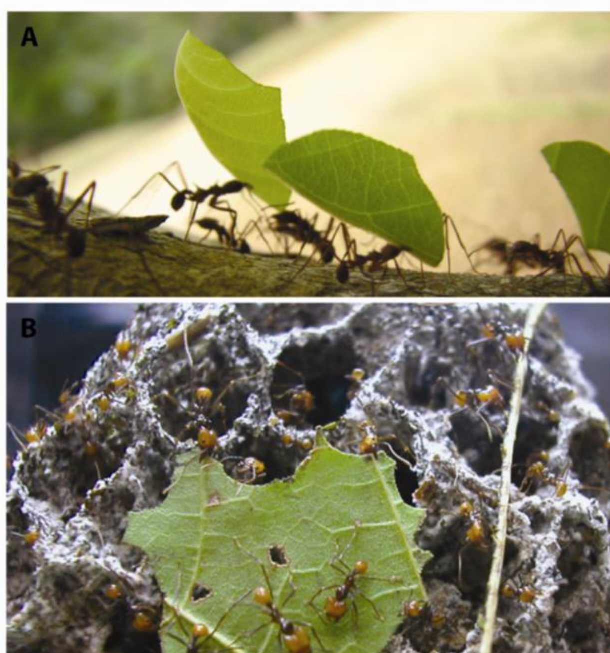
To examine how leaf-cutter ant fungus garden microbial communities differed from other environments, we predicted Clusters of Orthologous Groups (COGs, (28)) from all contigs and reads from the 3 fungus garden metagenomes and compared these with COG profiles from all other metagenomes available on the Integrated Microbial Genomes/Microbiomes database (IMG/M, (29)) (Figure 2.3). COG profiles for the three fungus garden metagenomes were found to be highly similar. Compared to all other metagenomes on IMG, many COG categories were over-represented in fungus gardens (Fisher's exact test, $P < 0.01$), including Amino Acid Transport and Metabolism, Carbohydrate Transport and Metabolism, and Inorganic Ion Transport and Metabolism (Figure 2.3). Specific COGs involved in Carbohydrate Transport and Metabolism were analyzed in more detail to investigate possible bacterial roles in polysaccharide degradation, and sugar transporters and phosphotransferase system components in particular were found to be significantly over-represented in the fungus garden metagenomes (Fisher's exact test, $P < 0.01$).

**Table 2.2.** Family-level classification of partial-length 16S sequences recovered from *Atta colombica* and *Atta cephalotes* fungus gardens.

| | ***A. colombica* Top** | | | ***A. colombica* Bottom** | | | ***A. cephalotes*** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Family** | **OTUs** | **Clones** | **% OTUs** | **OTUs** | **Clones** | **% OTUs** | **OTUs** | **Clones** | **% OTUs** |
| *Acetobacteraceae* | 1 | 3 | 0.49 | 2 | 36 | 0.73 | 1 | 2 | 4.00 |
| *Aeromonadaceae* | 5 | 13 | 2.45 | 0 | 0 | 0.00 | 0 | 0 | 0.00 |
| *Alcaligenaceae* | 1 | 7 | 0.49 | 2 | 3 | 0.73 | 0 | 0 | 0.00 |
| *Aurantimonadaceae* | 0 | 0 | 0.00 | 0 | 0 | 0.00 | 0 | 0 | 0.00 |
| *Bacillaceae* | 5 | 45 | 2.45 | 8 | 156 | 2.92 | 1 | 5 | 4.00 |
| *Carnobacteriaceae* | 3 | 197 | 1.47 | 6 | 273 | 2.19 | 1 | 2 | 4.00 |
| *Clostridiaceae* | 3 | 3 | 1.47 | 1 | 1 | 0.36 | 0 | 0 | 0.00 |
| *Comamonadaceae* | 7 | 474 | 3.43 | 5 | 296 | 1.82 | 0 | 0 | 0.00 |
| *Enterobacteriaceae* | 129 | 6,502 | 63.24 | 202 | 10,706 | 73.72 | 19 | 10,496 | 76.00 |
| *Enterococcaceae* | 5 | 66 | 2.45 | 0 | 0 | 0.00 | 0 | 0 | 0.00 |
| *Flavobacteriaceae* | 3 | 33 | 1.47 | 2 | 21 | 0.73 | 0 | 0 | 0.00 |
| *Moraxellaceae* | 3 | 167 | 1.47 | 2 | 64 | 0.73 | 1 | 2 | 4.00 |
| *Paenibacillaceae* | 2 | 15 | 0.98 | 3 | 3 | 1.09 | 0 | 0 | 0.00 |
| *Pseudomonadaceae* | 10 | 335 | 4.90 | 9 | 345 | 3.28 | 1 | 3 | 4.00 |
| *Ruminococcaceae* | 3 | 7 | 1.47 | 4 | 14 | 1.46 | 0 | 0 | 0.00 |
| *Sphingomonadaceae* | 3 | 11 | 1.47 | 2 | 4 | 0.73 | 0 | 0 | 0.00 |
| *Staphylococcaceae* | 4 | 13 | 1.96 | 4 | 6 | 1.46 | 0 | 0 | 0.00 |
| *Veillonellaceae* | 0 | 0 | 0.00 | 4 | 96 | 1.46 | 0 | 0 | 0.00 |
| *Xanthomonadaceae* | 4 | 134 | 1.96 | 2 | 60 | 0.73 | 1 | 1 | 4.00 |
| Other | 13 | 163 | 6.00 | 16 | 86 | 6.00 | 0 | 0 | 0.00 |
| Total | 204 | 8,188 | 100 | 274 | 12,170 | 100 | 25 | 10,511 | 100 |

To further investigate potential bacterial roles in plant polymer deconstruction, we compared all predicted proteins in the three community metagenomes to the Carbohydrate Active Enzymes database (CAZy, (30)) and identified numerous enzymes potentially involved in this process (Table 2.4). The largest proportion of the identified proteins were most similar to oligosaccharide-degrading enzymes (176–566 CAZymes, 28–30%), and relatively few were found to be predicted cellulases (4–5 CAZymes, 0.2–0.6%). Compared to other well-known

lignocellulose-degrading communities such as the Tammar wallaby foregut (31) and termite hindgut (32), fungus gardens contained relatively fewer cellulases and hemicellulases, but similar numbers of oligosaccharide-degrading enzymes.



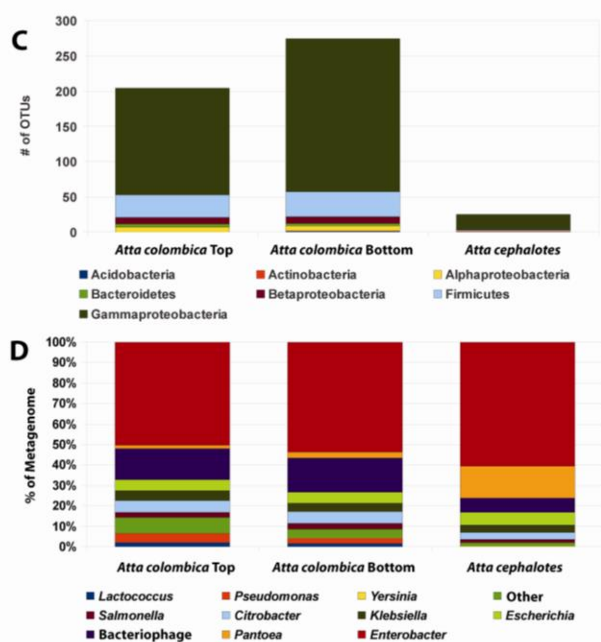




**Figure 2.1.** Leaf cutter ants forage on plant material (A) that they use as manure for specialized fungus gardens (B). Pyrosequencing of bacterial 16S genes from fungus gardens of the leaf-cutter ants *Atta colombica* and *Atta cephalotes* recovered 8,000-12,000 sequences representing 25-274 operational taxonomic units (OTUs, 97% identity cutoff) (C). Microbial community composition was then investigated by directly pyrosequencing 382-441 Mbp of DNA from the same leaf-cutter ant fungus gardens and using the Genome Abundance and Average Size tool to estimate the relative abundance of different microbial groups (D). [Photo credits: A; Jarrod J. Scott, B; Austin D. Lynch.]

### 2.4.4 Comparison of *Enterobacter* Populations

To identify similarities between the *Enterobacter* populations across different metagenomes, we performed a fragment recruitment analysis comparing all predicted genes from the *Enterobacter* FGI 35 phylogenetic bins from each metagenome to the draft *Enterobacter* FGI 35 genome (Figure 2.4). The fragment

recruitment analysis identified near-uniform coverage of >95% nucleic acid identity BLAST hits across the 33 *Enterobacter* FGI 35 contigs, with the exception of 4 regions between 18-66 Kb large that we termed Variable Regions (VR) I-IV. Moreover, we found that there was also near-uniform coverage of 70-85% identity BLASTN hits across the draft genome. Investigation of the coding potential in these conserved regions identified genes required for the synthesis of thiamine, pyridoxine, nicotinate, nicotinamide, pantothenate, folate, and 19 amino acids. Only the later stages of the histidine biosynthetic pathway could be identified, although the full pathway is present in other *Enterobacter* contigs. These regions also encoded ABC transporters and phosphotransferase system components predicted to uptake cellobiose, xylose, glucose, sucrose, β-glucosides, arbutin/salicin, N-acetylmuramic acid, mannitol, mannose, sorbitol, galactitol, L-ascorbate, fructose, ribose, L-arabinose, methyl-galactoside, sulfate, sulfonate, spermidine/putrescine, 2-Aminoethylphosphonate, iron, and other nutrients. The variable regions were found to contain primarily hypothetical genes and genes of unknown function, although some phage integrases were also identified.

### 2.4.5 Metaproteomics

Individual searches of the metaproteomic data against the predicted protein databases of each community metagenome recovered a total of 1186 redundant and 869 non-redundant peptides. Of all distinct peptides recovered, 129 were found in both lab and field samples, while 351 were unique to the laboratory sample and 389 were unique to the field sample. A total of 747, 238, and 201 peptides were recovered for the searches against the *Atta cephalotes, Atta colombica* top, and *Atta colombica* bottom datasets, respectively. These peptides were mapped onto a total of 653 proteins, 354 of which were predicted from contigs or singletons that were

**Table 2.3.** Phylogenetic classification of all assembled contigs and singletons in the leaf-cutter ant fungus garden metagenomes. BLASTN was used to compare all sequences to NCBI's non-redundant nucleotide database.

| | ***A. cephalotes*** | ***A. colombica* Top** | ***A. colombica* Bottom** |
|---|---|---|---|
| **Classification** | | Kb of Sequence (% of Assembly) | |
| Bacteria | 29,058.8 (71.5%) | 78,823.2 (78.1%) | 66,803.1 (80.3%) |
| Proteobacteria | 28,985.3 (71.4%) | 75,187.6 (74.5%) | 63,493.9 (76.3%) |
| Gammaproteobacteria | 28,697.4 (70.6%) | 69,374 (68.8%) | 59,665.8 (71.7%) |
| *Enterobacteriaceae* | 28225.4 (69.5%) | 53004.3 (52.5%) | 49280.1 (59.3%) |
| *Pseudomonadaceae* | 222.8 (0.55%) | 8973.2 (8.9%) | 6741.7 (8.1%) |
| Betaproteobacteria | 185.6 (0.45%) | 4,940.9 (4.9%) | 3171 (3.8%) |
| Alphaproteobacteria | 99.4 (0.24%) | 822.5 (0.82%) | 612.1 (0.74%) |
| Firmicutes | 29.1 (0.07%) | 3,167.7 (3.1%) | 3,043.64 (3.6%) |
| Actinobacteria | 19.4 (0.05%) | 106.6 (0.11%) | 82.5 (0.01%) |
| Bacteroidetes | 17.9 (0.04%) | 281.8 (0.28%) | 124.1 (0.15%) |
| | | | |
| Eukaryota | 661.3 (1.6%) | 145.1 (0.14%) | 133.5 (0.16%) |
| Fungi | 495.6 (1.2%) | 14.2 (0.01%) | 25.1 (0.03%) |
| Metazoa | 19.8 (0.05%) | 81.5 (0.08%) | 66.4 (0.08%) |
| Viridiplantae | 143.4 (0.35%) | 35.6 (0.04%) | 35.2 (0.04%) |
| | | | |
| Viruses | 102.2 (0.25%) | 727.6 (0.72%) | 947.6 (1.1%) |
| dsDNA viruses | 85.9 (0.21%) | 713.9 (0.71%) | 579.6 (0.7%) |
| ssDNA viruses | 15.3 (0.04%) | 11.6 (0.01%) | 367.25 (0.44%) |
| | | | |
| Other | 60.1 (<0.01%) | 132.8 (<0.01%) | 105.8 (<0.01%) |
| Unclassified | 10,741 (26.4%) | 21,076.2 (20.9%) | 15,177.1 (18.2%) |

phylogenetically binned as bacterial (see Appendix 1 for details on the phylogenetic binning procedure). The majority of bacterial proteins identified were predicted to belong to the *Enterobacteriaceae*, and functions predicted from these proteins included a variety of metabolic processes (Table 2.5). Figure 2.5 highlights the overlap observed between laboratory-reared and field-collected samples for one peptide mapped to a predicted glycosyl hydrolase.

**2.5 Discussion**

Leaf-cutter ants are dominant New World herbivores, foraging on up to 17% of the foliar biomass in some ecosystems (33). In 1874 Thomas Belt established that leaf-cutters do not consume leaf material directly, as had been previously assumed, but instead use it as manure to cultivate a fungus for food in specialized gardens (34). For over a hundred years after Belt's pioneering discovery it was believed that the fungus gardens of leaf-cutter ants represented a monoculture of the fungal cultivar that degraded plant cell wall material and converted it into nutrients for the ants (4, 12). However, both the lignocellulolytic capacity of the cultivar and the view that fungus gardens are composed solely of the fungal mutualist have been recently challenged (11, 13, 15, 18). In this study, we explore the hypothesis that bacteria are common constituents of fungus gardens that could be participating in plant biomass degradation and nutrient cycling.

Our work demonstrates that a distinct community of bacteria resides in the fungus gardens of *A. colombica* and *A. cephalotes* leaf-cutter ants. Our identification of similar bacterial groups in fungus garden samples taken from different ant species and garden strata supports this conclusion. Moreover, this is consistent with our finding that relatively few bacterial genera comprise the majority of the metagenomic sequences recovered in this study (see below). This, combined with previous work on nitrogen fixation, plant biomass degradation, and membrane lipid profiles in these ecosystems, indicates that bacteria are long-term residents of fungus gardens and not merely allochthonous organisms introduced from leaf material or the surrounding soil (7, 11, 18, 19). Thus, the term "fungus garden" may be misleading, as these environments are in fact composed of a fungus-bacteria community.

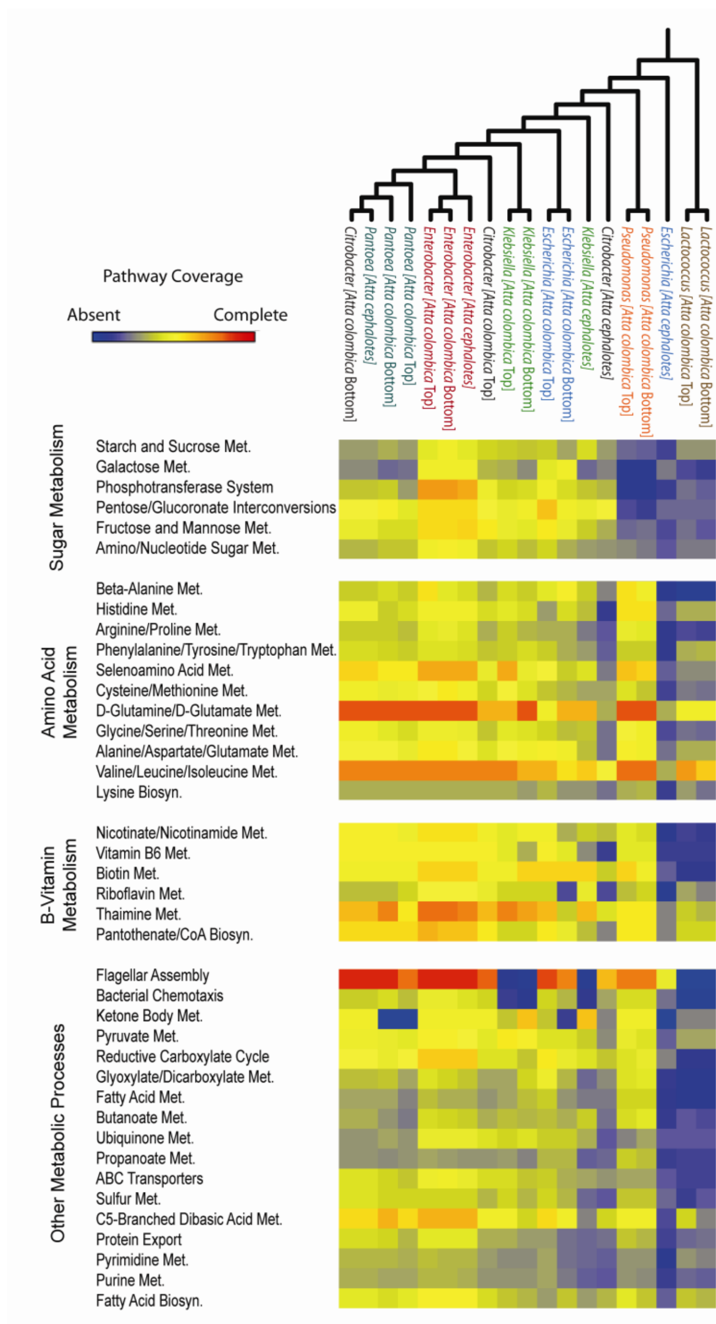



**Figure 2.2.** Reconstruction of KEGG pathways recovered from phylogenetic bins generated from the leaf-cutter ant fungus garden metagenomes. KEGG profiles normalized by the number of predicted proteins in each phylogenetic bin were used for the clustering analysis. Pathways involved in the metabolism of carbohydrates, amino acids, and B-vitamins were among the most highly represented, and are shown here.

The bacterial component of the microbial ecosystem in fungus gardens appears to be dominated by only a few groups. Specifically, the genera *Enterobacter, Klebsiella, Citrobacter, Escherichia,* and *Pantoea* represent over two-thirds of the bacterial component in each of the community metagenomes (Figure 2.1d). This narrow genus-level diversity is likely the result of both the nutrient composition of the plant-fungal matrix and the meticulous hygienic practices of the ants. For example, leaf-cutters continuously weed their gardens to remove areas infected

with microbial pathogens (35), and also apply antimicrobials derived from both glandular secretions and symbiotic Actinobacteria (36, 37). The extent of plant biomass degradation could also affect microbial diversity, but if this was a critical factor we would expect to find distinct communities between top and bottom garden strata, which contain fresh leaf material and largely

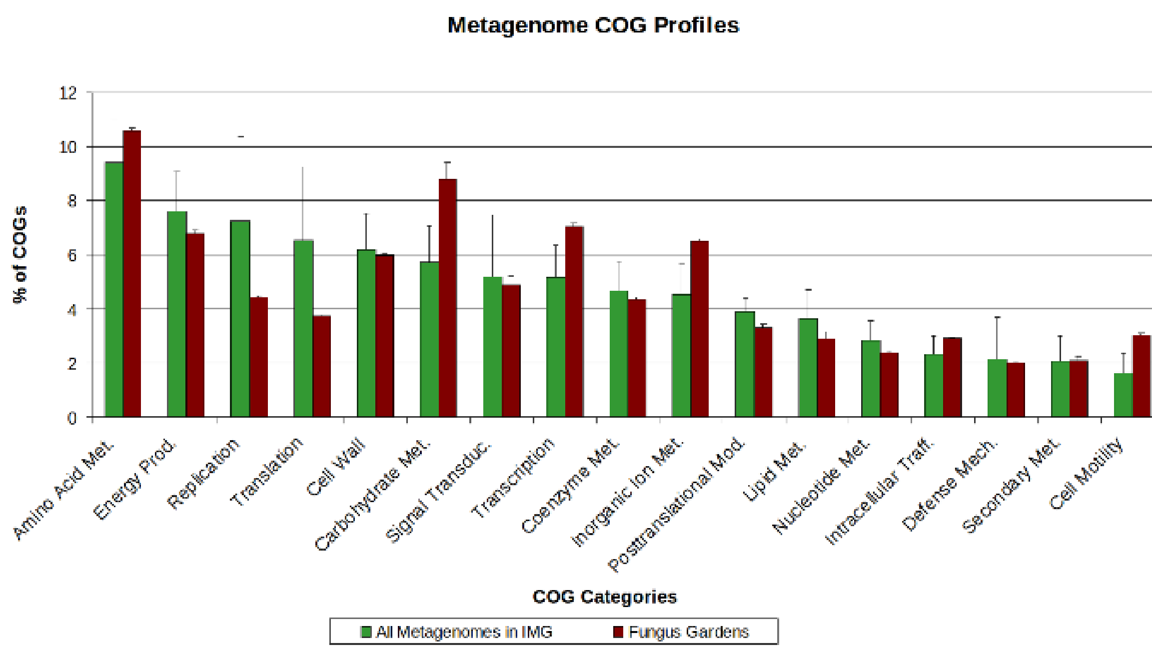



**Figure 2.3.** Comparison of the COG category distributions of the three combined fungus garden metagenomes (*A. colombica* Top, *A. colombica* Bottom, and *A. cephalotes* combined) and all other metagenomes available in IMG. The average COG values are shown ±SD.

degraded biomass, respectively. The similarity between different strata observed here, consistent with previous work reporting little difference between 16S libraries constructed from these two regions (11), indicates that the extent of plant biomass degradation is not a major contributor to community structuring. The consistent presence of bacterial groups within the *Enterobacteriaceae* throughout different garden strata and leaf-cutter ant species implicates them

as having a consistent role in fungus gardens, and suggests that these environments represent highly-structured communities rather than a random collection of opportunistic microbes. Although it remains a possibility that while removing the fungal matrix and plant debris from fungus gardens our analysis excluded microbial groups adhering to fungal or plant biomass, thereby skewing the composition of the metagenomes, our results are generally consistent with previous culture-independent investigations that either analyzed whole fungus gardens or utilized different methods to isolate bacterial cells (11, 18). Moreover, our processing of fresh rather than frozen fungus garden material may be partially responsible for our success in removing fungal or plant debris from our samples.

Bacteria of the genus *Enterobacter* appear to be particularly prevalent in the fungus gardens. In contrast to the narrow genus-level diversity observed in these environments, multiple species of *Enterobacter* appear to be present in all gardens analyzed. Our fragment recruitment analysis demonstrates that populations of bacteria with >95% and 70-85% nucleic acid identity to the reference *Enterobacter* FGI 35 genome exist in these environments (Figure 2.4). The four large gaps identified in the recruitment plot likely represent prophage or other variable elements in the reference genome. Because *Enterobacter* FGI 35 was isolated from an *Atta colombica* fungus garden, the near-uniform coverage of genes at > 95% identity across all metagenomes indicates that highly similar strains of *Enterobacter* are present in all of the samples analyzed. The near-uniform coverage of genes at 70-85% identity likely represents multiple distinct

**Table 2.4.** Partial list of CAZymes identified in the leaf-cutter ant fungus garden metagenomes, as compared with those found in the termite hindgut and wallaby foregut. The raw number of enzymes is given, as well as the percent of all CAZymes identified in individual metagenomes.

| | ***At. col* Top** | ***At. col.* Bottom** | ***At. ceph.*** | **Wallaby** | **Termite** | **Known Activity** |
|---|---|---|---|---|---|---|
| **Cellulases** | | | | | | |
| GH5 | 5 | 2 | 2 | 20 | 97 | Cellulase, mannnosidase |
| GH6 | 0 | 0 | 2 | 0 | 0 | Endocellulse, cellobiohydrolase |
| GH9 | 0 | 2 | 0 | 4 | 39 | Endocellulase, cellobiohydrolase |
| GH44 | 0 | 0 | 0 | 0 | 4 | Endoglucanase, xyloglucanase |
| GH45 | 0 | 0 | 0 | 0 | 6 | Endoglucanase |
| **Total** | **5 (0.26%)** | **4 (0.24%)** | **4 (0.63%)** | **24 (2.3%)** | **146 (10.1%)** | |
| **Hemicellulases** | | | | | | |
| GH8 | 36 | 30 | 13 | 2 | 17 | Cellulase, xylanase, chitosanase |
| GH10 | 0 | 1 | 3 | 18 | 92 | Xylanase |
| GH11 | 0 | 0 | 0 | 0 | 18 | Xylanase |
| GH28 | 13 | 11 | 3 | 10 | 13 | Polygalacturonase |
| GH26 | 1 | 0 | 0 | 8 | 19 | Xylanase, mannanase |
| GH53 | 1 | 4 | 0 | 3 | 5 | Endogalactanase |
| **Total** | **51 (2.6%)** | **46 (2.8%)** | **19 (2.9%)** | **41 (3.9%)** | **164 (11.3%)** | |
| **Debranching enzymes** | | | | | | |
| GH51 | 4 | 2 | 0 | 18 | 26 | Arabinofuranosidase |
| GH67 | 2 | 0 | 0 | 0 | 6 | Glucuronidase |
| GH78 | 25 | 14 | 4 | 52 | 7 | Rhamnosidases |
| **Total** | **31 (1.6%)** | **16 (0.98%)** | **6 (0.63%)** | **70 (6.8%)** | **39 (2.7%)** | |
| **Oligosaccharide degrading enzymes** | | | | | | |
| GH1 | 256 | 243 | 77 | 84 | 27 | Glucosidase, galactosidase, mannosidase |
| GH2 | 23 | 22 | 8 | 33 | 30 | Galactosidase, mannosidase, glucuronidase |
| GH3 | 95 | 76 | 29 | 98 | 108 | Glucosidase, xylosidase |
| GH4 | 61 | 55 | 35 | 3 | 17 | Galactosidase, glucosidase |
| GH29 | 20 | 4 | 0 | 5 | 12 | Fucosidase |
| GH35 | 11 | 1 | 0 | 10 | 6 | Galactosidase, glucosaminidase, |
| GH36 | 26 | 24 | 9 | 29 | 4 | Galactosidase, N-acetylgalactosaminidase |
| GH38 | 21 | 11 | 4 | 3 | 26 | Mannosidase |
| GH39 | 1 | 0 | 0 | 3 | 11 | Xylosidase, iduronidase |
| GH42 | 13 | 14 | 3 | 17 | 34 | Galactosidase |
| GH43 | 39 | 36 | 11 | 3 | 57 | Arabinase, xylosidase |
| GH52 | 0 | 0 | 0 | 0 | 3 | Xylosidase |
| **Total** | **566 (28.9%)** | **486 (29.6%)** | **176 (27.6%)** | **288 (27.9%)** | **335 (23.1%)** | |

species, since it is improbable that a single population of bacteria would have such a large range of nucleotide identifies to a single reference genome. Genes 70-85% identical to the *Enterobacter* FGI 35 genome may represent divergent *Enterobacter* species or even novel *Enterobacteriaceae* for which an appropriate reference for phylogenetic binning does not exist. That different species of leaf-cutter ant harbor abundant *Enterobacter* populations indicates that this group may be an important constituent of the fungus garden community.

The overall functional potential of the metagenomes includes a diversity of bacterial genes associated with plant biomass degradation, supporting previous work that has suggested bacteria play a role in this process. The vast majority of CAZymes identified in the metagenomes are associated with oligosaccharide degradation or simple sugar metabolism, suggesting that bacteria are processing partially-degraded plant material. We also found KEGG pathways involved in hexose and pentose sugar metabolism to be highly represented in the *Enterobacteriaceae*, indicating that sugar monomers can be readily metabolized by many of these bacteria. Moreover, our KEGG, COG, and metaproteomic analyses recovered numerous sugar transporters (Figure 2.2, Table 2.5), including a large number of cellobiose-specific phosphotransferase system components that are known to be involved in the uptake of the by-products of cellulose hydrolysis (Figure 2.1, Table 2.5). Together these data suggest that bacterial community members are metabolizing predominantly partially-degraded plant material, although it remains a possibility that unidentified bacterial lignocellulases are also playing a role in the degradation of plant biomass.

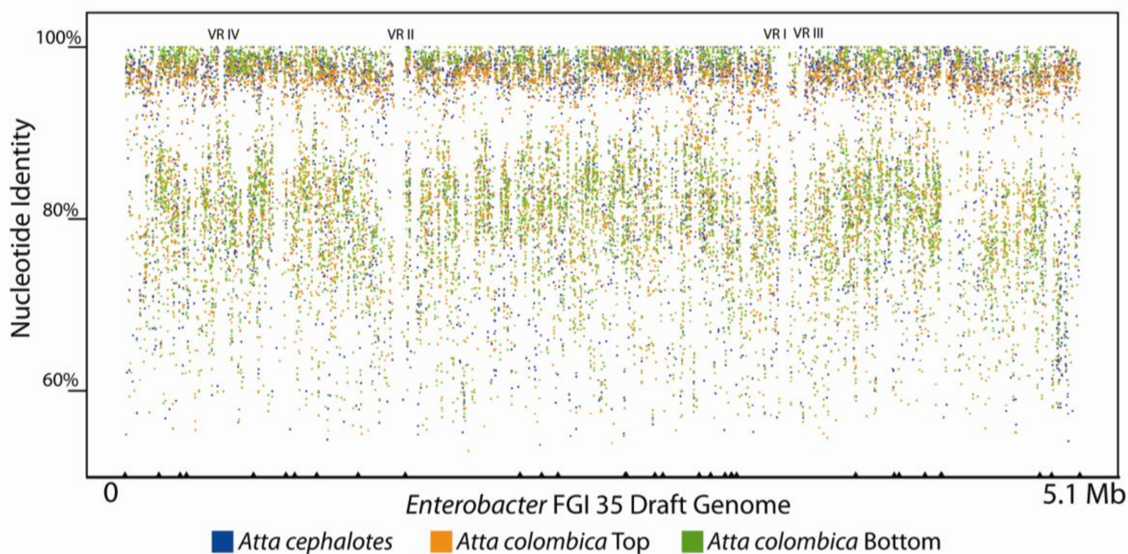



**Figure 2.4.** Fragment recruitment analysis of genes phylogenetically binned to *Enterobacter* FGI 35 against the draft *Enterobacter* FGI 35 genome. Each point indicates the best BLASTN match of a gene. Tick marks on the bottom indicate contig boundaries of the FGI 35 draft genome. Regions showing little or no coverage in the recruitment are marked on the top.

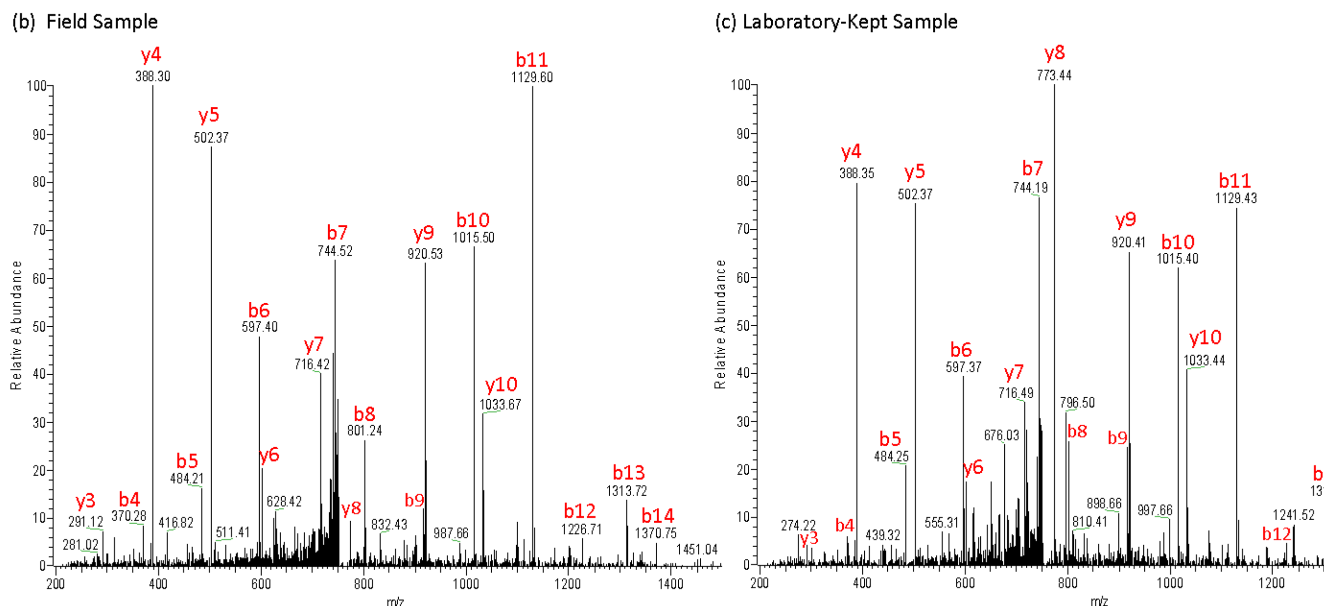
Bacterial lineages in fungus gardens were also found to possess diverse biosynthetic pathways. Pathways involved in amino acid and B-vitamin metabolism were particularly well-represented in the detected *Enterobacteriaceae* and *Pseudomonas* sequences, and biosynthetic pathways for thiamin, pyridoxine, nicotinate, nicotinamide, pantothenate, folate, and all 20 amino acids could be reconstructed from the *Enterobacter* bins. As mentioned above, enzymes involved in the metabolism of oligosaccharides and simple sugars were also identified in many of these groups, indicating that they may convert carbon-rich plant biomass into amino acids, B-vitamins, proteins, or other nutrients. Previous work has indicated that bacteria play a role in the introduction and cycling of nitrogen in fungus gardens (7). Together with our work, this suggests that the combined metabolism of resident bacteria may enrich the nutrient composition of fungus

gardens through the conversion of carbohydrate-rich oligosaccharides into a variety of other nutrients that could promote the growth of the fungal cultivar or even nourish the ants themselves.

Our metaproteomic analysis recovered peptides mapping to bacterial proteins predicted to participate in biomass degradation and nutrient biosynthesis, supporting the results of our metagenomic characterization and further indicating that bacteria are involved in these processes (Table 2.5). Our manual inspection of the metaproteomic data identified multiple peptides belonging to glycoside hydrolases, sugar transporters, and amino acid and B-vitamin biosynthetic pathways. That multiple peptides could be assigned to proteins with similar predicted functions indicates that these processes may be prevalent in fungus gardens. Moreover, many of the mapped peptides originated from both laboratory-reared and field-collected samples, including one which belonged to a family 3 glycosyl hydrolase (Figure 2.5). Although these data should be interpreted cautiously due to the few bacterial proteins identified overall, this may indicate physiological similarities between bacteria in laboratory-reared versus field-collected colonies.

Not all bacteria in fungus gardens were found to have substantial biosynthetic capacity, and in particular the *Lactococcus* groups appear to have limited coding potential in the majority of pathways analyzed. This may be a result of lower sequencing coverage, as only a relatively small fraction of the metagenomes was predicted to belong to these groups. Alternatively, these groups may not be contributing substantially to nutrient cycling and are able to subsist on free sugars and other nutrients available in fungus gardens. Importantly, the by-products of *Lactococci* metabolism may acidify fungus gardens and contribute to the maintenance of the lower pH in these ecosystems, which has previously been observed at 4.4–5.0 (38). The

regulation of the pH of fungus gardens to this narrow range has been hypothesized to be critical to the growth of the fungal cultivar, but the mechanism through which this occurs has remained unknown (38). Few peptides from our metaproteomic datasets were recovered from this group, indicating they may be present in low abundance.



(a)

| # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| b ions | | 185.13 | 256.17 | 371.19 | 484.28 | 597.36 | 744.43 | 801.45 | 916.48 | 1015.55 | 1129.59 | 1226.64 | 1313.67 | 1370.7 |
| Peptide | A | I | A | D | L | L | F | G | D | V | N | P | S | G |
| y ions | 1516.80 | 1445.76 | 1332.68 | 1261.64 | 1146.61 | 1033.53 | 920.45 | 773.38 | 716.36 | 601.33 | 502.26 | 388.22 | 291.17 | 204.13 |
| # | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 |



**Figure 2.5:** Example of overlap between field-collected and laboratory-reared fungus garden samples for the glycoside hydrolase family 3 peptide N.AIADLLFGDVNPSGK.L. (a) Theoretical b and y ions, identified m/z values are highlighted in red. (b) Field sample MS/MS spectra; instrument, LTQ Orbitrap (high mass accuracy MS, low mass accuracy MS/MS); parent ion ppm error, 4.03; peptide charge state, +2; Retention Time, 42.72 minutes; XCorr, 4.19; MS-GF Spectra Probability, 1.42x10-13. (c) Laboratory-Kept sample MS/MS spectra; instrument, LTQ Orbitrap (high mass accuracy MS, low mass accuracy MS/MS); parent ion ppm error, 3.06; peptide charge state, +2; Retention Time, 33.36 minutes; XCorr, 4.08; MS-GF Spectra Probability, 1.67x10-12.

**Table 2.5.** A subset of bacterial proteins identified in leaf-cutter ant fungus gardens using liquid chromatography-tandem mass spectrometry. Proteins of interest are listed with their annotation, number of peptides matching, % coverage in the proteome, phyogenetic bin, and predicted KEGG pathway.

| Function | Unique | % Coverage | Phylogenetic Bin | KEGG Pathway Annotation |
|---|---|---|---|---|
| Maltooligosaccharide ABC transporter | 1 | 11.5 | *Enterobacter* | ABC transporters |
| ABC-type Fe3+-siderophore transporter | 1 | 15.0 | *Cronobacter* | ABC transporters |
| Molybdenum ABC transporter | 1 | 23.6 | *Enterobacter* | ABC transporters |
| ABC-type metal ion transport system | 1 | 17.5 | *Cronobacter* | ABC transporters |
| ABC-type sugar transport systems | 1 | 5.6 | *Enterobacter* | ABC transporters |
| Nickel ABC transporter | 1 | 13.7 | *Enterobacter* | ABC transporters |
| Nickel ABC transporter | 1 | 13.7 | *Enterobacter* | ABC transporters |
| Xylose-binding protein | 1 | 20.3 | *Enterobacter* | ABC transporters |
| Sulfate ABC transporter, permease protein | 1 | 11.3 | *Escherichia* | ABC transporters |
| ABC-type Fe2+-enterobactin transport system | 1 | 19.7 | *Citrobacter* | ABC transporters |
| Glutamate 5-kinase | 1 | 30.6 | *Citrobacter* | Arginine and proline met. |
| Arginine/lysine/ornithine decarboxylases | 1 | 18.3 | *Shigella* | Arginine and proline met. |
| Argininosuccinate lyase | 1 | 26.8 | *Escherichia* | Arginine and proline met. |
| Dihydrolipoamide dehydrogenase | 1 | 17.7 | *Escherichia* | Glycine, serine and threonine met. |
| Dihydrolipoamide dehydrogenase | 1 | 16.2 | *Pantoea* | Glycine, serine and threonine met. |
| Phosphoserine phosphatase | 1 | 9.9 | *Cronobacter* | Glycine, serine and threonine met. |
| Serine hydroxymethyltransferase | 2 | 26.8 | *Erwinia* | Glycine, serine and threonine met. |
| Serine hydroxymethyltransferase | 1 | 12.6 | *Citrobacter* | Glycine, serine and threonine met. |
| Serine hydroxymethyltransferase | 1 | 14.4 | *Pantoea* | Glycine, serine and threonine met. |
| Phosphoglycerate dehydrogenase | 1 | 5.3 | *Pantoea* | Glycine, serine and threonine met. |
| Phosphoglycerate dehydrogenase | 1 | 11.0 | *Pantoea* | Glycine, serine and threonine met. |
| N-formylglutamate amidohydrolase | 1 | 21.5 | *Pseudomonas* | Histidine met. |
| UDP-N-acetylmuramyl pentapeptide synthase | 1 | 9.9 | *Cronobacter* | Lysine biosyn.; Peptidoglycan biosyn. |
| Assimilatory nitrate reductase (NADH) beta subunit | 1 | 16.7 | *Enterobacter* | Nitrogen met. |
| Ketopantoate reductase | 1 | 11.5 | *Salmonella* | Pantothenate and CoA biosyn. |
| Dephospho-CoA kinase | 1 | 8.3 | *Cronobacter* | Pantothenate and CoA biosyn. |
| Dephospho-CoA kinase | 1 | 8.3 | *Cronobacter* | Pantothenate and CoA biosyn. |
| Ketol-acid reductoisomerase | 1 | 14.7 | *Cronobacter* | Pantothenate and CoA biosyn. |
| Ketol-acid reductoisomerase | 1 | 14.4 | *Enterobacter* | Pantothenate and CoA biosyn. |
| 2,3-dihydroxyphenylpropionate 1,2-dioxygenase | 1 | 16.3 | *Escherichia* | Phenylalanine met. |
| Phenylacetaldehyde dehydrogenase | 1 | 6.0 | *Enterobacter* | Phenylalanine met. |
| Phosphoribosylanthranilate isomerase | 1 | 23.5 | *Pseudomonas* | Phenylalanine, tyrosine and tryptophan met. |
| PTS system IIB component | 1 | 11.0 | *Enterobacter* | Phosphotransferase system |
| PTS system IIB component | 1 | 11.5 | *Enterobacter* | Phosphotransferase system |
| PTS system N-acetylglucosamine-specific | 1 | 8.4 | *Klebsiella* | Phosphotransferase system |
| Coproporphyrinogen III oxidase | 1 | 5.3 | *Klebsiella* | Porphyrin and chlorophyll met. |
| Alpha-1,4-glucan 6-glycosyltransferase | 1 | 4.7 | *Enterobacter* | Starch and sucrose met. |
| Glycosyl hydrolase family 3 | 1 | 11.5 | *Herbaspirillum* | Starch and sucrose met. |
| Glycosyl hydrolase family 4 | 1 | 8.1 | *Enterobacter* | Starch and sucrose met. |
| Sulfite reductase (NADPH) | 1 | 7.3 | *Cronobacter* | Sulfur met. |
| Cysteine synthase | 2 | 14.2 | *Enterobacter* | Sulfur met.; Cysteine and methionine met. |
| Cysteine synthase | 1 | 10.2 | *Cronobacter* | Sulfur met.; Cysteine and methionine met. |
| Cysteine synthase | 1 | 12.4 | *Enterobacter* | Sulfur met.; Cysteine and methionine met. |

In addition to bacteria, we found that fungus gardens contain substantial populations of bacteriophage (Figure 2.1). These organisms could be playing key roles by limiting bacterial abundance or decreasing ecosystem productivity. Moreover, because fungus gardens contain numerous closely-related genera in the *Enterobacteriaceae,* bacteriophage could provide a common mechanism for gene transfer between lineages. The presence of bacteriophage in fungus gardens adds to the number of organisms that are shaping these ecosystems and introduces a new layer of complexity into the ecology of fungus gardens.

Metagenomics and metaproteomics have previously been shown to be invaluable tools for analyzing microbial communities (39-46), including those associated with herbivores (31, 32, 45, 47). Here we use these techniques to provide insight into the fungus gardens of leaf-cutter ants. Our work establishes that relatively few genera dominate the bacterial fraction of these communities, and that the genus *Enterobacter* appears to be particularly prevalent. We show that bacteria have diverse metabolic potential associated with the degradation of plant biomass, and we confirm the production of two bacterial glycoside hydrolases *in situ*. Moreover, we show that bacteria in fungus gardens likely participate in the biosynthesis of amino acids, B-vitamins, and other nutrients that potentially enhance the growth or biomass processing efficiency of the fungal cultivar. This is consistent with a model of synergistic biomass degradation by a fungus-bacteria consortium. Our work enhances our knowledge of how leaf-cutter ants process massive quantities of plant biomass in their ancillary digestive systems, and underscores the importance of symbiotic communities on the evolution and ecology of herbivores.

**2.6 References**


1. **Hölldobler, B, and Wilson, EO.** 1990. The ants. Harvard University Press, Cambridge, MA.

2. **Hölldobler, B, and Wilson, EO.** 2010. The leafcutter ants: civilization by instinct. W. W. Norton and Company, Inc., New York, NY.

3. **Hölldobler, B, and Wilson, EO.** 2008. The superorganism: the beauty, elegance, and strangeness of insect societies. W. W. Norton & Company, New York, NY.

4. **Weber, NA.** 1966. Fungus-growing ants. *Science* **153:**587-604.

5. **Schultz, TR, and Brady, SG.** 2008. Major evolutionary transitions in ant agriculture. *Proc Natl Acad Sci U S A* **105:**5435-40.

6. **Chapela, IH, Rehner, SA, Schultz, TR, and Mueller, UG.** 1994. Evolutionary history of the symbiosis between fungus-growing ants and their fungi. *Science* **266:**1691-1694.

7. **Pinto-Tomas, AA, Anderson, MA, Suen, G, Stevenson, DM, Chu, FS, Cleland, WW, Weimer, PJ, and Currie, CR.** 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* **326:**1120-1123.

8. **De Fine Licht, HH, Schiott, M, Mueller, UG, and Boomsma, JJ.** 2010. Evolutionary transitions in enzyme activity of ant fungus gardens. *Evolution* **64:**2055-2069.

9. **Schiott, M, Rogowska-Wrzesinska, A, Roepstorff, P, and Boomsma, JJ.** 2010. Leaf-cutting ant fungi produce cell wall degrading pectinase complexes reminiscent of phytopathogenic fungi. *BMC Biol.* **8:**156.

10. **Semenova, TA, Hughes, DP, Boomsma, JJ, and Schiott, M.** 2011. Evolutionary patterns of proteinase activity in attine ant fungus gardens. *BMC Microbiol.* **11:**15.

11. **Suen, G, Scott, JJ, Aylward, FO, Adams, SM, Tringe, SG, Pinto-Tomas, AA, Foster, CE, Pauly, M, Weimer, PJ, Barry, KW, Goodwin, LA, Bouffard, P, Li, L, Osterberger, J, Harkins, TT, Slater, SC, Donohue, TJ, and Currie, CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6:**e1001129.

12. **Martin, MM, and Weber, NA.** 1969. The cellulose-utilizing capability of the fungus cultured by the attine ant *Atta colombica tonsipes*. *Ann. Entomol. Soc. Am.* **62:**1386-1387.

13. **Abril, AB, and Bucher, EH.** 2002. Evidence that the fungus cultured by leaf-cutting ants does not metabolize cellulose. *Ecol. Lett.* **5:**325-328.

14. **Schiott, M, De Fine Licht, HH, Lange, L, and Boomsma, JJ.** 2008. Towards a molecular understanding of symbiont function: identification of a fungal gene for the

degradation of xylan in the fungus gardens of leaf-cutting ants. *BMC Microbiol.* **8:**40.

15. **Gomes De Siqueira, C, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJA.** 1998. Metabolism of plant polysaccharides by *Leucoagaricus gongylophorus*, the symbiotic fungus of the leaf-cutting ant *Atta sexdens*. *Appl. Environ. Microbiol.* **64:**4820-4822.

16. **Silva, A, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJ.** 2006. Production of polysaccharidases in different carbon sources by *Leucoagaricus gongylophorus* Moller (Singer), the symbiotic fungus of the leaf-cutting ant *Atta sexdens* Linnaeus. *Curr. Microbiol.* **53:**68-71.

17. **Silva, A, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJ.** 2006. Starch metabolism in *Leucoagaricus gongylophorus*, the symbiotic fungus of leaf-cutting ants. *Microbiol. Res.* **161:**299-303.

18. **Scott, JJ, Budsberg, KJ, Suen, G, Wixon, DL, Balser, TC, and Currie, CR.** 2010. Microbial community structure of leaf-cutter ant fungus gardens and refuse dumps. *PLoS One* **5:**e9922.

19. **Bacci M, Ribeiro SB, Casarotto MEF, and Pagnocca FC.** 1995. Biopolymer-degrading bacteria from nests of the leaf-cutting ant *Atta sexdens Rubropilosa*. *Braz. J. Med. Biol. Res.* **28:**79-82.

20. **Santos, AV, Dillon, RJ, Dillon, VM, Reynolds, SE, and Samuels, RI.** 2004. Ocurrence of the antibiotic producing bacterium *Burkholderia* sp. in colonies of the leaf-cutting ant *Atta sexdens rubropilosa*. *FEMS Microbiol. Lett.* **239:**319-323.

21. **Haeder, S, Wirth, R, Herz, H, and Spiteller, D.** 2009. Candicidin-producing *Streptomyces* support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*. *Proc. Natl. Acad. Sci. USA* **106:**4742-4746.

22. **Mueller, U, Gerardo, N, Aanen, D, Six, D, and Schultz, T.** 2005. The evolution of agriculture in insects. *Annu. Rev. Ecol. Evol. Syst.* **36:**563-595.

23. **Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, YJ, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, ML, Jarvie, TP, Jirage, KB, Kim, JB, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, McDade, KE, McKenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth, GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A, Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF, and Rothberg, JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-380.

24. **Eng, J, McCormack, A, and Yates, J.** 1994. An approach to correlate tandem mass

spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5:**976-989.

25. **Kim, S, Gupta, N, and Pevzner, PA.** 2008. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **7:**3354-63.

26. **Angly, FE, Willner, D, Prieto-Davo, A, Edwards, RA, Schmieder, R, Vega-Thurber, R, Antonopoulos, DA, Barott, K, Cottrell, MT, Desnues, C, Dinsdale, EA, Furlan, M, Haynes, M, Henn, MR, Hu, Y, Kirchman, DL, McDole, T, McPherson, JD, Meyer, F, Miller, RM, Mundt, E, Naviaux, RK, Rodriguez-Mueller, B, Stevens, R, Wegley, L, Zhang, L, Zhu, B, and Rohwer, F.** 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput. Biol.* **5:**e1000593.

27. **Kanehisa, M, Araki, M, Goto, S, Hattori, M, Hirakawa, M, Itoh, M, Katayama, T, Kawashima, S, Okuda, S, Tokimatsu, T, and Yamanishi, Y.** 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36:**D480-4.

28. **Tatusov, RL, Natale, DA, Garkavtsev, IV, Tatusova, TA, Shankavaram, UT, Rao, BS, Kiryutin, B, Galperin, MY, Fedorova, ND, and Koonin, EV.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:**22-8.

29. **Markowitz, VM, Ivanova, NN, Szeto, E, Palaniappan, K, Chu, K, Dalevi, D, Chen, IM, Grechkin, Y, Dubchak, I, Anderson, I, Lykidis, A, Mavromatis, K, Hugenholtz, P, and Kyrpides, NC.** 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36:**D534-8.

30. **Cantarel, BL, Coutinho, PM, Rancurel, C, Bernard, T, Lombard, V, and Henrissat, B.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37:**D233-8.

31. **Pope, PB, Denman, SE, Jones, M, Tringe, SG, Barry, K, Malfatti, SA, McHardy, AC, Cheng, JF, Hugenholtz, P, McSweeney, CS, and Morrison, M.** 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc. Natl. Acad. Sci. USA* **107:**14793-8.

32. **Warnecke, F, Luginbuhl, P, Ivanova, N, Ghassemian, M, Richardson, TH, Stege, JT, Cayouette, M, McHardy, AC, Djordjevic, G, Aboushadi, N, Sorek, R, Tringe, SG, Podar, M, Martin, HG, Kunin, V, Dalevi, D, Madejska, J, Kirton, E, Platt, D, Szeto, E, Salamov, A, Barry, K, Mikhailova, N, Kyrpides, NC, Matson, EG, Ottesen, EA, Zhang, X, Hernandez, M, Murillo, C, Acosta, LG, Rigoutsos, I, Tamayo, G, Green, BD, Chang, C, Rubin, EM, Mathur, EJ, Robertson, DE, Hugenholtz, P, and Leadbetter, JR.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:**560-565.

33. **Costa, AN, Vasconcelos, HL, Vieira-Neto, EHM, and Bruna, EM.** 2009. Do herbivores exert top-down effects in Neotropical savannas? Estimates of biomass consumption by leaf-cutter ants. *J. Veg. Sci.* **19:**849-854.

34. **Belt, T.** 1874. The Naturalist in Nicaragua. E. Bumpus, London.

35. **Currie, CR, and Stuart, AE.** 2001. Weeding and grooming of pathogens in agriculture by ants. *P. Roy. Soc. B-Biol. Sci.* **268:**1033-1039.

36. **Currie, CR, Scott, JA, Summerbell, RC, and Malloch, D.** 1999. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398:**701-704.

37. **Fernández-Marín, H, Zimmerman, JK, Rehner, SA, and Wcislo, WT.** 2006. Active use of the metaplural glands by ants in controlling fungal infection. *P. Roy. Soc. B-Biol. Sci.* **273:**1689-1695.

38. **Powell, RJ, and Stradling, DJ.** 1986. Factors influencing the growth of *Attamyces bromatificus*, a symbiont of the attine ants. *Trans. Br. Mycol. Soc.* **87:**205-213.

39. **Verberkmoes, NC, Russell, AL, Shah, M, Godzik, A, Rosenquist, M, Halfvarson, J, Lefsrud, MG, Apajalahti, J, Tysk, C, Hettich, RL, and Jansson, JK.** 2009. Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* **3:**179-89.

40. **Woyke, T, Teeling, H, Ivanova, NN, Huntemann, M, Richter, M, Gloeckner, FO, Boffelli, D, Anderson, IJ, Barry, KW, Shapiro, HJ, Szeto, E, Kyrpides, NC, Mussmann, M, Amann, R, Bergin, C, Ruehland, C, Rubin, EM, and Dubilier, N.** 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443:**950-955.

41. **Gill, SR, Pop, M, Deboy, RT, Eckburg, PB, Turnbaugh, PJ, Samuel, BS, Gordon, JI, Relman, DA, Fraser-Liggett, CM, and Nelson, KE.** 2006. Metagenomic analysis of the human distal gut microbiome. *Science* **312:**1355-1359.

42. **Kalyuzhnaya, MG, Lapidus, A, Ivanova, N, Copeland, AC, McHardy, AC, Szeto, E, Salamov, A, Grigoriev, IV, Suciu, D, Levine, SR, Markowitz, VM, Rigoutsos, I, Tringe, SG, Bruce, DC, Richardson, PM, Lidstrom, ME, and Chistoserdova, L.** 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. *Nature Biotech.* **26:**1029-34.

43. **Allgaier, M, Reddy, A, Park, JI, Ivanova, N, D'Haeseleer, P, Lowry, S, Sapra, R, Hazen, TC, Simmons, BA, VanderGheynst, JS, and Hugenholtz, P.** 2009. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* **5:**e8812.

44. **Ram, RJ, Verberkmoes, NC, Thelen, MP, Tyson, GW, Baker, BJ, Blake, RC, 2nd, Shah, M, Hettich, RL, and Banfield, JF.** 2005. Community proteomics of a natural microbial biofilm. *Science* **308:**1915-20.

45. **Burnum, KE, Callister, SJ, Nicora, CD, Purvine, SO, Hugenholtz, P, Warnecke, F, Scheffrahn, RH, Smith, RD, and Lipton, MS.** 2011. Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *ISME J.* **5:**161-4.

46. **Wilmes, P, Wexler, M, and Bond, PL.** 2008. Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS One* **3:**e1778.

47. **Brulc, JM, Antonopoulos, DA, Miller, ME, Wilson, MK, Yannarell, AC, Dinsdale, EA, Edwards, RE, Frank, ED, Emerson, JB, Wacklin, P, Coutinho, PM, Henrissat, B, Nelson, KE, and White, BA.** 2009. Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proc. Natl. Acad. Sci. USA* **106:**1948-1953.

# Chapter 3: *Leucoagaricus gongylophorus* Produces Diverse Enzymes for the Degradation of Recalcitrant Plant Polymers in Leaf-Cutter Ant Fungus Gardens

**Previously Published**: Frank O Aylward, Kristin E Burnum-Johnson, Susannah G Tringe, Clotilde Teiling, Daniel M Tremmel, Joseph A Moeller, Jarrod J Scott, Kerrie W Barry, Paul D Piehowski, Carrie D Nicora, Stephanie A Malfatti, Matthew E Monroe, Samuel O Purvine, Lynne A Goodwin, Richard D Smith, George M Weinstock, Nicole M Gerardo, Garret Suen, Mary S Lipton, Cameron R Currie (2013). *Leucoagaricus gongylophorus* produces diverse enzymes for the degradation of recalcitrant plant polymers in leaf-cutter ant fungus gardens. *Applied and Environmental Microbiology*, 79 (12), 3770-3778.

**Author Contributions**: For this chapter I performed all computational experiments, and I conceived of all analyses with Cameron R. Currie, Kristin Burnum, and Mary Lipton. Proteomic analyses were conducted at Pacific Northwest National Laboratories, the fungal genome was sequenced by Roche Biosciences, and metagenomes were generated by the Joint Genome Institute. Other authors provided helpful advice for analyses, and helped with writing.

## 3.1 Abstract

Plants represent a large reservoir of organic carbon comprised primarily of recalcitrant polymers that most metazoans are unable to deconstruct. Many herbivores gain access to nutrients in this material indirectly by associating with microbial symbionts, and leaf-cutter ants are a paradigmatic example. These ants use fresh foliar biomass as manure to cultivate gardens composed primarily of *Leucoagaricus gongylophorus*, a basidiomycetous fungus that produces specialized hyphal swellings that serve as a food source for the host ant colony. Although leaf-cutter ants are conspicuous herbivores that contribute substantially to carbon turnover in Neotropical ecosystems, the process through which plant biomass is degraded in their fungus gardens is not well understood. Here we present the first draft genome of *L. gongylophorus,* and using genomic and metaproteomic tools we investigate its role in lignocellulose degradation in the gardens of both *Atta cephalotes* and *Acromyrmex echinatior* leaf-cutter ants. We show that *L. gongylophorus* produces a diversity of lignocellulases in ant gardens, and is likely the primary driver of plant biomass degradation in these ecosystems. We also show that this fungus produces

distinct sets of lignocellulases throughout the different stages of biomass degradation, including numerous cellulases and laccases that likely play an important role in lignocellulose degradation. Our study provides a detailed analysis of plant biomass degradation in leaf-cutter ant fungus gardens as well as insight into the enzymes underlying the symbiosis between these dominant herbivores and their obligate fungal cultivar.

**3.2 Introduction**

The ecology and evolution of metazoans is shaped, at least in part, by microbial symbionts (1, 2). The formation of symbiotic associations with beneficial microbes that confer novel physiological capacities onto their hosts has been described as a form of evolutionary innovation, and even allows some animals to occupy ecological niches that would otherwise be unavailable (1-4). The ability of animals to gain access to nutrients in plant biomass, an abundant energy source in terrestrial ecosystems, is facilitated by the formation of symbioses with microbes, as plant cell walls are largely composed of recalcitrant polymers that most animals are unable to deconstruct (5, 6). In herbivorous mammals, especially ruminants, the role of microbes in mediating plant biomass degradation has been described (6, 7). By contrast, detailed studies of lignocellulolytic microbes associated with insect herbivores, the most species-diverse and dominant plant-feeding animals in most ecosystems, are limited.

Leaf-cutter ants are hallmark examples of insect herbivores that gain access to nutrients in plant material through symbioses with microbes (Figure 3.1a,b). Through the cultivation of fungus-bacteria "gardens" on fresh foliar material, these ants are able to access nutrients in plant biomass that would otherwise be unavailable (8). Despite the central importance of plant biomass degradation to the ecology and evolution of leaf-cutter ants, this process is poorly understood. It has long been assumed that the dominant basidiomycetous cultivar in these gardens,

*Leucoagaricus gongylophorus,* can readily degrade plant biomass (9). However, reports that *L. gongylophorus* can grow on cellulose in pure culture have varied (10, 11), and the full extent of this organism's lignocellulolytic capabilities has not been determined. Culture-based studies have shown that a diversity of bacteria, yeasts, and microfungi can also be found in ant gardens (12-14), and recent culture-independent investigations have revealed that a distinct community of bacteria resides in the fungus gardens of leaf-cutter ants (15, 16). Although the genetic potential of bacteria in *Atta* gardens has been shown to be consistent with the degradation of less-recalcitrant plant material (17), the relative contribution of these microbes to biomass degradation remains unknown.

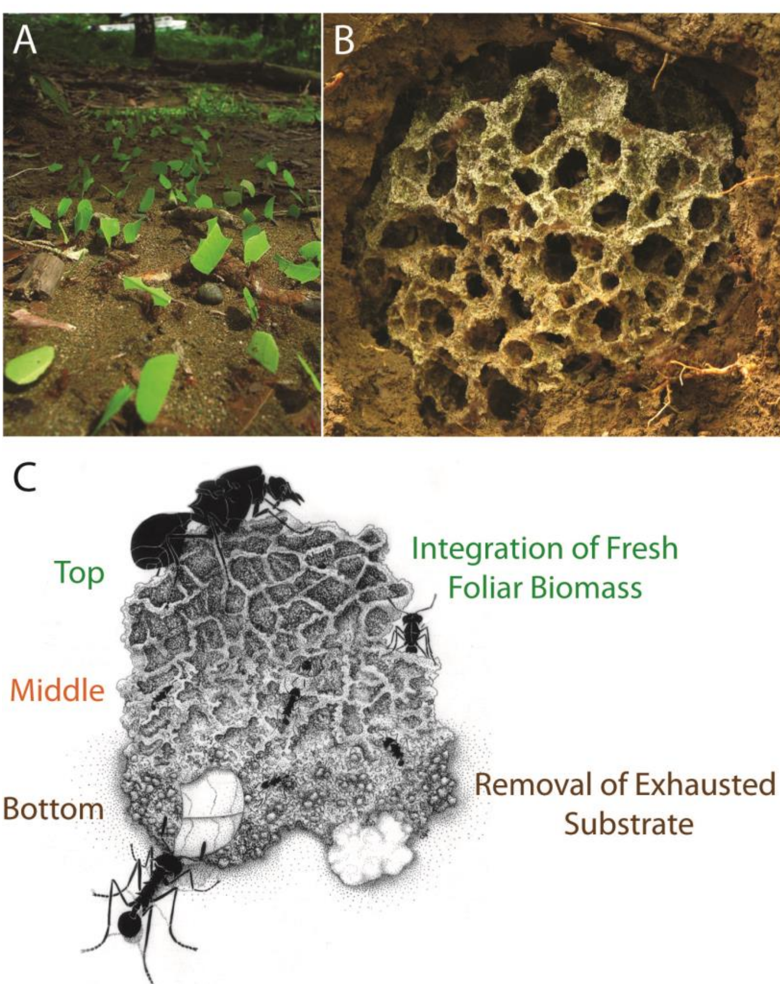



**Figure 3.1.** Leaf-cutter ants forage on fresh foliar biomass (A) and use it as manure to cultivate symbiotic microbial gardens (B) they consume for food. Fresh biomass is progressively degraded after it is integrated into the top strata of leaf-cutter ant gardens, creating a vertical gradient of biomass degradation (C) (Photo credits: A) Used under the GNU Free Documentation License, Version 1.2; B) Jarrod J. Scott; C) Original line drawing by Cara Gibson).

The symbiotic gardens of leaf-cutter ants are meticulously tended by their hosts to maintain a stable microbial assemblage (18, 19). Fresh biomass is integrated into the top strata of fungus gardens and progressively degraded as it moves into lower strata. After 4-6 weeks, the ants remove spent fungus garden material from the bottom strata of gardens. Thus, gardens represent a gradient of biomass degradation whereby fresh biomass, composed of diverse polymers, plant toxins, and other compounds, is degraded in top strata while partially-degraded plant biomass, from which many usable nutrients have been exhausted, comprises the bottom strata (Figure 3.1c) (20). These gardens are continuously tended by the ants to facilitate biomass degradation, the removal of pests, and the prevention of infection from fungal pathogens (19). Although the two genera of leaf-cutter ants, *Atta* and *Acromyrmex*, cultivate the same phylogenetic group of *L. gongylophorus* as the primary cultivar in their gardens (21), these ant genera have distinct life histories. While species of *Acromyrmex* maintain colonies containing ~100,000 ants, *Atta* species can achieve colonies with millions of workers and hundreds of fungus garden chambers (8). Both *Atta* and *Acromyrmex* are broadly distributed throughout Neotropical ecosystems, with species of *Atta* often dominating herbivorous niches (22).

In this study we use genomic and metaproteomic tools to investigate plant biomass degradation in the fungus gardens of both *Atta cephalotes* and *Acromyrmex echinatior* leaf-cutter ants. We sequence the first draft genome of a *L. gongylophorus* isolate and provide an inventory of its lignocellulolytic potential. We then classify the lignocellulases produced in fungus gardens throughout the different stages of biomass degradation *in situ* by performing metaproteomic analyses on different strata of *At. cephalotes* and *Ac. echinatior* gardens. Lastly, to investigate the degree of conservation between lignocellulases encoded in different *L. gongylophorus*

strains, we compare the sequences of 4 glycoside hydrolases sequenced from isolates associated with different ant hosts and geographic areas.

## 3.3 Materials and Methods

### 3.3.1 Fungal Genome Sequencing and Annotation

The *Leucoagaricus gongylophorus* Ac12 strain used for genome sequencing was isolated from a nest of *At. cephalotes* in Gamboa, Panama in July 2010. Cultures of *L. gongylophorus* were maintained in the dark at 30 ºC on plates of potato dextrose agar (BD Difco, Franklin Lakes, NJ). DNA was purified using a modified CTAB extraction. Briefly, mycelia were scraped directly from plates into 2 mL tubes containing 0.25 mg of silica beads and 1X PBS and shaken for 2 min in a bead-beater. Lysed fungal cells were then combined with CTAB buffer and DNA was extracted as previously described (23). DNA was sequenced using one plate each of shotgun and 8 kb paired-end Roche 454 Titanium pyrosequencing (24), and subsequently assembled using Newbler v. 2.1 (Details in Table 3.1).

Proteins were predicted from the assembly using GeneMark-ES (25) and Augustus (26) (details in Appendix 2). Each protein prediction dataset was annotated independently, and proteins predicted by both tools were identified using BLASTP (27). Only proteins identified in our metaproteomic analyses (at least one peptide matching with < 10 ppm error, see below for details) were included in the final protein prediction dataset. Eukaryotic Orthologous Group (KOG) (28) and Pfam (29) annotations were constructed using RPS-BLAST (30) (E-value: 1e-5), Carbohydrate Active Enzymes (CAZymes) (31) were identified using methods previously described (16), Fungal Oxidative Lignin-degrading enzymes (FOLymes) (32) were identified by comparing all predictions to proteins on the FOLy database (downloaded: 11/01/2012) using BLASTP (E-value: 1e-10), and proteases were predicted using the MEROPs database (33). The

annotations for all predicted CAZymes, FOLymes, and proteases were inspected manually. Signal peptide and protein localization predictions were generated for all fungal proteins using WolfPSort (34) and SignalP (35) (both the hidden Markov model and neural network implementation). All FOLymes and proteases not predicted to be secreted were excluded from subsequent analyses. A list of the most abundant *L. gongylophorus* lignocellulases identified in the metaproteomic data can be found in Table 3.2.

### 3.3.2 Metaproteomic Sample Preparation, Processing, and Analysis

Metaproteomic analysis was conducted on garden material collected from one *At. cephalotes* and one *Ac. echinatior* leaf-cutter ant colonies reared in the laboratory on diets of Oak (*Quercus*) and Maple (*Acer*). Three samples were collected across the vertical gradient of biomass degradation in each fungus garden (top, middle, and bottom strata). Details for proteomic sample preparation can be found in the Appendix 2. Briefly, 150 mg of each sample were added to an ice-cold mortar with liquid nitrogen and ground for ~2 minutes with a pestle. Then 2 mL of water was added to the sample with continuous grinding until the mixture was thawed. The samples were pipetted into 0.6 mL centrifuge tubes that contained 0.1 mm zirconia/silica beads and bead-beat for 3 minutes. The tubes were centrifuged at 4 °C at 10,000g for 5 minutes and the supernatant removed. The collected supernatant was then centrifuged at 4000g for 5 minutes to remove debris. Two proteomic samples were prepared from each biological sample, one using a Rapigest SF surfactant (Waters, Milford, MA) protocol and the other using a Filter Aided Sample Preparation (FASP) (36) protocol, and resulting peptides were analyzed by liquid chromatography high mass accuracy tandem mass spectrometry (LC-MS/MS) (LTQ Orbitrap Velos & LTQ Orbitrap, Thermo Fisher Corporation, San Jose, CA). The peptide tandem mass spectra resulting from these procedures were then mapped onto appropriate protein

prediction datasets. An outline of the metaproteomic workflow is provided in Appendix 2. Peptide matches were filtered using Sequest scores (37), MS-GF spectral probabilities (38), and false discovery rates, as previously described (17). Moreover, all peptides with > 10 ppm mass error were discarded. Proteins with only a single unique peptide matching were retained, but those mapping to lignocellulases were inspected manually.

### 3.3.3 Protein Prediction Databases Used for Spectral Mapping

Both the Augustus and GeneMark-ES protein predictions of the draft *L. gongylophorus* genome were used independently for spectral mapping. To compare fungal proteins to those of bacteria in fungus gardens, 8 databases corresponding to bacterial proteins were also used. Four of these databases correspond to bacterial metagenomes constructed from *At. colombica, At. cephalotes*, and *Ac. echinatior* fungus gardens, while 6 correspond to the genomes of bacteria isolated from fungus gardens (*Enterobacter* FGI 35, *Pseudomonas* FGI 182, *Serratia* FGI 94 (39), *Enterobacteriaceae* strain FGI 157 (40), *Klebsiella variicola* AT-22, and *Pantoea* at-9b). Proteins were predicted using IMG-ER (41) and Prodigal (42) for the metagenomic and genomic datasets, respectively, and CAZymes were predicted from these datasets in a manner identical to that of the *L. gongylophorus* protein predictions.

### 3.3.4 Statistical Analyses of Spectral Profiles for *L. gongylophorus* Lignocellulases

The number of mass spectra mapped to *L. gongylophorus* lignocellulases was compared between samples to identify enzymes enriched in a given stratum. The number of spectra mapping to both GeneMark-ES and Augustus proteins were combined, if predicted by both programs, and final spectral profiles were clustered using Pearson's r values calculated in R (http://www.R-project.org). In the case of individual lignocellulases (see Appendix 2), only those proteins having at least 10 spectra mapping were considered in order to reduce the chances of

spurious clustering due to low spectral coverage. In the case of lignocellulase families, all spectra mapping to proteins of a given family were combined, and all CAZy, FOLy, and MEROPs protein families are shown (Figure 3.2). Fisher's Exact Test was performed on the raw number of spectra mapping to enzymes in different samples to identify lignocellulases enriched in one sample compared to another. Comparisons were performed only between samples corresponding to the same strata in *Ac. echinatior* and *At. cephalotes* gardens.

### 3.3.5 PCR Amplification, Sequencing, and Phylogenetic Analysis of *L. gongylophorus* CAZymes

All *L. gongylophorus* strains used for phylogenetic analysis were obtained from leaf-cutter ant colonies collected from different locations in South and Central America and maintained in the Currie laboratory (details in Table 3.3). Cultures were isolated on plates of potato dextrose agar (BD Difco, Franklin Lakes, NJ) and maintained at 30 ºC in the dark. DNA was extracted using a modified phenol/chloroform extraction. Primers used for the amplification of genes encoding one GH6 and two GH10 glycoside hydrolases were designed in this study from contigs present in previously-reported community metagenomes of ant gardens (details in Appendix 2), which contained a small amount of *L. gongylophorus* sequences in addition to the bacterial component (17). Amplification of partial-length GH7-encoding genes was performed using the primers fungcbhIF and fungcbhIR designed in a previous study (43). All sequences were amplified using PCR on a MJ Research PRD-200 Peltier Thermal Cycler with a 2 min 94 ºC denaturing step followed by 29 cycles of a 1 min 94 ºC denaturing step, a 2 min annealing step (see individual temperatures used below), and a 2 min 72 ºC elongation step, and finally a 6 min 72 ºC elongation step. The annealing temperatures used for the GH7, GH6, and two GH10 amplicons (termed CBH1, CBHII, xynI, and xynII, respectively) were 50 ºC, 44 ºC, 52 ºC, and

45 ºC, respectively. Amplicons were run on 1% agarose gels containing EtBr for visualization and subsequently sequenced using an ABI 3730xl DNA Analyzer. The resulting chromatograms were visualized using Sequencher v 4.5 (Gene Codes Inc., Ann Arbor, MI, USA). Positions for which the base calls were ambiguous due to conflicting chromatogram signals were designated "N".

Amino acid sequences for the nucleotide sequences of these CAZyme genes was obtained using a combination of 6-frame translation, FGENESH (http://linux1.softberry.com/ all.htm), and BLASTX (27) against the NCBI NR database (44). Nucleotide and amino acid sequences were compared among amplicons using standalone BLASTN and BLASTX (results in Table 3.4). A selection of top BLASTP hits was compiled for each CAZyme and used for subsequent phylogenetic analyses. MUSCLE (45) was used for amino acid alignments, FastTree (46) was used to infer maximum-likelihood phylogenetic trees with corresponding SH-like local support values (47). A small number of sequences for which regions were trimmed due to ambiguous base-calls in the chromatograms were excluded in the phylogenetic analyses, although they were used for the overall comparison of nucleotide and amino acid identity. Trees were visualized using the Interactive Tree of Life (ITOL, (48)).

## 3.4 Results

### 3.4.1 *L. gongylophorus* Draft Genome Assembly and Annotation

Assembly of the reads from two full plates of Roche 454 Titanium pyrosequencing of purified *L. gongylophorus* DNA yielded 92,785 contigs comprising 101 Mbp of total sequence (Table 3.1). The largest contig in the assembly was 100,988 bp, and 2,368 contigs were larger than 5 kb. Analysis of the GC content of all contigs revealed a bimodal distribution with peaks at 29% and 47%, with primarily mitochondrial sequences comprising the former and chromosomal

sequences the latter (see Appendix 2). Gene prediction using a combination of Augustus and GeneMark-ES yielded 12,132 non-redundant predictions, and the majority of these had homology to proteins in other sequence databases (details in the Appendix 2). Individual spectral mapping of metaproteomic data to the raw Augustus and GeneMark-ES protein prediction datasets confirmed 4,567 protein predictions. Upon annotation of these confirmed proteins using the CAZy, FOLy, and MEROPs databases we identified 145 predicted biomass-degrading enzymes, including 81 glycoside hydrolases (GH), 6 polysaccharide lyases (PL), 9 carbohydrate esterases (CE), 9 laccases, 5 glyoxal oxidases, 4 aryl-alcohol oxidases, and 26 secreted proteases (details in Appendix 2). Additionally, we amplified and sequenced a single GH7 cellulase from *L. gongylophorus* that was not identified in the draft genome.

**Table 3.1**. *L. gongylophorus* draft genome sequencing and annotation statistics.

| | |
|---|---|
| Number of contigs in the *L. gongylophorus* assembly | 92,785 |
| Total bp in assembly | 101,584,4 |
| N50 contig size | 1,793 |
| Predicted proteins | 12,132 |
| Proteins verified by spectral mapping | 4,567 |
| % with KOG Annotations | 71.1 |
| % with Pfam Annotations | 71.4 |
| Total mass spectra mapped onto proteins | 484,059 |

### 3.4.2 Metaproteomics

Our metaproteomic analysis of six samples comprising the top, middle, and bottom strata of *Ac. echinatior* and *At. cephalotes* fungus gardens identified 505,566 spectra that could be confidently mapped to amino acid sequences in at least one of our protein prediction databases (Table 3.1, Appendix 2). Of these, 484,059 could be mapped to proteins belonging to *L. gongylophorus*, and 4,309 could be mapped to proteins in one of our bacterial databases. Of the 145 lignocellulases identified in the *L. gongylophorus* genome, 137 were represented in the

metaproteome of at least one *Ac. echinatior* sample, and 138 in at least one *At. cephalotes* sample (details in Appendix 2). In total, 44,347 spectra were mapped to all *L. gongylophorus* lignocellulases in the *Ac. echinatior* samples, while 27,878 were mapped to these same enzymes in the *At. cephalotes* samples. Inspection of CAZymes present in the bacterial component of the metaproteomic data identified only a single mass spectrum mapping to a GH8.

Clustering analysis of spectral profiles mapping to the *L. gongylophorus* lignocellulase protein families yielded distinct patterns throughout both the *Ac. echinatior* and *At. cephalotes* samples (Figure 3.2). When the spectral profiles of individual enzymes were analyzed, separate sets of lignocellulases were identified as abundant in the top, middle, and bottom strata in both ant species (see Appendix 2). Over 46% of the lignocellulases identified in the metaproteome were over-represented in at least one *Ac. echinatior* stratum when compared to the corresponding stratum in *At. cephalotes*, while the reverse was found for 30% of the enzymes (Fisher's Exact Test, $P < 0.05$). Comparison of the spectral profiles recovered for all *L. gongylophorus* proteins in the six samples resulted in separate clustering of the *At. cephalotes* and *Ac. echinatior* samples (Figure 3.3).

The most abundant *L. gongylophorus* biomass-degrading enzymes in the six metaproteomic datasets comprised 15 CAZymes, 4 FOLymes, and 13 proteases (Table 3.2). Specifically, these enzymes comprised 10 GHs, 3 PLs, 2 CEs, and 2 laccases. Of the abundant CAZymes identified, the majority belong to families predicted to hydrolyze hemicelluloses, pectins, and starch. Inspection of the spectral profiles of the most abundant lignocellulases revealed many to be most abundant in a particular stratum of fungus garden (Figure 3.4). Of the two most abundant laccases identified, one (LAG_2404) was found to be most abundant in the top strata of both *Ac. echinatior* and *At. cephalotes* gardens, while a nearly opposite trend was

**Table 3.2.** Fungal lignocellulases with high spectral abundance in the metaproteomic datasets. The spectral abundances shown represent the sum recovered from the top, middle, and bottom strata for both the *Ac. echinatior* and *At. cephalotes* samples.

| LAG ID | Protein Family | Annotation | Spectra Mapped | |
|---|---|---|---|---|
| **CAZymes** | | | *Ac. echinatior* | *At. cephalotes* |
| LAG_992 | GH15 | Glycoamylase, glycodextranase | 3,812 | 1,428 |
| LAG_4755 | PL1 | Pectin/Pectate lyase | 3,425 | 1,121 |
| LAG_3581 | CE5 | Acetyl-xylan esterase, cutinase | 305 | 1,440 |
| LAG_1450 | CE8 | Pectin Methylesterase | 998 | 543 |
| LAG_543 | GH28 | Polygalacturonase | 1,050 | 387 |
| LAG_3369 | PL4 | Rhamnogalacturonan lyase | 889 | 462 |
| LAG_3001 | GH35 | β-galactosidase | 495 | 557 |
| LAG_81 | PL4 | Rhamnogalacturonan lyase | 116 | 791 |
| LAG_420 | GH18 | Chitinase, acetylglucosaminidase | 186 | 688 |
| LAG_2062 | GH3 | Glucosidase, xylosidase | 325 | 252 |
| LAG_1651 | GH78 | Rhamnosidase | 163 | 403 |
| LAG_2564 | GH3 | Glucosidase, xylosidase | 116 | 445 |
| LAG_2638 | GH13, CBM20 | Amylase, pollulanase, glucosidase | 449 | 58 |
| LAG_5098 | GH3 | Glucosidase, xylosidase | 335 | 131 |
| LAG_4224 | GH10 | Xylanase | 249 | 146 |
| **FOLymes** | | | | |
| LAG_2404 | LO1 | Laccase | 3,483 | 1,947 |
| LAG_2639 | LO1 | Laccase | 2,562 | 2,013 |
| LAG_2522 | LDA3 | Glyoxal oxidase | 1,356 | 641 |
| LAG_5549 | LO1 | Laccase | 264 | 373 |
| **Proteases** | | | | |
| LAG_2622 | A01A | Aspartyl protease | 7,717 | 3,371 |
| LAG_3716 | M36 | Metalloprotease | 2,861 | 1,564 |
| LAG_2011 | S09X | Serine protease | 1,998 | 1,262 |
| LAG_2465 | S53 | Serine protease | 1,578 | 712 |
| LAG_3735 | S08A | Serine protease | 1,599 | 337 |
| LAG_981 | S10 | Serine protease | 1,038 | 310 |
| LAG_5096 | S08A | Serine protease | 471 | 783 |
| LAG_439 | A01A | Aspartyl protease | 433 | 693 |
| LAG_7402 | A01A | Aspartyl protease | 379 | 220 |
| LAG_3725 | M28E | Metalloprotease | 388 | 120 |
| LAG_2527 | S53 | Serine protease | 307 | 122 |
| LAG_3512 | S08A | Serine protease | 145 | 247 |
| LAG_1757 | S10 | Serine protease | 243 | 91 |

identified for the other (LAG_2639). Of the three cellulases identified, the cellobiohydrolases (GH6 and GH7) were found to be most abundant in the bottom strata of both ant gardens, while no trend could be found for the endoglucanase (GH9). Of the most abundant hemicellulases and

pectinases, a glucoamylase (GH15) and xylan esterase/cutinase (CE5) were most abundant in the top strata, while a pectin methylesterase (CE8) displayed an opposite trend. No trend consistent between *Ac. echinatior* and *At. cephalotes* samples could be identified for the other abundant lignocellulases.

### 3.4.3 Analysis of Fungal Lignocellulases from Different Isolates

To investigate the similarity of *L. gongylophorus* lignocellulases isolated from different ant hosts and geographic regions, we sequenced and compared the partial-length nucleotide sequences of four GHs from 11 fungal strains isolated from *Atta* and *Acromyrmex* colonies collected in Panama, Costa Rica, Peru, and Argentina (Table 3.3). Comparison of the nucleotide sequences revealed that each gene was on average 98.8-99.5% identical to corresponding genes in other isolates (Table 3.4). Moreover, the corresponding amino acid sequences for the genes were on average 98.9-100% identical between isolates. Phylogenetic analysis of these amino acid sequences revealed homology to enzymes encoded by numerous other saprotrophic fungi (see Appendix 2).

## 3.5 Discussion

In Neotropical ecosystems, leaf-cutter ants and their symbiotic fungus play important roles in nutrient cycling and plant biomass turnover (8). Despite their importance to leaf-cutter ants, it has remained unknown how the fungus gardens cultivated by these insects convert fresh foliar biomass into nutrients for their hosts. Here, using genomic and metaproteomic tools, we provide detailed insight into the lignocellulases used by the fungal cultivar *L. gongylophorus* to degrade plant biomass in the symbiotic gardens of leaf-cutter ants.

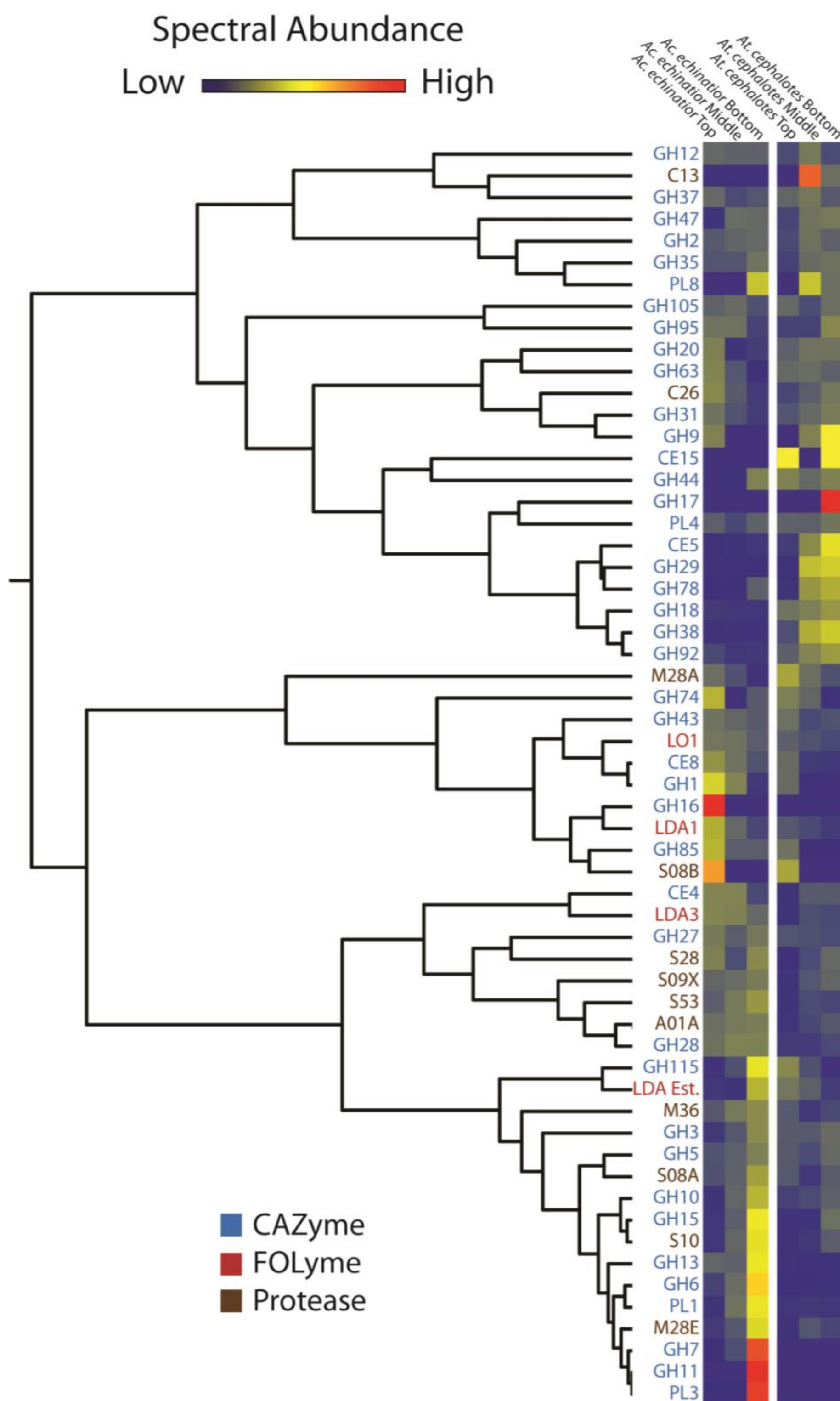



**Figure 3.2.** Heatmap representing the relative number of mass spectra matching to *L. gongylophorus* CAZyme (blue), FOLyme (magenta), and MEROPs (brown) protein families in the top, middle, and bottom strata of *Ac. echinatior* and *At. cephalotes* gardens. Rows have been normalized to unity. The dendrogram represents clustering based on the Pearson correlation of the spectral profiles for each protein family.

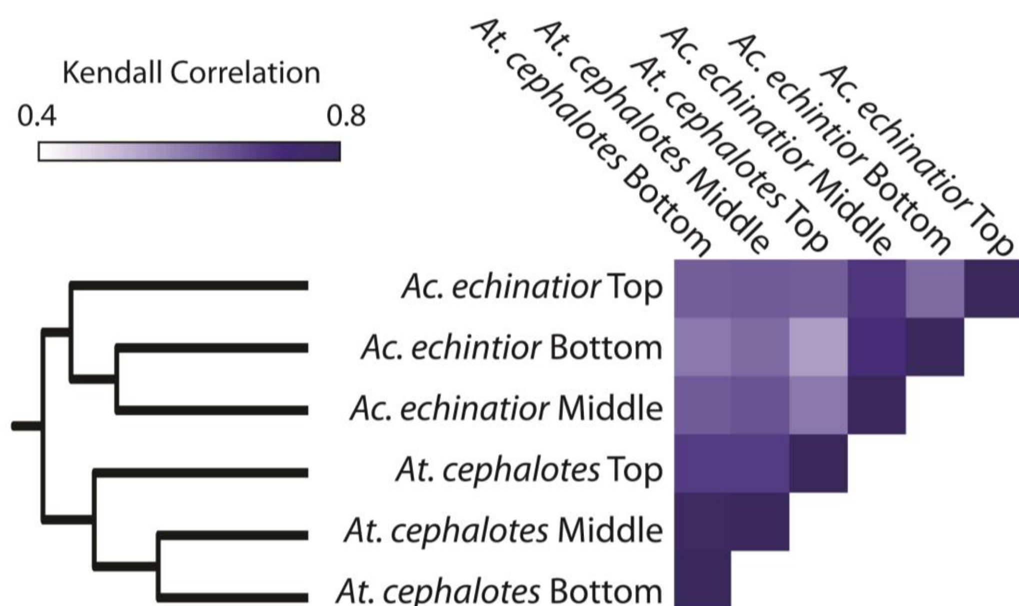



**Figure 3.3** Comparison of the spectral profiles of six metaproteomic datasets recovered from mapping all mass spectra against the *L. gongylophorus* protein predictions.

The combination of our genomic and metaproteomic analyses supports the hypothesis that *L. gongylophorus* is the dominant driver of biomass degradation in leaf-cutter ant gardens. In our analysis of the draft genome of this fungus as well as our metaproteomic experiments we identified a total of 145 lignocellulases, including numerous pectinases, xylanases, amylases, and cellulases (Figures 3.2, Table 3.2, and Appendix 2). Future transcriptomic analyses and additional genome sequencing will likely identify more enzymes used by this fungus for biomass degradation in leaf-cutter ant fungus gardens.

Previous metagenomic characterizations of bacterial communities in *Atta* and *Acromyrmex* gardens recovered mainly oligosaccharide-degrading enzymes and few CAZy

**Table 3.3.** Summary of *Leucoagaricus gongylophorus* isolates, their country of origin, host ant species, and genes that could be successfully amplified and sequenced from their purified DNA. All genes could be amplified, but only those genes marked with an X were successfully sequenced.

| Isolate ID | Country of Origin | Host Ant Species | CBHI | CBHII | XynI | XynII |
|---|---|---|---|---|---|---|
| LG_Peru1 | Peru | *Acromyrmex* sp. | X | X | X | X |
| LG_CR1 | Costa Rica | *Atta cephalotes* | X | X | X | X |
| LG_PN1 | Panama | *Acromyrmex octospinosus* | X | X | X | X |
| LG_CR2 | Costa Rica | *Atta cephalotes* | X | X | X | X |
| LG_CR3 | Costa Rica | *Acromyrmex echinatior* | X | X | X | X |
| LG_CR4 | Costa Rica | *Acromyrmex echinatior* | X | X | X | X |
| LG_CR5 | Costa Rica | *Atta cephalotes* | X | X | X | X |
| LG_Pe2 | Peru | *Acromyrmex* sp. | X | X | X | X |
| LG_CR6 | Costa Rica | *Atta cephalotes* | X | X | X | X |
| LG_ARG1 | Argentina | *Acromyrmex laticeps* | X | | | X |
| LG_ARG2 | Argentina | *Acromyrmex niger* | X | X | | X |

families associated with the degradation of recalcitrant polysaccharides (17), suggesting that bacteria are not playing a prominent role in biomass degradation in fungus gardens. Moreover, our metaproteomic results revealed that although > 81,000 spectra could be confidently mapped to a diversity of *L. gongylophorus* lignocellulases, only a single spectra could be mapped to a bacterial GH8, a CAZyme family that in bacteria is often associated with cell wall modification rather than plant biomass degradation (49). Previous work identified a bacterial glucosidase produced in fungus gardens (17), however, and it remains a possibility that bacterial enzymes not present in the protein databases used in this study could be playing a role in biomass degradation.

Most of the highly abundant *L. gongylophorus* enzymes identified in the metaproteomic data are predicted to degrade pectin, xylan, starch, and proteins (Table 3.2, Figure 3.4a). This is consistent with studies showing that both hemicellulases and proteases are produced by this organism (20, 50, 51), and that activity against these substrates can be consistently detected in whole fungus gardens (52). Moreover, culture-based work has shown that *L. gongylophorus* can produce enzymes active against a variety of polysaccharides when grown in pure culture (53,

54). Our finding that enzymes targeting hemicelluloses are among the most highly abundant in our metaproteomic datasets indicates that these polymers are likely the primary polysaccharides degraded by *L. gongylophorus* in fungus gardens. The abundance of proteases throughout all strata of our samples also suggests that plant proteins may be an important nutrient source for *L. gongylophorus* throughout the degradation process.

**Table 3.4.** Nucleotide and amino acid similarity of four CAZymes sequenced from 11 *L. gongylophorus* isolates.

| Amplicon | CAZy family | Length of nt alignment | Length of aa alignment | Average nt Identity | Average gaps in nt alignment | Average aa identity | Average gaps in aa alignment |
|---|---|---|---|---|---|---|---|
| Xyn I | GH 10 | 647 | 94 | 99.1 | 0.92 | 99.17 | 0 |
| Xyn II | GH 10 | 429 | 152 | 98.9 | 1.04 | 100 | 0 |
| CBH I | GH 7 | 507 | 147 | 98.8 | 1.02 | 98.94 | 0 |
| CBH II | GH 6 | 838 | 237 | 99.5 | 0.91 | 99.9 | 0 |

In addition to numerous hemicellulases and proteases, our genomic and metaproteomic analyses also identified putative cellulases (GH6, GH7, and GH9) produced by *L. gongylophorus*. Interestingly, this fungus has been reported to be incapable of growth on cellulose alone in pure culture (10). Moreover, recent investigations in cellulose degradation in fungus gardens have given conflicting results (16, 55), raising the question of whether *L. gongylophorus* or other microbes in fungus gardens can degrade this polymer. Our genomic and metaproteomic work is the first sequence-based evidence that *L. gongylophorus* both encodes cellulases and produces them in fungus gardens, suggesting that this fungus is contributing to some amount of cellulose degradation in these ecosystems. Results of our metaproteomic data also indicate that the abundance of the GH6 and GH7 cellulases increases in the lower strata of ant gardens (Figure 3.4b), suggesting that they may be produced primarily when less recalcitrant polymers have been exhausted. Interactions with bacteria or compounds produced by ants in

gardens may stimulate *L. gongylophorus* to produce these enzymes, potentially explaining why reports of the cellulolytic ability of this fungus in pure culture have varied (10, 11).

Our identification of numerous laccases and accessory oxidases with secretion signals in the genome of *L. gongylophorus* indicates that these enzymes may also be important for lignocellulose degradation in the ant-fungus symbiosis. Two laccases, an aryl-alcohol oxidase, and a glyoxal oxidase were among the most well-represented enzymes in our metaproteomic data (Table 3.2, Figure 3.4b), suggesting *L. gongylophorus* produces large quantities of these enzymes in ant gardens. Similar enzymes in other basidiomycetous fungi have been shown to play important roles in the degradation of plant polymers, especially lignin (56). Although recent work has indicated that significant amounts of lignin are not degraded in fungus gardens (16), it is likely that this polymer is physically linked to polysaccharides in plant cell walls, and partial degradation of lignin may therefore make otherwise unavailable polysaccharides more accessible to other lignocellulases. The use of laccases and redox-active enzymes may thus increase the efficiency of biomass degradation even if lignin itself is not used as a carbon source.

Alternatively, laccases may be necessary for the degradation of secondary metabolites in plant tissue that may be toxic to *L. gongylophorus* or the host ants, as indicated by recent work (57). This may explain why some laccases were found to be more abundant in top garden strata (Figure 3.4a), where plant toxins would first be encountered by *L. gongylophorus* and lignin degradation would not yet be necessary to access additional polysaccharides. It has previously been suggested that secondary metabolites produced by plants, or even by endophytic fungi living in plant tissue, may be important factors influencing the choice of foliar biomass foraged

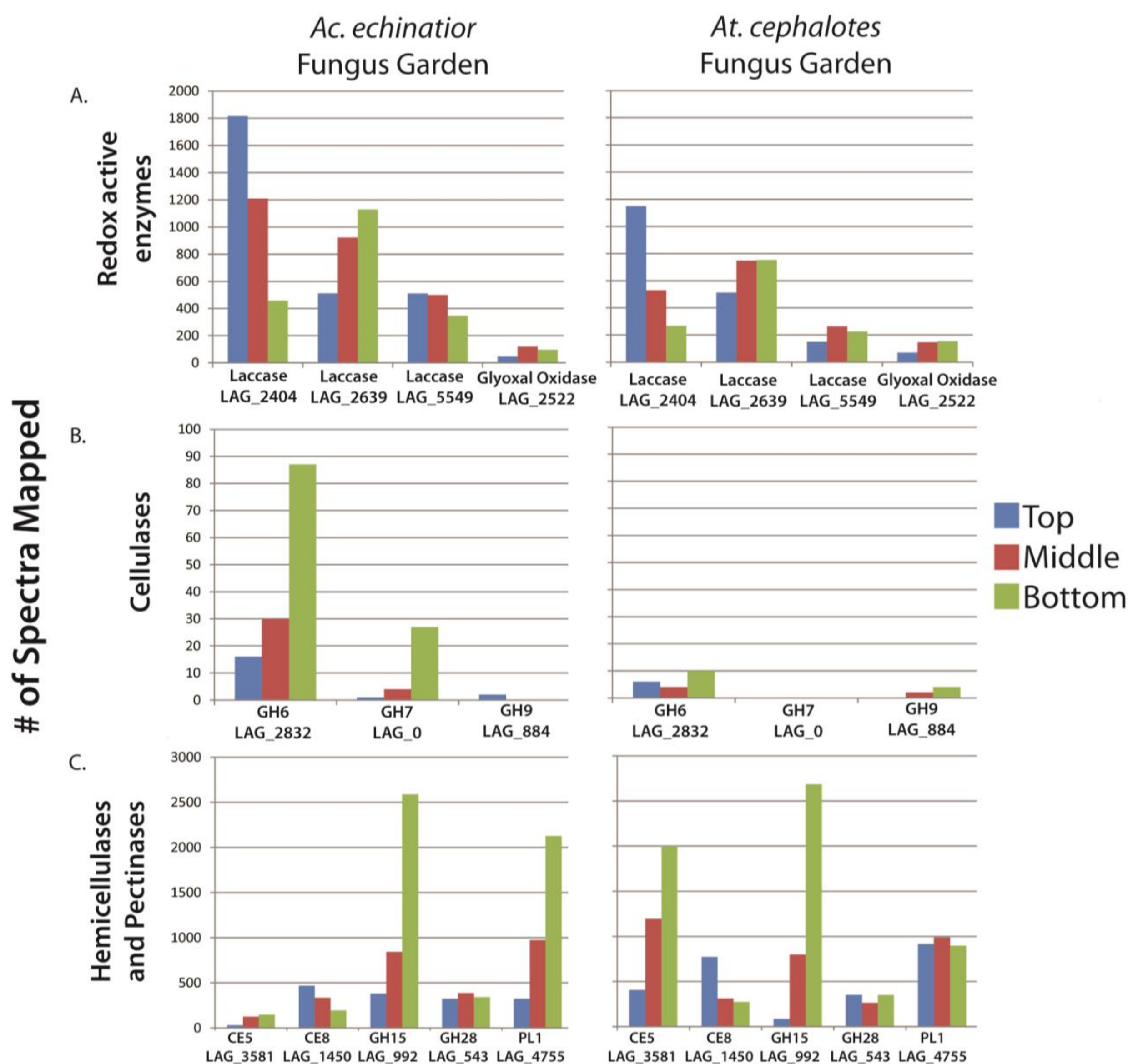



**Figure 3.4**. Bar chart showing the raw number of mass spectra that could be mapped to a subset of the most abundant lignocellulases identified in the six samples. Predicted FOLymes (a), cellulases (b), and hemicellulases and pectinases (c) are shown.

by the ants (58). The laccases and redox-active enzymes produced by *L. gongylophorus* may thus play an important role in detoxifying these compounds and broadening the range of plants that can be harvested and efficiently degraded in the ant-fungus symbiosis.

The composition of plant biomass in ant gardens is highly variable due to the diversity of plants foraged by the ants and changes in substrate composition throughout the degradation process (59). Therefore, the ability of *L. gongylophorus* to quickly alter the production and secretion of lignocellulases in response to nutrient availabilities is likely critical for efficient biomass processing. Consistent with this, we found distinct stratification in the lignocellulase profiles of the metaproteomes of both *Ac. echinatior* and *At. cephalotes* gardens, indicating that separate enzymatic cocktails are used by *L. gongylophorus* for the degradation of lignocellulose at different stages of biomass breakdown (Figures 3.2 and Appendix 2). A distinct cocktail of abundant CAZymes, laccases, and proteases was identified in the gardens throughout all stages of biomass degradation (see Appendix 2), suggesting that *L. gongylophorus* uses different sets of these enzymes to acquire nutrients from different plant substrates depending on the stage of biomass degradation.

Despite the similar overall patterns of stratified lignocellulase profiles between ant species, significantly different quantities of mass spectra were recovered from 65% of the lignocellulases when strata were compared directly (Fisher's Exact Test, $P < 0.05$). Both nests had been fed the same mixture of oak and maple leaves prior to the time of sampling, indicating that this difference is not due to diet. Our clustering analysis of overall mass spectra profiles for all *L. gongylophorus* proteins suggests that the physiology of *L. gongylophorus* may be influenced by the host ants (Figure 3.3). Although both *Acromyrmex* and *Atta* species culture *L. gongylophorus*, a number of factors could contribute to differences in the physiology of the fungal symbiont between nests. For example, differences in the hygienic practices of the host ants or interactions with bacteria co-inhabiting fungus gardens are possible explanations.

Species of *Atta* and *Acromyrmex* are distributed across the Americas in numerous distinct ecosystems, and their selection of plants varies depending on the ant species, location, and season (59). Because of this ecological variability, we explored the possibility that lignocellulases varied with the range or species of ant host. Contrary to our expectations, our comparison of four CAZymes reveals that the lignocellulases encoded by this fungus are highly conserved across both host ant species and geographic range (Figure 3.5, Tables 3.3 and 3.4). This is also surprising given previous reports identifying genetic diversity in *L. gongylophorus* cultures obtained from the same geographic area (60). Interestingly, two of the genes analyzed here encode predicted cellulases (families GH6 and GH7), consistent with previous work indicating that the degradation of cellulose is an important process in fungus gardens (16).

In this work, we demonstrate that the fungal symbiont of the ants, *L. gongylophorus*, encodes a diversity of plant biomass-degrading enzymes and is likely the primary driver of lignocellulose degradation in fungus gardens. Our metaproteomic analysis provides evidence that cellulases and redox-active enzymes produced by *L. gongylophorus* may be playing critical roles in this symbiosis by both degrading recalcitrant plant polymers and detoxifying secondary metabolites in plant tissue. Moreover, our phylogenetic analysis of *L. gongylophorus* CAZymes indicates that different species of leaf-cutter ants, which may inhabit different ecosystems and have colony sizes differing by millions of workers, appear to use the same highly conserved enzymes of *L. gongylophorus* to degrade their plant forage and convert it into usable nutrients. Our work highlights the importance of microbes to the herbivory of a dominant herbivore, as well as their importance to nutrient cycling and carbon turnover in Neotropical ecosystems.

### 3.6 References


1. **Douglas, AE.** 2010. The symbiotic habit. Princeton University Press, Princeton, New Jersey, USA.

2. **Moran, NA.** 2006. Symbiosis. *Curr. Biol.* **16:**R866.

3. **Klepzig, KD, Adams, AS, Handelsman, J, and Raffa, KF.** 2009. Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environ. Entomol.* **38:**67-77.

4. **Zilber-Rosenberg, I, and Rosenberg, E.** 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32:**723-735.

5. **Douglas, A.** 2009. The microbial dimension in insect nutritional ecology. *Funct. Ecol.* **23:**38-47.

6. **Mackie, RI.** 2002. Mutualistic fermentative digestion in the gastrointestinal tract: diversity and evolution. *Integr. Comp. Biol.* **42:**319-26.

7. **Hobson, PN, and Wallace, RJ.** 1982. Microbial ecology and activities in the rumen: part 1. *Crit. Rev. Microbiol.* **9:**165-225.

8. **Hölldobler, B, and Wilson, EO.** 2010. The leafcutter ants: civilization by instinct. W. W. Norton and Company, Inc., New York, NY.

9. **Martin, MM, and Weber, NA.** 1969. The cellulose-utilizing capability of the fungus cultured by the attine ant *Atta colombica tonsipes*. *Ann. Entomol. Soc. Am.* **62:**1386-1387.

10. **Abril, AB, and Bucher, EH.** 2002. Evidence that the fungus cultured by leaf-cutting ants does not metabolize cellulose. *Ecol. Lett.* **5:**325-328.

11. **Bacci, M, Anversa, MM, and Pagnocca, FC.** 1995. Cellulose degradation by *Leucocoprinus gongylophorus*, the fungus cultured by the leaf-cutting ant *Atta sexdens rubropilosa*. *A. Van Leeuw. J. Microb.* **67:**385-386.

12. **Bacci M, Ribeiro SB, Casarotto MEF, and Pagnocca FC.** 1995. Biopolymer-degrading bacteria from nests of the leaf-cutting ant *Atta sexdens Rubropilosa*. *Braz. J. Med. Biol. Res.* **28:**79-82.

13. **Currie, CR.** 2001. Prevalence and impact of a virulent parasite on a tripartite mutualism. *Oecologia* **128:**99-106.

14. **Rodrigues, A, Mueller, UG, Ishak, HD, Bacci Jr, M, and Pagnocca, FC.** 2011. Ecology of microfungal communities in gardens of fungus-growing ants (Hymenoptera: Formicidae): a year-long survey of three species of attine ants in central Texas. *FEMS Microbiol. Ecol.* **78:**244-255.

15. **Scott, JJ, Budsberg, KJ, Suen, G, Wixon, DL, Balser, TC, and Currie, CR.** 2010. Microbial community structure of leaf-cutter ant fungus gardens and refuse dumps. *PLoS One* **5:**e9922.

16. **Suen, G, Scott, JJ, Aylward, FO, Adams, SM, Tringe, SG, Pinto-Tomas, AA, Foster, CE, Pauly, M, Weimer, PJ, Barry, KW, Goodwin, LA, Bouffard, P, Li, L, Osterberger, J, Harkins, TT, Slater, SC, Donohue, TJ, and Currie, CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6:**e1001129.

17. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **6:**1688–1701.

18. **Currie, CR, Mueller, UG, and Malloch, D.** 1999. The agricultural pathology of ant fungus gardens. *Proc. Natl. Acad. Sci. USA* **96:**7998-8002.

19. **Currie, CR, and Stuart, AE.** 2001. Weeding and grooming of pathogens in agriculture by ants. *P. Roy. Soc. B-Biol. Sci.* **268:**1033-1039.

20. **Schiott, M, De Fine Licht, HH, Lange, L, and Boomsma, JJ.** 2008. Towards a molecular understanding of symbiont function: identification of a fungal gene for the degradation of xylan in the fungus gardens of leaf-cutting ants. *BMC Microbiol.* **8:**40.

21. **Chapela, IH, Rehner, SA, Schultz, TR, and Mueller, UG.** 1994. Evolutionary history of the symbiosis between fungus-growing ants and their fungi. *Science* **266:**1691-1694.

22. **Hölldobler, B, and Wilson, EO.** 2008. The superorganism: the beauty, elegance, and strangeness of insect societies. W. W. Norton & Company, New York, NY.

23. **Wilson, K.** 1987. Unit 2.4: Preparation of genomic DNA from bacteria, Current protocols in molecular biology. Wiley, New York, NY.

24. **Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, YJ, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, ML, Jarvie, TP, Jirage, KB, Kim, JB, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, McDade, KE, McKenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth, GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A, Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF, and**

**Rothberg, JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-80.

25. **Borodovsky, M, and Lomsadze, A.** 2011. Current protocols in bioinformatics, vol. Unit 4.6, eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Wiley Online Library.

26. **Stanke, M, Schoffmann, O, Morgenstern, B, and Waack, S.** 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7:**62.

27. **Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:**403-410.

28. **Tatusov, RL, Fedorova, ND, Jackson, JD, Jacobs, AR, Kiryutin, B, Koonin, EV, Krylov, DM, Mazumder, R, Mekhedov, SL, Nikolskaya, AN, Rao, BS, Smirnov, S, Sverdlov, AV, Vasudevan, S, Wolf, YI, Yin, JJ, and Natale, DA.** 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:**41.

29. **Finn, RD, Tate, J, Mistry, J, Coggill, PC, Sammut, SJ, Hotz, HR, Ceric, G, Forslund, K, Eddy, SR, Sonnhammer, EL, and Bateman, A.** 2008. The Pfam protein families database. *Nucleic Acids Res.* **36:**D281-8.

30. **Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W, and Lipman, DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:**3389-402.

31. **Cantarel, BL, Coutinho, PM, Rancurel, C, Bernard, T, Lombard, V, and Henrissat, B.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37:**D233-8.

32. **Levasseur, A, Piumi, F, Coutinho, PM, Rancurel, C, Asther, M, Delattre, M, Henrissat, B, Pontarotti, P, Asther, M, and Record, E.** 2008. FOLy: an integrated database for the classification and functional annotation of fungal oxidoreductases potentially involved in the degradation of lignin and related aromatic compounds. *Fungal Genet. Biol.* **45:**638-45.

33. **Rawlings, ND, Barrett, AJ, and Bateman, A.** 2011. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **40:**D343-50.

34. **Horton, P, Park, KJ, Obayashi, T, Fujita, N, Harada, H, Adams-Collier, CJ, and Nakai, K.** 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35:**W585-7.

35. **Petersen, TN, Brunak, S, von Heijne, G, and Nielsen, H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8:**785-6.

36. **Wisniewski, JR, Zougman, A, Nagaraj, N, and Mann, M.** 2009. Universal sample preparation method for proteome analysis. *Nat. Methods* **6:**359-62.

37. **Eng, J, McCormack, A, and Yates, J.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5:**976-989.

38. **Kim, S, Gupta, N, and Pevzner, PA.** 2008. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **7:**3354-63.

39. **Aylward, FO, Tremmel, DM, Starrett, GJ, Bruce, DC, Chain, P, Chen, A, Davenport, KW, Detter, C, Han, CS, Han, J, Huntemann, M, Ivanova, NN, Kyrpides, NC, Markowitz, V, Mavrommatis, K, Nolan, M, Pagani, I, Pati, A, Pitluck, S, Teshima, H, Deshpande, S, Goodwin, L, Woyke, T, and Currie, CR.** 2013. Complete genome of *Serratia* sp. strain FGI 94, a strain associated with leaf-cutter ant fungus gardens. *Genome Announc.* **1:**e0023912.

40. **Aylward, FO, Tremmel, DM, Bruce, DC, Chain, P, Chen, A, Walston Davenport, K, Detter, C, Han, CS, Han, J, Huntemann, M, Ivanova, NN, Kyrpides, NC, Markowitz, V, Mavrommatis, K, Nolan, M, Pagani, I, Pati, A, Pitluck, S, Deshpande, S, Goodwin, L, Woyke, T, and Currie, CR.** 2013. Complete genome of *Enterobacteriaceae* bacterium strain FGI 57, a strain associated with leaf-cutter ant fungus gardens. *Genome Announc.* **1:** e0023812.

41. **Markowitz, VM, Mavromatis, K, Ivanova, NN, Chen, IM, Chu, K, and Kyrpides, NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:**2271-8.

42. **Hyatt, D, Chen, GL, Locascio, PF, Land, ML, Larimer, FW, and Hauser, LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:**119.

43. **Edwards, IP, Upchurch, RA, and Zak, DR.** 2008. Isolation of fungal cellobiohydrolase I genes from sporocarps and forest soils by PCR. *Appl. Environ. Microbiol.* **74:**3481-9.

44. **Pruitt, KD, Tatusova, T, and Maglott, DR.** 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35:**D61-5.

45. **Edgar, RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32:**1792-7.

46. **Price, MN, Dehal, PS, and Arkin, AP.** 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26:**1641-50.

47. **Anisimova, M, and Gascuel, O.** 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55:**539-52.

48. **Letunic, I, and Bork, P.** 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:**127-8.

49. **Mazur, O, and Zimmer, J.** 2011. Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. *J. Biol. Chem.* **286:**17601-6.

50. **Semenova, TA, Hughes, DP, Boomsma, JJ, and Schiott, M.** 2011. Evolutionary patterns of proteinase activity in attine ant fungus gardens. *BMC Microbiol.* **11:**15.

51. **Schiott, M, Rogowska-Wrzesinska, A, Roepstorff, P, and Boomsma, JJ.** 2010. Leaf-cutting ant fungi produce cell wall degrading pectinase complexes reminiscent of phytopathogenic fungi. *BMC Biol.* **8:**156.

52. **De Fine Licht, HH, Schiott, M, Mueller, UG, and Boomsma, JJ.** 2010. Evolutionary transitions in enzyme activity of ant fungus gardens. *Evolution* **64:**2055-2069.

53. **Gomes De Siqueira, C, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJA.** 1998. Metabolism of plant polysaccharides by *Leucoagaricus gongylophorus*, the symbiotic fungus of the leaf-cutting ant *Atta sexdens*. *Appl. Environ. Microbiol.* **64:**4820-4822.

54. **Silva, A, Bacci, M, Jr., Pagnocca, FC, Bueno, OC, and Hebling, MJ.** 2006. Production of polysaccharidases in different carbon sources by *Leucoagaricus gongylophorus* Moller (Singer), the symbiotic fungus of the leaf-cutting ant *Atta sexdens* Linnaeus. *Curr. Microbiol.* **53:**68-71.

55. **Moller, IE, De Fine Licht, HH, Harholt, J, Willats, WG, and Boomsma, JJ.** 2011. The dynamics of plant cell-wall polysaccharide decomposition in leaf-cutting ant fungus gardens. *PLoS One* **6:**e17506.

56. **Baldrian, P.** 2006. Fungal laccases - occurrence and properties. *FEMS Microbiol. Rev.* **30:**215-42.

57. **Licht, HHDF, Schiøtt, M, Rogowska-Wrzesinska, A, Nygaard, S, Roepstorff, P, and Boomsma, JJ.** 2013. Laccase detoxification mediates the nutritional alliance between leaf-cutting ants and fungus-garden symbionts. *Proc. Natl. Acad. Sci. USA* **110:**583-587.

58. **Bael, SAV, Estrada, C, and Wcislo, WT.** 2011. Fungal-fungal interactions in leaf-cutting ant agriculture. *Psyche***:** doi:10.1155/2011/617478.

59. **Wirth, R, Herz, H, Ryel, RJ, Beyschlag, W, and Hoelldobler, B.** 2003. Herbivory of leaf-cutting ants: A case study on *Atta colombica* in the tropical rainforest of Panama. Springer, New York, NY.

60. **Mikheyev, AS, Mueller, UG, and Boomsma, JJ.** 2007. Population genetic signatures of diffuse co-evolution between leaf-cutting ants and their cultivar fungi. *Mol. Ecol.* **16:**209-216.

# Chapter 4: Convergence of Simple and Highly-Specific Microbiomes in Fungus-Growing Insects

**Publication in Preparation**: Frank O. Aylward, Garret Suen, Aaron S. Adams, Peter Biedermann, Jarrod J. Scott, Stephanie Malfatti, Tijana Glavina del Rio, Susannah G. Tringe, Michael Poulsen, Kenneth F. Raffa, Kier Klepzig, Cameron R. Currie.

**Author Contributions**: For this chapter I performed all computational experiments, and I conceived of all analyses with Cameron R. Currie. Other authors collected the samples, generated the metagenomic data, provided helpful advice for analyses, and helped with writing.

## 4.1 Abstract

Host niche and phylogeny driven factors are both important influences on the symbiotic microbiota of metazoans, but the relative impact of these forces on microbiome structure remain unclear. To understand the extent to which similar host lifestyles can lead to convergent microbiome structure, we characterized the microbiota associated with lineages of ants, beetles, and termites that have each independently evolved obligate symbioses with fungi they grow for food. We demonstrate that different fungus-growing insects associate with simple and highly similar bacterial communities comprising the genera *Enterobacter*, *Rahnella*, and *Pseudomonas* that appear to be even more specific to their host insects than even the obligate fungal cultivars. Our work demonstrates the drastic effect that microbial symbioses can have on the entire microbiomes of their hosts, and is an important advance to our understanding of the factors responsible for the structure and function of the symbiotic microbiota of metazoans.

## 4.2 Introduction

Symbioses between metazoans and microbial communities are ubiquitous in nature and have contributed to many of the watershed events in the history of life on Earth (1-3). These microbiomes, which range from simple consortia of relatively few species to highly complex and dynamic communities, benefit their hosts through the defense against pathogens (4), degradation

of recalcitrant dietary material (5, 6), and biosynthesis of essential nutrients (7). The vast physiological potential of microbes also plays a role in the acquisition of novel ecological strategies in metazoans, and microbial symbionts have even been argued to play an important role in host speciation (8, 9). Despite the importance of microbiomes for the evolution and physiology of their host, the forces that shape the composition and function of these microbial communities are not well understood.

Although microbiomes are highly dynamic systems, the forces that shape their structure and function can broadly be divided into two categories that describe their variation relative to their hosts (10, 11), which we refer to here as niche vs. phylogeny-driven factors. Niche-driven factors refer to the lifestyle and behavior of the host and include influences such as its diet and geographic distribution, while phylogeny-driven factors refer to co-evolutionary forces between the host and its microbiome that are strictly dependent on the evolutionary history of the host rather than its lifestyle. While phylogeny-driven factors lead to microbiome compositions that mirror the branching evolutionary histories of their host lineages, niche-driven forces lead to the emergence of convergent microbiome structures in distantly-related hosts, providing those hosts occupy similar ecological niches.

Attempts to examine the relative importance of these influences in shaping microbiome structure have led to conflicting results. In mammals, where this topic has been explored extensively, studies of the microbial communities in the feces of animals have indicated that diet is a critical factor that determines the overall phylogenetic structure of microbial communities residing in the gastrointestinal tract (12, 13). Other studies of closely-related wild primates, however, have suggested that host phylogeny is the most important factor that determines gut microbiome structure (11). Moreover, analyses of specific bacterial groups residing in the

mammalian gut have also hinted at the importance of phylogeny-driven factors (14). In non-mammalian systems, such as the honey bee gut and termite hindgut, an interplay between host phylogeny and diet is also likely to be responsible for highly-specific microbiomes of these insects (5, 15, 16).

In this study we sought to address the relative importance of niche vs. phylogeny-driven forces in shaping host microbiome structure by comparing the microbial communities associated with lineages of fungus-growing insects that have independently evolved mutualistic fungal symbioses. Fungus-growing ants (Tribe: Attini), beetles (family: Scolytinae), and termites (subfamily: Macrotermitinae) all engage in obligate nutritional symbioses with mutualistic fungi (17), while pine beetles (genus *Dendroctonus*), although not considered strict fungal farmers, also associate with mutualistic fungi that they consume for food (18-20). Unlike in the mammalian gut, where common ancestry precludes the complete exclusion of phylogeny-driven factors as a determinant of microbiome structure, the independent origins of fungus growing behavior in insects provides a unique opportunity to assess the extent to which similar lifestyles influence host microbiome structure. Here, we use a comparative metagenomic approach to characterize the microbial communities associated with these different fungus-growing insect lineages to better understand the evolutionary dynamics of host microbiomes.

## 4.3 Materials and Methods

### 4.3.1 Sample Collection, Processing, and Sequencing

Details regarding the time and location in which samples were acquired can be found in Figure 4.1 and Table 4.1. Sampling of the leaf-cutter ant species *Atta cephalotes* and *Atta colombica* and the mountain pine beetle *Dendroctonus ponderosae* have been previously described (21, 22) (see also Chapter 2 and Appendix 4). For all fungus garden, gallery, or whole

insect samples the bacterial fraction was isolated by gently vortexing the samples in 1% Phosphate Buffered Saline (PBS) and 0.1% Tween before conducting a modified differential centrifugation procedure, as previously described (21). For the Ambrosia beetles, fungus-growing termites, mountain pine beetles, and southern pine beetles whole-insect or larval samples were also prepared for sequencing. For these samples, 90-300 whole insects were pooled for each sample. Once the bacterial fraction of each sample was isolated, DNA was extracted using the bacterial extraction protocol available in the Qiagen DNeasy plant maxikit (Qiagen Sciences, Germantown, MD), which has been shown to yield an accurate representation of community DNA (22). Community metagenomes were subsequently generated from the using Roche 454 Titanium pyrosequencing (23), and assemblies were generated using Newbler v.2.1 with default parameters. Details of the metagenomes and assembly statistics can be found in Table 4.2.

**4.3.2 Phylogenetic Binning and Relative Abundance Estimation**

Phylogenetic bins from all metagenomes were generated using a combination of BLASTN (24) and PhymmBL (25). All contigs and singletons were first compared to a reference data set containing all completely sequenced bacterial and archaeal genomes available in the NCBI as of 1 January 2012. All contigs having BLASTN hits with E values < $1e^{-10}$ were classified according to their best hit, while all other contigs were subsequently classified using PhymmBL. Contigs with no BLASTN matches that were classified by PhymmBL, with

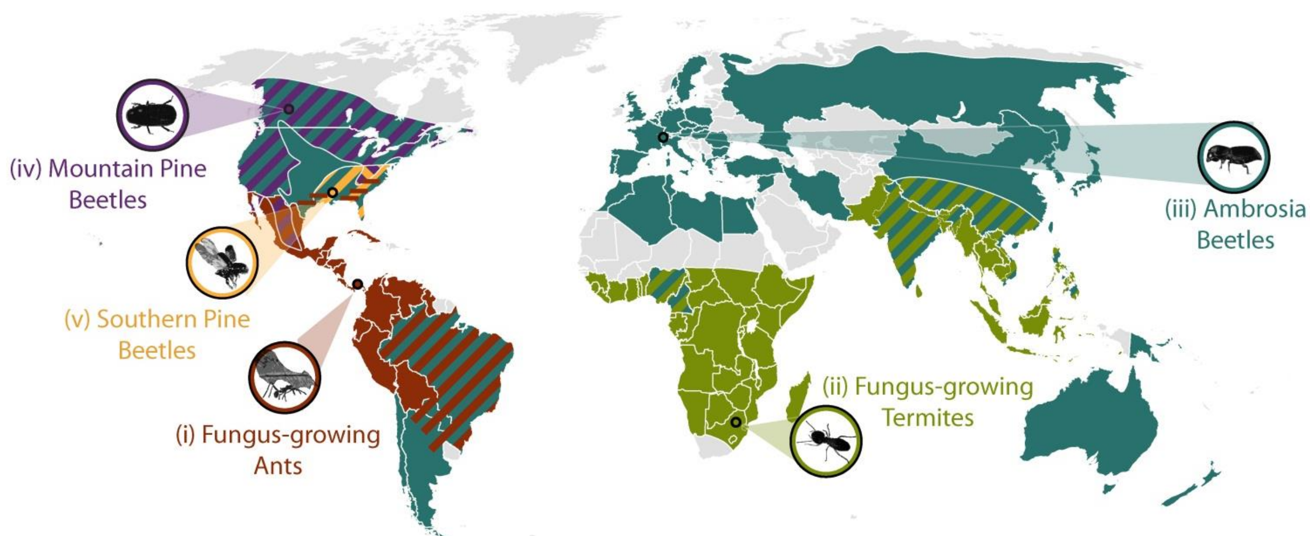



**Figure 4.1**. Map showing the worldwide distribution of fungus-growing insects together with the specific locations in which samples were acquired for this study. Collection details can be found in Table 1. Colored regions indicate the range of the insects, and areas where these ranges overlap are indicated with cross-hatching.

confidence scores < 50 considered "unclassified". A relative abundance value for each contig was calculated by multiplying the length of the contig by its percent coverage. The relative abundance of each phylogenetic group in the metagenomes was then calculated by summing the abundance values for each contig classified in a particular group.

### 4.3.3 Composite Genome Reconstruction

Composite genomes were reconstructed from the assembled community metagenomes using a combination of BLASTN-based homology to reference genomes, PhymmBL binning, and differential coverage binning. All contigs > 800 bp were first binned at the genus level based on their BLASTN or PhymmBL assignment, and plots of contig length vs. coverage were then generated for each genus in each metagenome using R. Contigs having similar coverage were then placed in the same bin, and the best BLASTN hit of each contig was again cross-referenced

to ensure each bin had consistent best-BLASTN hits. Only bins with > 1 Mb of sequence were considered for subsequent analysis (details in Table 4.3). Methods similar to this have recently been shown to produce consistent and high-fidelity genome bins in other metagenomes (26).

### 4.3.4 Average Nucleotide Identity (ANI) Comparisons and BLASTN Mapping

ANI values were generated for pairs of complete or composite genomes by averaging the % identity of reciprocal best BLASTN hits of each gene using the parameters –X 150, -q -1, -F F, -e $1e^{-5}$. These BLASTN parameters have previously been shown to provide accurate ANI values (27). The genes used for these analyses were predicted independently from each genome using Prodigal (28). BLASTN was also used to map genes and contigs onto a subset of *Enterobacteriaceae* or *Pseudomonas* genomes using the same parameters, and only best hits were retained.

### 4.3.5 Phylogenetic Analysis of Complete Genomes

Phylogenetic trees of select *Enterobacteriaceae* and *Pseudomonas* genomes were constructed from a concatenated alignment of the highly-conserved single-copy housekeeping genes *rpoB*, *rplB*, *pyrG*, *leuS*, *lepA*, *ileS*, and *recA*, which have previously been shown to be useful phylogenetic markers (29). Genes were identified by comparing genes in each genome to the corresponding genes in *Escherichia coli* K12 or *Pseudomonas putida* KT2440 as available on the Ribosomal Database Project (http://rdp.cme.msu.edu/) (30) using BLASTN (e-value of $1e^{-100}$). Alignments for each gene were constructed separately using MAFFT and combined prior to

**Table 4.1**. Details regarding the time, location, and species of insect collected in this study. A map of collections is presented in Figure 1.

| Sample | Sample Sequenced | Insect Host | Insect Species | Site of Sampling | GPS Coord. | Date Sampled |
|---|---|---|---|---|---|---|
| Ambrosia Beetle Adults | Whole Insect | Ambrosia Beetle | *Xyleborinus saxesenii* | Spilwald Forest, Switzerland | 47.0, 7.3 | 11/14/09 |
| Ambrosia Beetle Gallery | Fungus Gallery | Ambrosia Beetle | *Xyleborinus saxesenii* | Spilwald Forest, Switzerland | 47.0, 7.3 | 11/14/09 |
| Ambrosia Beetle Larvae | Insect Larvae | Ambrosia Beetle | *Xyleborinus saxesenii* | Spilwald Forest, Switzerland | 47.0, 7.3 | 11/14/09 |
| *Atta cephalotes* Fungus Garden | Fungus Garden | Fungus-growing Ant | *Atta cephalotes* | Gamboa, Panama | 9.1, -79.7 | 04/17/09 |
| *Atta colombica* Fungus Garden Bottom | Fungus Garden (Bottom strata) | Fungus-growing Ant | *Atta colombica* | Gamboa, Panama | 9.1, -79.7 | 04/17/09 |
| *Atta colombica* Fungus Garden Top | Fungus Garden (Top Strata) | Fungus-growing Ant | *Atta colombica* | Gamboa, Panama | 9.1, -79.7 | 04/17/09 |
| *Acromyrmex echinatior* Fungus Garden | Fungus Garden | Fungus-growing Ant | *Acromyrmex echinatior* | Gamboa, Panama | 9.1, -79.7 | 04/17/09 |
| *Apterostigma dentigerum* Fungus Garden | Fungus Garden | Fungus-growing Ant | *Apterostigma dentigerum* | Gamboa, Panama | 9.1, -79.7 | 04/17/09 |
| *Cyphomyrmex longiscapus* Fungus Garden | Fungus Garden | Fungus-growing Ant | *Cyphomyrmex longiscapus* | Gamboa, Panama | 9.1, -79.7 | 04/17/09 |
| *Trachymyrmex zeteki* Fungus Garden | Fungus Garden | Fungus-growing Ant | *Trachymyrmex zeteki* | Gamboa, Panama | 9.1, -79.7 | 04/19/10 |
| Fungus-growing Termite Fugus Comb | Fungus Comb | Fungus-growing Termite | *Macrotermes natalensis* | Mookgopong, South Africa | -24.7, 28.8 | 01/15/10 |
| Fungus-growing Termite Workers | Whole Insect | Fungus-growing Termite | *Macrotermes natalensis* | Mookgopong, South Africa | -24.7, 28.8 | 01/15/10 |
| Mountain Pine Beetle Hybrid Pine Gallery | Fungus Gallery | Mountain Pine Beetle | *Dendroctonus ponderosae* | McBride, Canada | 53.3, -120.3 | 06/30/09 |
| Mountain Pine Beetle Hybrid Pine Beetle | Whole Insect | Mountain Pine Beetle | *Dendroctonus ponderosae* | McBride, Canada | 53.3, -120.3 | 06/30/09 |
| Mountain Pine Beetle Lodgepole Beetle | Whole Insect | Mountain Pine Beetle | *Dendroctonus ponderosae* | Grand Prairie, Canada | 54.7, -119.7 | 06/30/09 |
| Mountain Pine Beetle Lodgepole Gallery | Fungus Gallery | Mountain Pine Beetle | *Dendroctonus ponderosae* | Grand Prairie, Canada | 54.7, -119.7 | 06/30/09 |
| Southern Pine Beetle | Whole Insect | Southern Pine Beetle | *Dendroctonus frontalis* | Homochitto National Forest, USA | 31.5, -90.9 | 08/15/09 |
| Southern Pine Beetle Gallery | Fungus Gallery | Southern Pine Beetle | *Dendroctonus frontalis* | Homochitto National Forest, USA | 31.5, -90.9 | 08/15/09 |

**Table 4.2**. Sequencing and assembly statistics for the 18 fungus-growing insect metagenomes analyzed in this study.

| Sample | IMG ID | Mbp of Assembly | Largest Contig | N50 Contig Size | Mbp of Reads | No. of Reads After QC |
|---|---|---|---|---|---|---|
| Ambrosia Beetle Adults | 2043231000 | 32.3 | 42,934 | 1,793 | 265.3 | 651,722 |
| Ambrosia Beetle Gallery | 2084038008 | 40.5 | 170,590 | 1,537 | 343.7 | 861,709 |
| Ambrosia Beetle Larvae | 2228664020 | 14.4 | 127,605 | 5,550 | 156.3 | 415,585 |
| *Atta cephalotes* Fungus Garden | 2029527004 | 42.3 | 168,281 | 3,812 | 385.3 | 1,068,791 |
| *Atta colombica* Fungus Garden Bottom | 2029527006 | 85.9 | 361,872 | 1,443 | 355.7 | 862,360 |
| *Atta colombica* Fungus Garden Top | 2029527005 | 105.1 | 359,777 | 1,411 | 407.9 | 998,047 |
| *Acromyrmex echinatior* Fungus Garden | 2035918000 | 62.8 | 290,993 | 1,722 | 354.1 | 863,365 |
| *Apterostigma dentigerum* Fungus Garden | 2029527003 | 101.3 | 515,338 | 1,748 | 431.9 | 1,229,061 |
| *Cyphomyrmex longiscapus* Fungus Garden | 2030936005 | 90.7 | 264,203 | 3,180 | 384.8 | 952,219 |
| *Trachymyrmex zeteki* Fungus Garden | 2084038018 | 67.7 | 216,696 | 2,154 | 453.0 | 1,248,690 |
| Fungus-growing Termite Fugus Comb | 2065487014 | 13.7 | 286,238 | 53,321 | 698.1 | 1,849,062 |
| Fungus-growing Termite Workers | 2065487013 | 46.5 | 129,579 | 2,293 | 625.1 | 1,736,328 |
| Mountain Pine Beetle Hybrid Pine Gallery | 2029527007 | 30.2 | 142,658 | 10,633 | 335.4 | 821,117 |
| Mountain Pine Beetle Hybrid Pine Beetle | 2032320009 | 20.6 | 59,558 | 3,499 | 267.2 | 740,695 |
| Mountain Pine Beetle Lodgepole Beetle | 2035918003 | 61.3 | 140,467 | 5,627 | 153.2 | 423,929 |
| Mountain Pine Beetle Lodgepole Gallery | 2032320008 | 32.7 | 47,283 | 1,470 | 182.3 | 492,214 |
| Southern Pine Beetle | 2044078006 | 22.7 | 249,649 | 23,099 | 411.1 | 1,155,933 |
| Southern Pine Beetle Gallery | 2044078007 | 45.7 | 348,445 | 7,064 | 585.5 | 1,578,386 |

**Table 4.3.** Summary of the 42 composite genomes reconstructed in this study and the sequenced genomes to which they were found to have the highest Average Nucleotide Identity (ANI).

| Bin ID | Metagenome | Size (Mbp) | Closest Sequenced Relative | % Complete | ANI with Closest Relative |
|---|---|---|---|---|---|
| ***Enterobacter*** | | | | | |
| Acro. Ent. 1 | *Acromyrmex echinatior* | 2.1 | *Enterobacter* sp. strain FGI 35 | 45.1 | 98.6 |
| Acro. Ent. 2 | *Acromyrmex echinatior* | 1.4 | *Enterobacter cloacae* ATCC 13047 | 22.8 | 88.6 |
| Aptero. Ent. 1 | *Apterostigma dentigerum* | 1.1 | *Enterobacter* sp. strain FGI 35 | 37.0 | 90.7 |
| Aptero Ent. 2 | *Apterostigma dentigerum* | 4.0 | *Enterobacter cloacae* ATCC 13047 | 70.4 | 87.9 |
| ACEF Ent. 1 | *Atta cephalotes* | 4.4 | *Enterobacter cloacae* ATCC 13047 | 74.7 | 87.8 |
| ACEF Ent. 2 | *Atta cephalotes* | 4.1 | *Enterobacter* sp. strain FGI 35 | 81.5 | 96.9 |
| ACOLB Ent. 1 | *Atta colombica* Garden Bottom | 5.7 | *Enterobacter cloacae* ATCC 13047 | 58.0 | 87.6 |
| ACOLB Ent. 2 | *Atta colombica* Garden Bottom | 2.3 | *Enterobacter cloacae* ATCC 13047 | 35.8 | 87.2 |
| ACOLB Ent. 3 | *Atta colombica* Garden Bottom | 2.6 | *Enterobacter* sp. strain FGI 35 | 52.5 | 98.0 |
| ACOLT. Ent. 2 | *Atta colombica* Garden Top | 2.8 | *Enterobacter cloacae* ATCC 13047 | 41.4 | 87.5 |
| ACOLT Ent. 3 | *Atta colombica* Garden Top | 3.8 | *Enterobacter cloacae* ATCC 13047 | 37.0 | 87.0 |
| ACOLT Ent. 1 | *Atta colombica* Garden Top | 4.9 | *Enterobacter* sp. strain FGI 35 | 80.9 | 96.5 |
| Cypho. Ent. 1 | *Cyphomyrmex longiscapus* | 4.5 | *Enterobacter cloacae* ATCC 13047 | 96.3 | 87.9 |
| FGT Worker Ent. 1 | Fungus-growing Termite Worker | 4.4 | *Enterobacter cloacae* ATCC 13047 | 63.6 | 89.3 |
| Trachy. Ent. 1 | *Trachymyrmex zeteki* | 2.6 | *Enterobacter* sp. strain FGI 35 | 74.7 | 98.6 |
| Trachy. Ent. 2 | *Trachymyrmex zeteki* | 5.4 | *Enterobacter asburiae* LF7a | 86.4 | 97.9 |
| ***Rahnella*** | | | | | |
| Amb. Adult Rn | Ambrosia Beetle Adult | 6.4 | *Rahnella aquatilis* HX2 | 85.8 | 88.3 |
| Amb Gal. Rn | Ambrosia Beetle Gallery | 5.3 | *Rahnella aquatilis* HX2 | 100.0 | 87.0 |
| Amb. Lar. Rn | Ambrosia Beetle Larvae | 3.7 | *Rahnella aquatilis* HX2 | 92.0 | 87.3 |
| FGTG Rn. 1 | Fungus-growing Termite Garden | 5.0 | *Rahnella aquatilis* HX2 | 100.0 | 87.2 |
| FGTW. Rn. 2 | Fungus-growing Termite Worker | 5.2 | *Rahnella aquatilis* HX2 | 100.0 | 87.2 |
| MPB HB Rn. | MPB Hybrid Beetle | 2.9 | *Rahnella aquatilis* HX2 | 76.5 | 87.4 |
| MPB HG Rn. | MPB Hybrid Gallery | 4.8 | *Rahnella aquatilis* HX2 | 98.8 | 87.2 |
| MPB LPB Rn. | MPB Lodgepole Beetle | 5.2 | *Rahnella aquatilis* HX2 | 96.9 | 87.2 |
| MPB LPG Rn. | MPB Lodgepole Gallery | 5.0 | *Rahnella aquatilis* HX2 | 99.4 | 87.2 |
| SPB Gallery Rn. | SPB Gallery | 3.7 | *Rahnella aquatilis* HX2 | 93.8 | 83.4 |
| ***Pseudomonas*** | | | | | |
| ACOL Ps. | *Atta colombica* Top and Bottom | 1.2 | *Pseudomonas* sp. strain FGI 182 | 69.8 | 97.7 |
| Acro. Ps. | *Acromyrmex echinatior* | 5.8 | *Pseudomonas* sp. strain FGI 182 | 100.0 | 96.8 |
| Amb. Ad. Ps. | Ambrosia Beetle Adult | 5.9 | *Pseudomonas fluorescens* SBW25 | 98.8 | 88.0 |
| Amb. Gallery Ps. | Ambrosia Beetle Gallery | 4.6 | *Pseudomonas fluorescens* SBW25 | 86.4 | 92.3 |
| Amb. Larvae Ps. | Ambrosia Beetle Larvae | 6.5 | *Pseudomonas fluorescens* SBW25 | 98.1 | 87.9 |
| Aptero. Ps. | *Apterostigma dentigerum* | 4.0 | *Pseudomonas* sp. strain FGI 182 | 90.1 | 90.1 |
| Cypho. Ps. | *Cyphomyrmex longiscapus* | 2.0 | *Pseudomonas* sp. strain FGI 182 | 78.4 | 90.7 |
| FGT Worker Ps. | Fungus-growing Termite Worker | 2.7 | *Pseudomonas syringae* B728a | 65.4 | 84.3 |
| MPB HB Ps. 1 | MPB Hybrid Beetle | 4.1 | *Pseudomonas fluorescens* SBW25 | 53.1 | 90.1 |
| MPB HB Ps. 2 | MPB Hybrid Beetle | 6.5 | *Pseudomonas fluorescens* SBW25 | 51.2 | 91.1 |
| MPB HG Ps. | MPB Hybrid Gallery | 5.5 | *Pseudomonas fluorescens* SBW25 | 92.0 | 87.9 |
| MPB LP Beetle Ps. 1 | MPB Lodgepole Beetle | 2.6 | *Pseudomonas syringae* B728a | 45.7 | 84.3 |
| MPB LP Beetle Ps. 2 | MPB Lodgepole Beetle | 6.0 | *Pseudomonas fluorescens* SBW25 | 85.2 | 87.7 |
| SPB Beetle Ps. 1 | SPB Beetle | 6.8 | *Pseudomonas fluorescens* Pf5 | 67.3 | 86.9 |
| SPB Beetle Ps. 2 | SPB Beetle | 6.3 | *Pseudomonas fluorescens* SBW25 | 82.1 | 88.9 |
| SPB Gallery Ps. | SPB Gallery | 5.4 | *Pseudomonas fluorescens* SBW25 | 56.2 | 89.4 |

phylogenetic analysis. Maximum-likelihood phylogenies were constructed using FastTree (31), and local support values were calculated using the Shimodaira-Hasegawa (SH) test (32).

**4.3.6 Functional Comparisons of Metagenomes**

The 18 fungus-growing insect metagenomes generated here were compared to 24 gut-associated metagenomes and 33 non-host associated metagenomes that are publicly available on IMG/M (33) or MG-RAST (34) (details in Table 4.4). Predicted proteins obtained from either IMG/M or MG-RAST were compared to the Clusters of Orthologous Groups (COG) (35) and Protein Families (Pfam) (36) databases using RPS-BLAST (37) (e-value < $1e^{-5}$). Best RPS-BLAST hits were compiled in a matrix and normalized by the total number of COG or Pfam hits. COG or Pfam families representing < 0.1% of total annotated proteins were excluded from subsequent analyses. The normalized matrices were then used to for Principle Component Analyses (PCA) using the R module FactoMineR (http://factominer.free.fr/).

**4.4 Results**

**4.4.1 General Properties of the Metagenomes and Phylogenetic Classification**

Sequencing of the 18 fungus-growing insect metagenomes yielded between 0.4-1.8 million high-quality reads comprising between 153.2-698.1 Mpb of total sequence, and after assembly the community metagenomes comprised between 14.4-105.1 Mbp (Table 4.2). Analysis of the relative abundance of phylogenetic groups in the metagenomes revealed that the families *Enterobacteriaceae* and *Pseudomonadaceae* and the genera *Pseudomonas*, *Enterobacter*, and *Rahnella* were particularly well-represented in the samples (Figure 4.2A, B). A genus-level rank-abundance curve of the 18 combined metagenomes demonstrates that *Pseudomonas*, *Enterobacter*, and *Rahnella* are overall substantially more abundant than other

genera. Generally, the genus *Enterobacter* was more commonly identified in the fungus-growing ant samples, while the genus *Rahnella* was more abundant in the termite- and

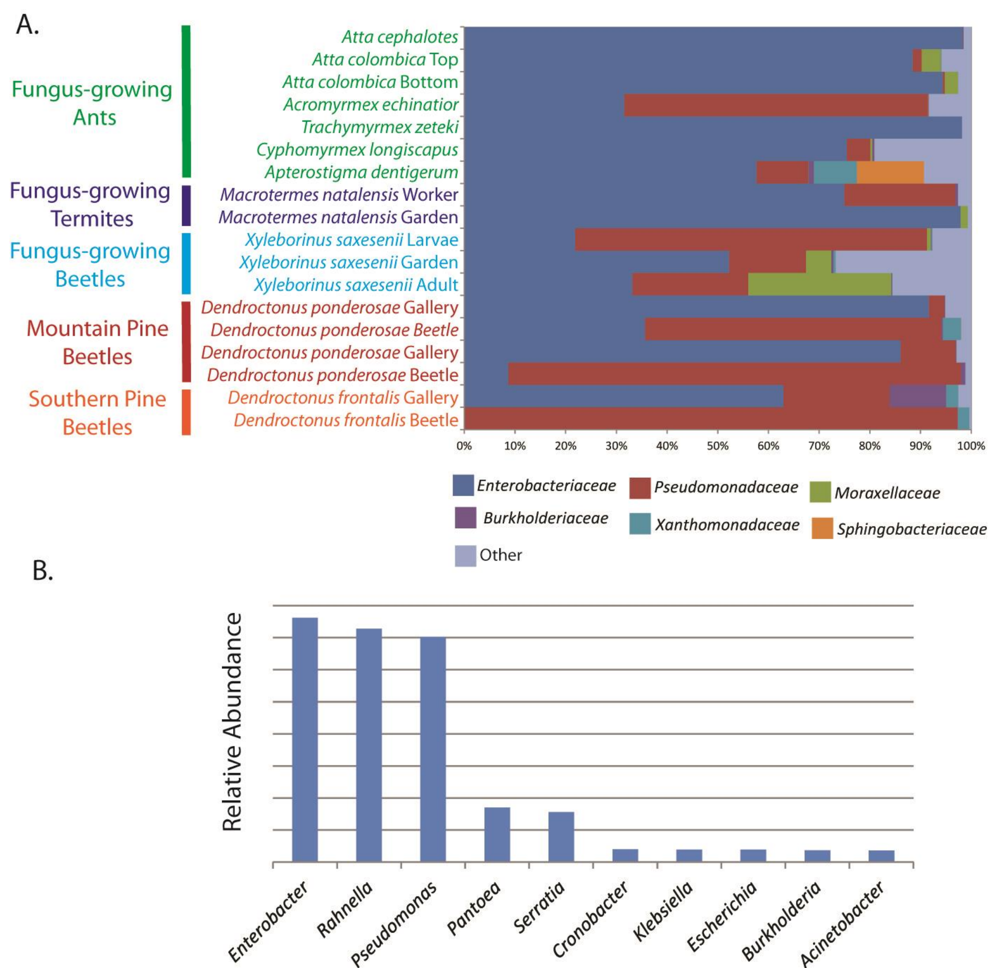



**Figure 4.2.** Relative abundance of bacterial families identified in the 18 fungus-growing insect-associated community metagenomes (A). Rank-abundance curve of genera identified in the metagenomes (B). Contigs were binned using both PhymmBL and BLASTN, and relative abundance values were calculated by multiplying the length of each contig by its coverage.

beetle-associated samples. Sequences matching to *Pseudomonas* were found in ant, beetle, and termite-associated samples.

The mountain pine beetle, southern pine beetle, fungus-growing termite, and Ambrosia beetle metagenomes were sequenced using both the whole insects or larvae in addition to fungus gardens. Additionally, the four mountain pine beetle metagenomes analyzed here were constructed from whole beetles and galleries associated with native and non-native tree hosts (*Pinus contorta* and hybrid *P. banksiana-P. contorta*, respectively) from two different geographic regions (Table 4.1). The phylogenetic composition of the whole insect samples generally coincided with the associated fungus garden samples, although the abundance of different phylogenetic groups differed in some cases. For example, the abundance of the genus *Pseudomonas* was higher in all whole-insect samples compared to their associated fungus gardens, while the opposite trend was observed for the family *Enterobacteriaceae*.

### 4.4.2 Phylogenetic Mapping of Sequences

Because the majority of all sequences in the metagenomes were classified as belonging to the family *Enterobacteriaceae* and genus *Pseudomonas*, we sought to gain more detailed insight into the phylogenetic groups that were present in each sample by mapping these sequences onto a phylogeny of available genomes from these groups. This analysis revealed a high degree similarity between the overall phylogenetic profiles of the Ambrosia beetle, termite, and pine beetle samples, with numerous sequences similar to *Rahnella aquitalis* and *Pseudomonas fluorescens* dominating the phylogenetic profile of these metagenomes (Figure 4.3). Although the fungus-growing ant samples were also dominated by sequences similar to *Enterobacteriaceae* and *Pseudomonas*, these microbiomes primarily consisted of sequences mapping to the species *Enterobacter cloacae, Enterobacter aerogenes*, and *Pseudomonas putida*.

### 4.4.3 Analysis of Composite Genomes

To investigate the similarity of these microbiomes in more detail, we reconstructed composite genomes of the dominant phylogenetic groups in each metagenome (Table 4.3). Overall we were able to construct 42 composite genomes of at least 1 Mb in size from the genera *Enterobacter*, *Rahnella*, and *Pseudomonas*. Comparison of the average nucleotide identity (ANI)

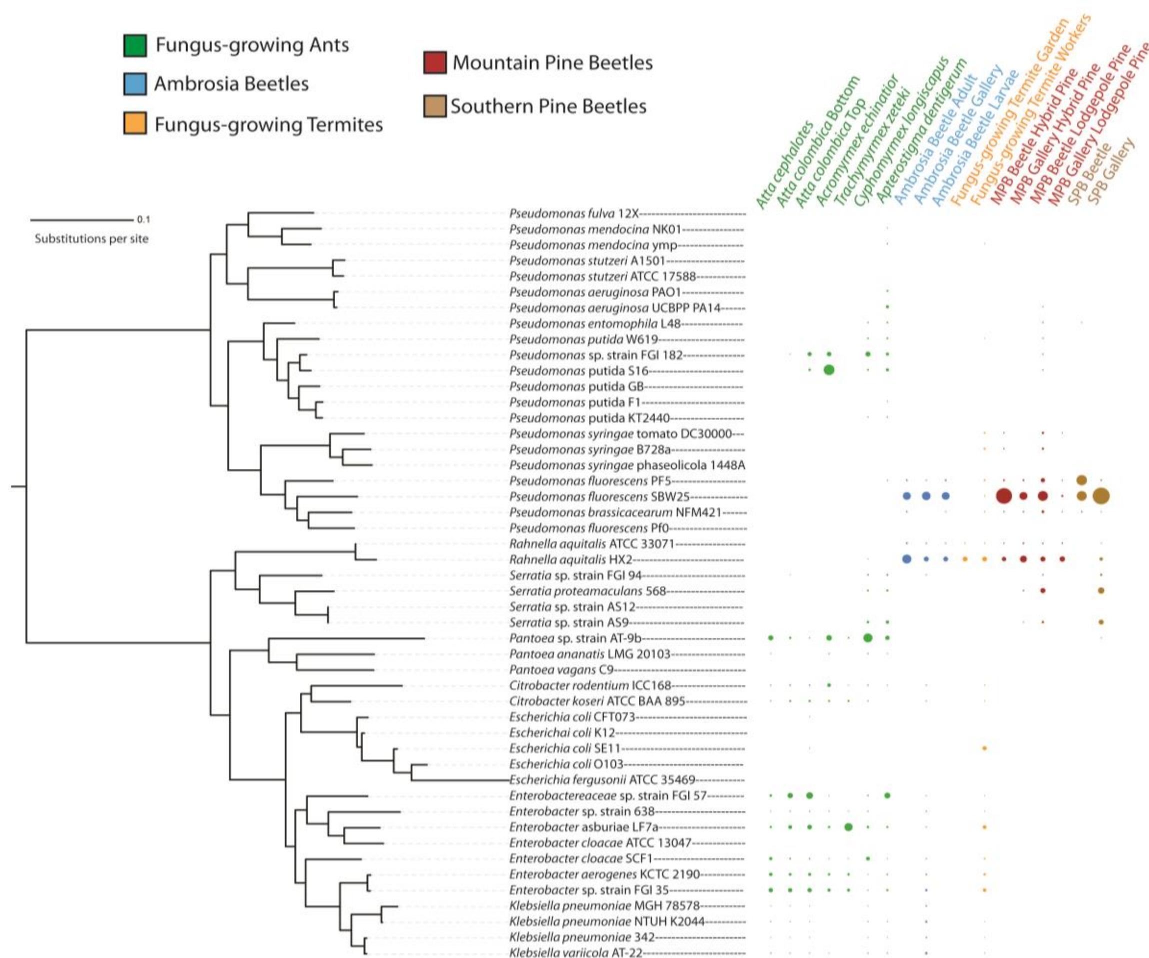



**Figure 4.3**. Phylogenetic context of genes predicted from contigs classified as belonging to either the family *Enterobacteriaceae* or genus *Pseudomonas* in the 18 fungus-growing insect metagenomes. Genes were predicted from contigs using Prodigal and mapped onto select genomes currently available for these phylogenetic groups using BLASTN. Only best BLASTN hits were retained. The size of the dots indicates the number of genes mapping. The phylogeny is based on a nucleotide alignment of 7 concatenated housekeeping genes (see Materials and Methods for details).

of these genomes to each other revealed that highly similar subgroups were present in multiple different insect systems (Figure 4.4). *Enterobacter* genomes were primarily identified in the fungus-growing ant samples, although we also reconstructed a composite genome from this genus from our sample of fungus-growing termite workers. Based on their ANI to previously-sequenced genomes we divided the composite *Enterobacter* genomes into an *Enterobacter aerogenes* group and an *Enterobacter cloacae* group. Composite genomes in the *Enterobacter aerogenes* group were identified in all of the fungus-growing ant samples except that of *Cyphomyrmex longiscapus* fungus garden, and pairwise ANI of the six genomes reconstructed revealed > 95% nucleotide identity between all of these genomes. A similar trend was observed in the composite genomes in the *Enterobacter cloacae* group, although 10 of these genomes were reconstructed, and they were identified in all ant samples as well as the termite worker sample. Pairwise ANI values of these genomes was slightly lower than that observed between *Enterobacter aerogenes* genomes (85-95%), suggesting the specificity of the *Enterobacter cloacae* group for fungus-growing ants may be lower than that of the *Enterobacter aerogenes* group.

The *Rahnella* composite genomes showed a similar level of whole-genome similarity between samples as the *Enterobacter* groups (Figure 4.4). The 10 *Rahnella* composite genomes that we reconstructed were represented in a broader diversity of insects, however, with termite, Ambrosia beetle, mountain pine beetle, and southern pine beetle samples all yielding composite genomes > 4 Mbp. The ANI value calculated between these genomes was > 93% in all cases except that of the southern pine beetle sample, which showed ~ 70% identity to all other genomes. Interestingly, the composite genomes reconstructed from the termite and mountain

pine beetle samples were observed to have > 98% ANI. Accounting for errors that would be expected to artificially reduce the ANI between these genomes, such as sequencing error and the potential inclusion of spurious contigs in the composite genomes during reconstruction, these *Rahnella* populations display an extremely high degree of similarity at the nucleic acid level.

We were also able to reconstruct 16 composite genomes from the genus *Pseudomonas* that we divided into 10 *P. fluorescens*, 4 *P. putida*, and 2 *P. syringae* groups based on their ANI to reference genomes (Figure 4.4, Table 4.3). The *P. putida* groups were only found in fungus-growing ant samples, while the *P. fluorescens* groups were only identified in the Ambrosia beetle or pine beetle samples, suggesting a degree of specificity in these interactions. The *P. syringae* genomes were reconstructed from the termite worker and mountain pine beetle samples. As with *Enterobacter* and *Rahnella*, *Pseudomonas* composite genomes that were placed in the same group displayed a high ANI to other each other (> 90%).

### 4.4.4 Functional Comparisons of the Metagenomes

To provide insight into the similarity of the function of the fungus-associated insect microbiomes we conducted COG and Pfam annotations of the associated metagenomes and compared the resulting functional profiles to those of 57 publicly available metagenomes generated from various host-associated and environmental samples (Figure 4.5, Table 4.4)). Consistent with their high phylogenetic similarity, the fungus-associated metagenomes clustered together in PCA plots of both functional metrics. Gut-associated metagenomes also seemed to cluster together, as did those generated from non-host associated environments. The foregut and midgut samples of *Panchlora* cockroaches, which feed specifically on the waste fungus garden material of leaf-cutter ants, also had similar functional profiles to fungus-associated insect

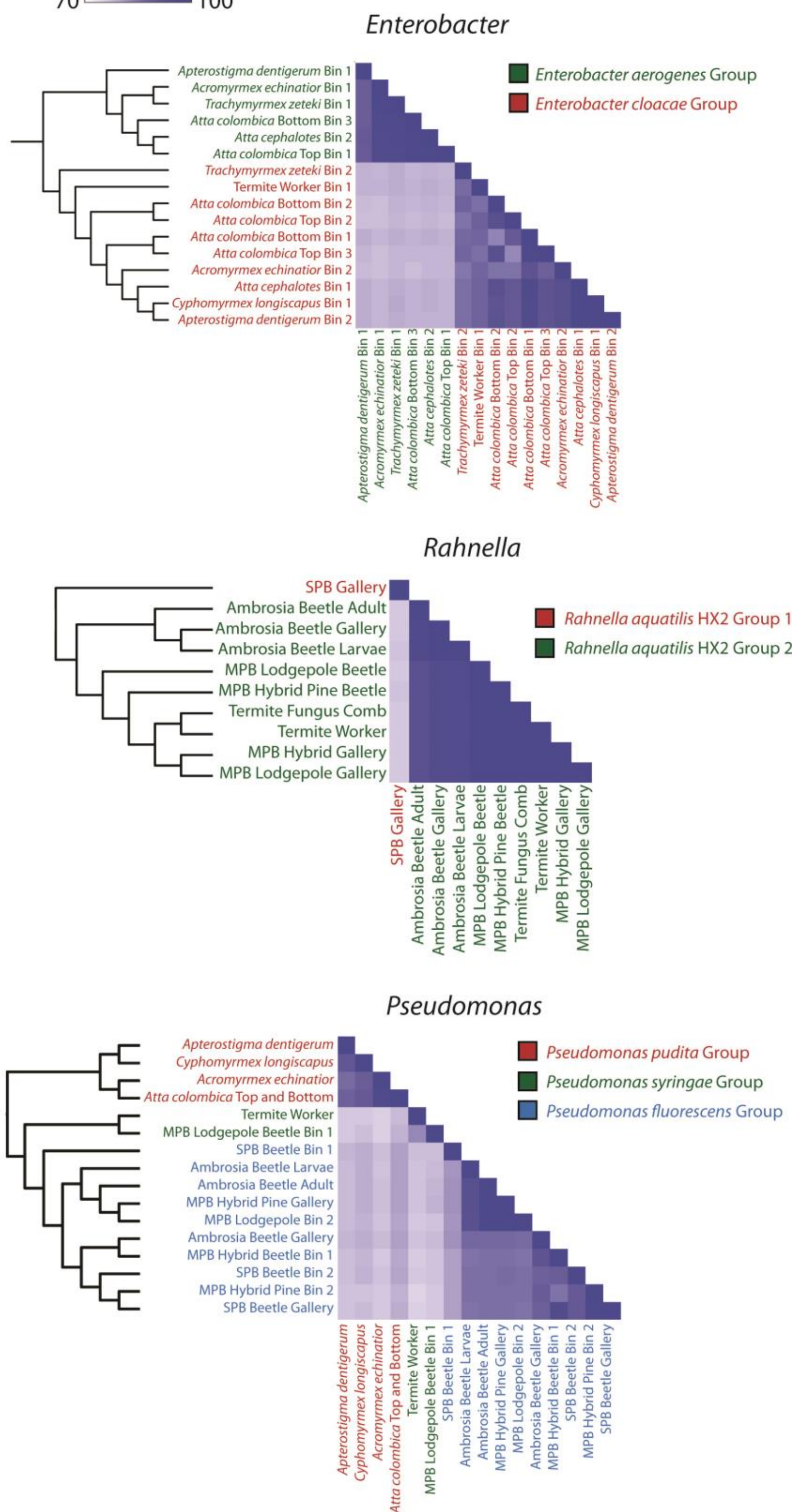



**Figure 4.4**. Average nucleotide identity (ANI) values for the composite *Enterobacter*, *Rahnella*, and *Pseudomonas* genomes reconstructed in this study. Composite genomes are colored according to their best ANI match in the sequenced genome collection. Values were calculated from pairwise best-BLASTN hits (parameters -X 150, -q -1, -F F). Dendrograms on the left-hand side represent neighbor-joining clustering performed on Pearson's r correlation of the ANI values.

microbiomes. A metagenome generated from the gut of the honey bee also clustered near the fungus-associated insect metagenomes (Figure 4.5).

### 4.5 Discussion

Our metagenomic analysis of the microbiomes of fungus-growing insects revealed that these hosts all associate with simple communities of bacteria belonging to the family *Enterobacteriaceae* and genus *Pseudomonas*. The microbiomes of Ambrosia beetles, fungus-growing termites, mountain pine beetles, and southern pine beetles in particular are comprised primarily of *Rahnella* and *Pseudomonas* populations that are strikingly similar at the whole-genome level. The microbiomes of fungus-growing ants are all markedly similar to each other and differ slightly but consistently from the microbiomes of the other insects analyzed, primarily in that different groups of *Pseudomonas* and *Enterobacteriaceae* predominated in these samples. Overall, the high degree of phylogenetic and functional similarity of all of the microbiomes analyzed demonstrates that convergent forces are the overriding factor that determines the structure of fungus-growing insect microbiota. Importantly, this work demonstrates that host niche and lifestyle can lead to a high degree of phylogenetic and functional convergence regardless of host phylogeny.

The fact that independent origins of fungus growing behavior in insects all converged on a similar bacterial community is not readily attributable to factors other than similarities in the fungus-growing lifestyles of the hosts. For example, the plant substrate harvested by these insects or surrounding environmental conditions are unlikely explanations of this pattern of convergence, as the samples analyzed in this study were collectively obtained from three cont-

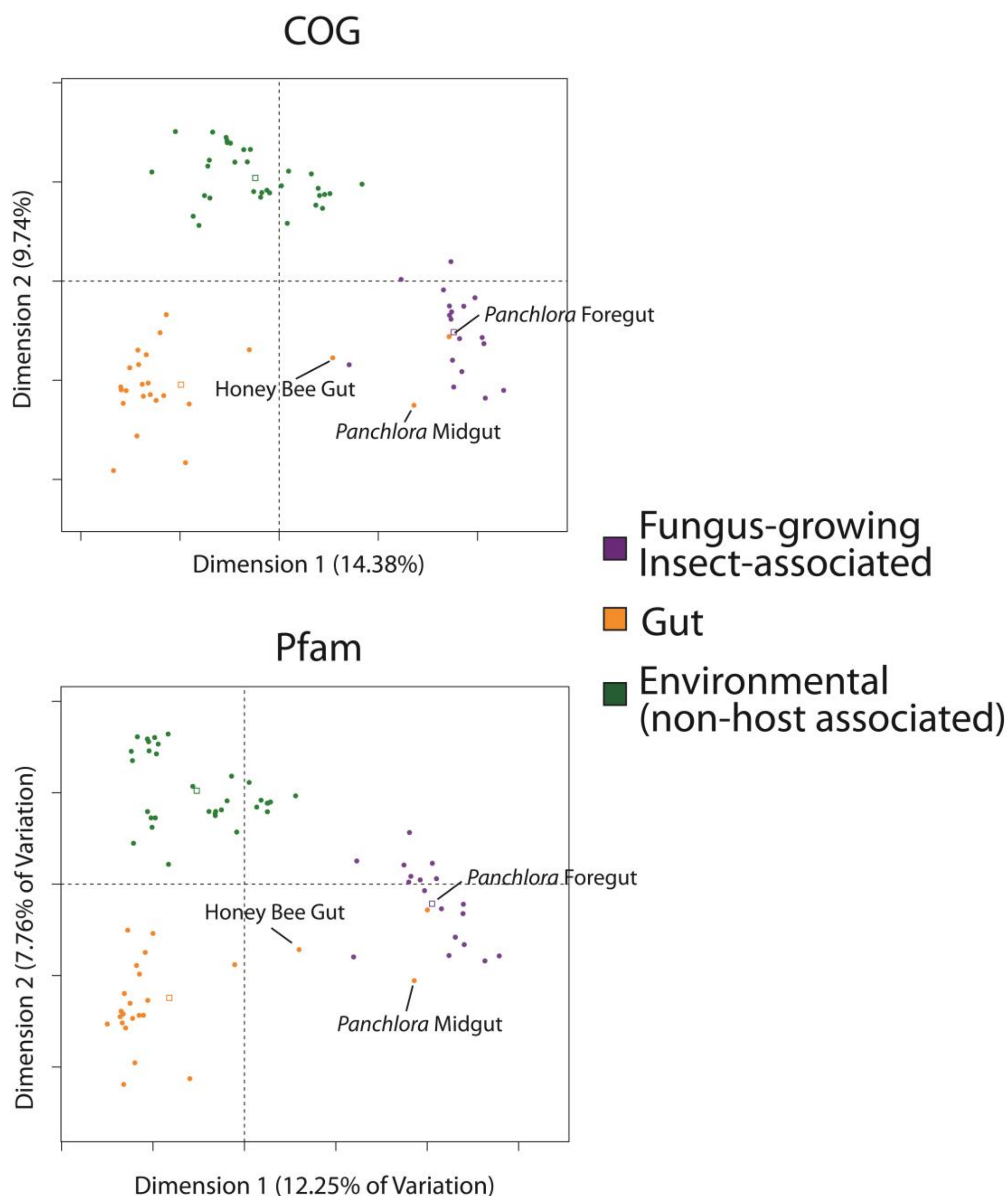



**Figure 4.5**. Principle Component Analyses of the functional potential of metagenomes representing 75 habitats, including the 18 fungus-growing insect samples, 33 non-host associated samples, and 24 gut-associated samples. Boxes represent the category averages. The PCA plot was created from a matrix of COG profiles normalized across metagenome sizes. The R module FactoMineR was used for calculations. The honey bee gut sample and the foregut and midgut samples of the *Panchlora* cockroach, an insect which feeds on the waste material of leaf-cutter ants, are given. Table 4.4 provides details of the metagenomes used.

-inents encompassing tropical and temperate ecosystems (Figure 4.1). Likewise, the large diversity of fungi collectively grown for food by these insects, encompassing gardens dominated by basidiomycete monocultures (ants, termites), mixed cultures of a primarily ascomycete cultivars (Ambrosia beetles), and multi-fungal assemblages (pine beetles) (17), suggests that the fungal symbionts are not a major determinant of bacterial community structure. Additionally, the microbiome of fungus-growing ant *Apterostigma dentigerum* was highly similar to the other ants sampled here despite the fact that this species is known to cultivate a pterulaceous fungus distantly related to the lepiotaceous fungi cultivated by the other ant species (38). Indeed, the high degree of similarity between the *Rahnella* and *Pseudomonas* groups identified in the termite, Ambrosia beetle, and pine beetles analyzed suggest that bacteria residing in fungus gardens may exhibit more stringent specificity for the host insects than the obligate fungal cultivars themselves.

The view that the fungus-growing lifestyle has independently transformed the microbiota of these insects is also supported by the high degree of similarity between the microbiomes of whole insects and their associated fungus garden microbiomes. Indeed, in fungus growing insects it appears that the microbiomes of the entire insects mirrors that of their fungus gardens. Most microbes associated with non-fungus growing insects generally reside in the gut and comprise a diversity of lineages in the phyla Firmicutes, Bacteroidetes, and Proteobacteria (39, 40), which is in stark contrast to that observed in fungus-growing insects. Because fungus-growing insects derive nourishment directly from their fungal cultivars and likely do not require the biodegradative or biosynthetic capabilities of gut microbes, the function of the microbes that

**Table 4.4**. Details regarding publicly-available metagenomes used for comparison in this study.

| Metagenome Sample | Environment Type | Number of Samples | IMG ID | MG-RAST ID |
|---|---|---|---|---|
| Canine Gut | Gut | 2 | 2019105001, 2019105002 | |
| Cockroach Gut | Gut | 3 | 2228664000, 3300000059, 3300000096 | |
| Cow Rumen | Gut | 1 | NA* | |
| Hoatzin Gut | Gut | 3 | 2088090016, 2088090036, 2100351002 | |
| Honey Bee Gut | Gut | 1 | 2166559020 | |
| Human Gut (Gill *et al* Study) | Gut | 2 | 2004002000, 2004002001 | |
| Human Gut (Kurokawa *et. al* Study) | Gut | 4 | | 4440939.3, 4440941.3, 4440945.3, 4440950.3 |
| Mouse Gut | Gut | 3 | 2004230001, 2004230004, 2004230000 | |
| Panda Gut | Gut | 3 | 2222084012, 2222084013, 2222084014 | |
| Termite Hindgut | Gut | 1 | 2004080001 | |
| Wallaby Foregut | Gut | 1 | 2021593001 | |
| Acid Mine Drainage | Non-host associated | 1 | 2001200000 | |
| Activated Sludge | Non-host associated | 2 | 2000000000, 2000000001 | |
| Antarctic Bacterioplankton | Non-host associated | 2 | 2008193000, 2008193001 | |
| Bath Hot Springs | Non-host associated | 1 | 2007309001 | |
| Bog | Non-host associated | 1 | 2124908038 | |
| Contaminated Groundwater | Non-host associated | 2 | 2007427000, 2006543007 | |
| Global Ocean Sampling Metagenomes | Non-host associated | 4 | | 4441126.3, 4441135.3, 4441152.3, 4441607.3 |
| HOT Samples | Non-host associated | 7 | | 4441041.3, 4441051.3, 4441055.3, 4441056.3, 4441057.3, 4441057.4, 4441062.3 |
| Lake Huron Sinkhole | Non-host associated | 2 | 2049941002, 2065487012 | |
| Lake Washington Sediment | Non-host associated | 4 | 2006207000, 2006207001, 2006207002, 2006207003 | |
| Minnesota Soil | Non-host associated | 1 | 2001200001 | |
| Permafrost | Non-host associated | 1 | 2124908043 | |
| Singapore Air | Non-host associated | 2 | 2003000006, 2003000007 | |
| Whalefall | Non-host associated | 3 | 2001200002, 2001200003, 2001200004 | |

* The data for this metagenome was downloaded from JGI via an alternative FTP: http://portal.nersc.gov/project/jgimg/CowRumenRawData/submission/

would typically reside in this environment appears to have been effectively externalized as fungus gardens. The initial symbiotic transition to a fungus-growing lifestyle thus appears to have drastically altered the composition of the entire microbiome of these insects, ultimately driving them towards a common structure.

Despite the overall pattern of convergence between the microbiomes of different fungus-growing insect lineages, a degree of host specificity was observed within the fungus-growing ants. In this group, the life history of the ants varies considerably; while leaf-cutter ants forage almost exclusively on fresh foliar material and can occupy colonies with > 1 million workers, nests of *Apterostigma* and *Cyphomyrmex* species contain ~50 workers and sometimes use dead plant material and insect frass as manure for their fungus gardens (41, 42). These considerable differences in host niche would be expected to lead to different bacterial assemblages if convergent forces were also responsible for driving fine-scale changes in microbiome structure. That similar bacterial groups inhabit the fungus gardens of all of these ants suggests that, although convergent forces appear to have driven all fungus-growing insect microbiomes towards a similar composition at broad scales, within each individual group phylogeny-driven factors determine fine-scale differences between these microbial communities. These conclusions are consistent with studies of the mammalian gut, where at large phylogenetic scales it has been proposed that diet is the key determinant of microbome structure (12, 13), while more fine-scale analyses of primate microbiomes have shown that host phylogeny is the overriding factor (11). Together with these studies, our work underscores the importance of taking the phylogenetic scale of the hosts into account when comparing microbiomes.

In addition to their similar phylogenetic compositions, the microbiomes of the fungus-growing insects also appear similar when analyzed at the functional level. Our comparison of these microbiomes with gut-associated and environmental microbial communities demonstrates that the fungus-growing insect microbiomes are distinct from previously-studied systems (Figure 4.4). Three notable exceptions to this are the honey bee gut and the foregut and midgut of the *Panchlora* cockroach, all of which clustered near the fungus-growing insect microbiomes in our functional comparisons. Given that *Panchlora* cockroaches are known to feed primarily on the fungus garden refuse of leaf-cutter ants (43), it is perhaps not surprising that the foregut and midgut of this insect displays similar phylogenetic and functional profiles to the fungus gardens themselves. The hindgut of this insect is most similar to other gut-associated microbiomes, however, and it is not known if the microbes in the foregut and midgut are viable or merely represent partially-degraded food material. Regardless, the foregut and midgut microbial communities do not represent true functional convergence, but are rather a consequence of overlapping ecologies of the hosts.

The similar functional composition of the honey bee gut and the fungus-growing insect microbiomes is less easily explained, however, as the honey bee gut is known to be composed of relatively few phylogenetic groups distantly related to any bacteria associated with fungus-growing insects (15). It is interesting to note that fungus-growing ants, Ambrosia beetles, and fungus-growing termites are all eusocial, as are honey bees. Although pine beetles do not have this social structure, it has been postulated that eusociality may promote fungus-growing behavior, as some level of division of labor is often required for insects to effectively manage fungus gardens (17). Pine beetles may thus represent a transition stage between fungus-

associated insects and true agriculturalists. In this case, the similarity between the functional profiles of the honey bee gut and the fungus-growing insect microbiomes may be a result of functional (but not phylogenetic) convergence of social insect microbiomes. Why sociality would result in this convergence is unknown, although it has been postulated that the management of microbiome structure is more important in eusocial insects because disease can spread more easily in densely packed insect communities (17). The different hygienic practices of these insects likely play an important role in excluding harmful microbes, thus potentially selecting for common functional potential in their microbiomes.

As the examination of the microbiomes of metazoans becomes more common, an important consideration in assessing the forces responsible for microbiome structure is the scale at which different comparisons are performed. In this study, we found that by comparing insect species with similar life histories from different orders we were able to reveal a distinct pattern of microbiome convergence. By examining more narrow phylogenetic scales in fungus-growing ants, however, we were also able to identify a host-specific pattern. It is unlikely that these general patterns are unique to fungus-growing insects; rather, it seems likely that both niche and phylogeny-driven forces can be broadly attributed to the microbiome dynamics of metazoans, with the differences in the scale at which comparisons are performed determining the relative influence of these two forces.

### 4.6 References

1. **Moran, NA.** 2006. Symbiosis. *Curr. Biol.* **16:**R866.

2. **Klepzig, KD, Adams, AS, Handelsman, J, and Raffa, KF.** 2009. Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environ. Entomol.* **38:**67-77.

3. **Douglas, AE.** 2010. The symbiotic habit. Princeton University Press, Princeton, New Jersey, USA.

4. **Ichinohe, T, Pang, IK, Kumamoto, Y, Peaper, DR, Ho, JH, Murray, TS, and Iwasaki, A.** 2011. Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl. Acad. Sci. USA* **108:**5354-9.

5. **Warnecke, F, Luginbuhl, P, Ivanova, N, Ghassemian, M, Richardson, TH, Stege, JT, Cayouette, M, McHardy, AC, Djordjevic, G, Aboushadi, N, Sorek, R, Tringe, SG, Podar, M, Martin, HG, Kunin, V, Dalevi, D, Madejska, J, Kirton, E, Platt, D, Szeto, E, Salamov, A, Barry, K, Mikhailova, N, Kyrpides, NC, Matson, EG, Ottesen, EA, Zhang, X, Hernandez, M, Murillo, C, Acosta, LG, Rigoutsos, I, Tamayo, G, Green, BD, Chang, C, Rubin, EM, Mathur, EJ, Robertson, DE, Hugenholtz, P, and Leadbetter, JR.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:**560-565.

6. **Hess, M, Sczyrba, A, Egan, R, Kim, TW, Chokhawala, H, Schroth, G, Luo, S, Clark, DS, Chen, F, Zhang, T, Mackie, RI, Pennacchio, LA, Tringe, SG, Visel, A, Woyke, T, Wang, Z, and Rubin, EM.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331:**463-467.

7. **Turnbaugh, PJ, Ley, RE, Mahowald, MA, Magrini, V, Mardis, ER, and Gordon, JI.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444:**1027-31.

8. **Brucker, RM, and Bordenstein, SR.** 2012. Speciation by symbiosis. *Trends Ecol. Evol.* **27:**443-51.

9. **Brucker, RM, and Bordenstein, SR.** 2012. The roles of host evolutionary relationships (genus: *Nasonia*) and development in structuring microbial communities. *Evolution* **66:**349-62.

10. **Ley, RE, Lozupone, CA, Hamady, M, Knight, R, and Gordon, JI.** 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* **6:**776-88.

11. **Ochman, H, Worobey, M, Kuo, CH, Ndjango, JB, Peeters, M, Hahn, BH, and Hugenholtz, P.** 2010. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol.* **8:**e1000546.

12. **Muegge, BD, Kuczynski, J, Knights, D, Clemente, JC, Gonzalez, A, Fontana, L, Henrissat, B, Knight, R, and Gordon, JI.** 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332:**970-4.

13. **Ley, RE, Hamady, M, Lozupone, C, Turnbaugh, PJ, Ramey, RR, Bircher, JS, Schlegel, ML, Tucker, TA, Schrenzel, MD, Knight, R, and Gordon, JI.** 2008. Evolution of mammals and their gut microbes. *Science* **320:**1647-51.

14. **Walter, J, Britton, RA, and Roos, S.** 2011. Host-microbial symbiosis in the vertebrate gastrointestinal tract and the *Lactobacillus reuteri* paradigm. *Proc. Natl. Acad. Sci. USA* **108 Suppl 1:**4645-52.

15. **Engel, P, Martinson, VG, and Moran, NA.** 2012. Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. USA* **109:**11002-7.

16. **Hongoh, Y, Deevong, P, Inoue, T, Moriya, S, Trakulnaleamsai, S, Ohkuma, M, Vongkaluang, C, Noparatnaraporn, N, and Kudo, T.** 2005. Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl. Environ. Microbiol.* **71:**6590-9.

17. **Mueller, U, Gerardo, N, Aanen, D, Six, D, and Schultz, T.** 2005. The evolution of agriculture in insects. *Annu. Rev. Ecol. Evol. Syst.* **36:**563-595.

18. **Six, DL, and Klepzig, KD.** 2004. Dendroctonus bark beetles as model systems for the study of symbioses. *Symbiosis* **37:**207-232.

19. **Adams, AS, and Six, DL.** 2007. Temporal variation in mycophagy and prevalence of fungi associated with developmental stages of Dendroctonus ponderosae (Coleoptera: Curculionidae). *Environ. Entomol.* **36:**64-72.

20. **Harrington, T.** 2005. Ecology and Evolution of Mycophagous Bark Beetles and Their Fungal Partners, p. 257-291. *In* F. Vega and M. Blackwell (ed.), Insect-Fungal Associations: Ecology and Evoluton. Oxford University Press, New York, NY.

21. **Adams, AS, Aylward, FO, Adams, SM, Erbilgin, N, Aukema, BH, Currie, CR, Suen, G, and Raffa, KF.** 2013. Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism. *Appl. Environ. Microbiol.* **79:**3468-75.

22. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **6:**1688–1701.

23. **Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, YJ, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, ML, Jarvie, TP, Jirage, KB, Kim, JB, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, McDade, KE, McKenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth,**

**GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A, Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF, and Rothberg, JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-80.

24. **Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:**403-410.

25. **Brady, A, and Salzberg, SL.** 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat. Methods* **6:**673-6.

26. **Albertsen, M, Hugenholtz, P, Skarshewski, A, Nielsen, KL, Tyson, GW, and Nielsen, PH.** 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31:**533-8.

27. **Konstantinidis, KT, and Tiedje, JM.** 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. USA* **102:**2567-72.

28. **Hyatt, D, Chen, GL, Locascio, PF, Land, ML, Larimer, FW, and Hauser, LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:**119.

29. **Santos, SR, and Ochman, H.** 2004. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ. Microbiol.* **6:**754-9.

30. **Cole, JR, Wang, Q, Cardenas, E, Fish, J, Chai, B, Farris, RJ, Kulam-Syed-Mohideen, AS, McGarrell, DM, Marsh, T, Garrity, GM, and Tiedje, JM.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37:**D141-145.

31. **Price, MN, Dehal, PS, and Arkin, AP.** 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26:**1641-50.

32. **Anisimova, M, and Gascuel, O.** 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst. Biol.* **55:**539-52.

33. **Markowitz, VM, Chen, IM, Chu, K, Szeto, E, Palaniappan, K, Grechkin, Y, Ratner, A, Jacob, B, Pati, A, Huntemann, M, Liolios, K, Pagani, I, Anderson, I, Mavromatis, K, Ivanova, NN, and Kyrpides, NC.** 2012. IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res.* **40:**D123-9.

34. **Glass, EM, Wilkening, J, Wilke, A, Antonopoulos, D, and Meyer, F.** 2010. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb. Protoc.* **2010:**pdb prot5368.

35. **Tatusov, RL, Fedorova, ND, Jackson, JD, Jacobs, AR, Kiryutin, B, Koonin, EV, Krylov, DM, Mazumder, R, Mekhedov, SL, Nikolskaya, AN, Rao, BS, Smirnov, S, Sverdlov, AV, Vasudevan, S, Wolf, YI, Yin, JJ, and Natale, DA.** 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:**41.

36. **Finn, RD, Tate, J, Mistry, J, Coggill, PC, Sammut, SJ, Hotz, HR, Ceric, G, Forslund, K, Eddy, SR, Sonnhammer, EL, and Bateman, A.** 2008. The Pfam protein families database. *Nucleic Acids Res.* **36:**D281-8.

37. **Altschul, SF, Madden, TL, Schaffer, AA, Zhang, J, Zhang, Z, Miller, W, and Lipman, DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:**3389-402.

38. **Villesen, P, Mueller, UG, Schultz, TR, Adams, RM, and Bouck, AC.** 2004. Evolution of ant-cultivar specialization and cultivar switching in Apterostigma fungus-growing ants. *Evolution* **58:**2252-65.

39. **Engel, P, and Moran, NA.** 2013. The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.*

40. **Colman, DR, Toolson, EC, and Takacs-Vesbach, CD.** 2012. Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* **21:**5124-37.

41. **Hölldobler, B, and Wilson, EO.** 2010. The leafcutter ants: civilization by instinct. W. W. Norton and Company, Inc., New York, NY.

42. **Weber, NA.** 1966. Fungus-growing ants. *Science* **153:**587-604.

43. **Gontang, EA, Aylward, FO, Glavina del Rio, T, Chovatia, M, Fern, A, Lo, C, Malfatti, S, Tringe, S, Currie, CR, and Kolter, R.** In preparation. Major shifts in microbial diversity and community composition across gut sections of a tropical cockroach.

**Summary and Future Outlook**

In this dissertation I have explored questions regarding the composition and function of the symbiotic microbial communities of fungus-growing insects, with particular emphasis on those of leaf-cutter ants. In Chapter 1 I provided a review of previous work on leaf-cutter ant fungus gardens and presented evidence that the symbiotic gardens cultivated by these insects represent a specialized community of microbes rather than merely a pure culture of the dominant fungal cultivar, *Leucoagaricus gongylophorus*. I explored this concept in more detail in Chapter 2, in which, using metagenomic and metaproteomic approaches, I presented a phylogenetic and functional characterization of bacterial communities residing in leaf-cutter ant fungus gardens. The results of this work provided evidence that bacteria in these ecosystems do not have the capacity for the degradation of plant biomass, and that they instead likely play a role in nutrient biosynthesis that potentially enhances the lignocellulolytic capacity of *L. gongylophorus*. To pursue the question of how plant biomass is degraded in fungus gardens in more detail, I subsequently conducted genomic and metaproteomic investigations of fungus gardens with an emphasis on *L. gongylophorus*, which is presented in Chapter 3. This work provides evidence that *L. gongylphorus* is the dominant driver of plant biomass degradation in fungus gardens, and further suggests that bacteria play ancillary roles in nutrient provisioning. Lastly, in Chapter 4, I presented a comparative metagenomic study of the microbial communities associated with fungus-growing ants, fungus-growing termites, Ambrosia beetles, mountain pine beetles, and southern pine beetles. This work establishes that the microbial communities of these insects have remarkably similar phylogenetic and functional compositions, suggesting the factors independent of host phylogeny have resulted in community-level convergence. Together with the analyses presented in earlier chapters, this suggests that the general principles underlying the leaf-cutter

ant fungus garden community, whereby the dominant fungal partner degrades the majority of plant biomass and bacteria play ancillary roles, may be broadly applicable to other fungus-growing insect systems as well.

Future experiments elucidating a more precise role for bacteria in the fungus gardens of leaf-cutter ants will be invaluable. Although I have conjectured in this dissertation that bacteria may somehow promote the lignocellulolytic capacity of *L. gongylophorus* in fungus gardens, there is no direct evidence supporting this assertion. *L. gongylophorus* and the dominant bacterial constituents of fungus gardens are all culturable, however, and future bioassay experiments analyzing individual fungus-bacteria pairings on artificial media may provide insight into the specific symbiotic dynamics that take place in fungus gardens. Other methods, such as infusing foliar biomass with antibacterials before feeding it to a colony of leaf-cutter ants, may provide evidence of a potential fitness benefit of bacteria to leaf-cutter ant colonies.

Another area of research deserving of more attention concerns the bacterial communities of fungus growing termites, Ambrosia beetles, and pine beetles. The remarkable similarity between bacterial communities associating with these insects suggests there is a high degree of specificity in these interactions. Are the compositions of these communities directly controlled by the insect hosts, or are they somehow selected for by the fungus-growing niche? What mechanism of dispersal could account for the high degree of similarity between *Pseudomonas* and *Rahnella* species that associate with insects that have non-overlapping ranges? Future investigations in which the specific host-microbe dynamics of these communities are probed in more detail will be necessary to answer these questions.

The challenges of analyzing the dynamics of microbial communities are not limited to fungus-growing insect systems, however. Indeed, a major challenge of microbiology in the future will be to understand how communities of microbes interact to produce systems-level phenomena such as harmful algal blooms, disease states in the human alimentary tract, or resilience to a changing climate. Numerous new technologies and bioinformatic techniques will be invaluable in disentangling the complex dynamics of microbial communities, but clever experimentation and collaboration between disparate scientific fields will also be necessary. It is my belief that this "complex systems" approach to biology will bring us one step closer to understanding the enormous diversity of microbial life on this planet as well as how it continues to shape the biosphere in which we live.

## Appendix 1: Supplementary Material for Chapter 2

**Previously Published as Supplementary Material**: Frank O Aylward, Kristin E Burnum, Jarrod J Scott, Garret Suen, Susannah G Tringe, Sandra M Adams, Kerrie W Barry, Carrie D Nicora, Paul D Piehowski, Samuel O Purvine, Gabriel J Starrett, Lynne A Goodwin, Richard D Smith, Mary S Lipton, Cameron R Currie (2012). Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME Journal*, 6: 1688-1701.

**Author Contributions are the same as in Chapter 2.**

### A1.1: 16S Pyrotags

16S pyrotag libraries were constructed and sequenced as previously described (1) from the same DNA samples used to generate the community metagenomes. All 16S pyrotags were quality checked, aligned, chimera checked, and clustered into 97% identity Operational Taxonomic Units (OTUs) using the program Pyrotagger (2). The length of all pyrotags was trimmed to 250bp using the online Pyrotagger interface before further processing. Representatives from each OTU were input into the Ribosomal Database Project's online Seqmatch tool (http://rdp.cme.msu.edu/seqmatch (3)) for classification. Only those matches having the highest Seqmatch score were used. All pyrotags can be found in the NCBI Short Read Archive under accession SRA036606.

### A1.2: Metagenome Sequencing, Assembly, and Annotaton

The three fungus garden samples analyzed were subjected to standard 454-titanium library construction. For both the *A. colombica* garden top and bottom samples, only 500 ng DNA were used for library construction, which did not initially yield sufficient stock for sequencing. For these samples, the bead-immobilized DNA was subjected to an additional 8 rounds of PCR with adapter primers to generate sufficient material for sequencing. One single plate of a Roche 454 FLX GS Titanium pyrosequencer (4) was used to sequence each library,

yielding between 382-441 Mbp of sequence data per run. Each run was assembled with Newbler version 2.3, with overlap requirements of minimum length 40 and minimum identity 95%, yielding between 26-46 Mb assembled contigs and between 16-52 Mb unassembled singletons for each community metagenome. Assembly was performed individually on each metagenome using the program Newbler (v. 2.3) with default parameters (minimum identity of 95%, minimum length of 40 bp). Sequencing statistics for the community metagenomes are presented in Table 2.1. Open reading frame (ORF) prediction for all metagenomes was done using a combination of MetaGene (5) and BLASTX (6). ORFs were then annotated via the IMG/ER pipeline (7) and loaded into IMG/M (http://img.jgi.doe.gov/cgi-bin/m/main.cgi (8)). These protein databases, which include predictions made from potentially arthropod, plant, or fungal sequences (Table 2.3), were subsequently used in all metaproteomic spectral mapping (see below for details). These databases included both partial- and full length protein predictions. The *Atta cephalotes*, *Atta colombica* top, and *Atta colombica* bottom community metagenomes can be found in GenBank under the accessions AEWW00000000, AGFS00000000, AGFT00000000.

**A1.3: Taxonomic Binning**

Trimmed and quality-checked pyrosequencing reads from the community metagenomes were analyzed using the Genome Abundance and Average Size (GAAS) program (9) for investigation of relative microbial abundances. All complete bacterial and archael genomes available on NCBI (accessed 03/01/2010) were used for classification. Default parameters were used (-p = 50, -q = 50) and data were parsed from the output using custom Perl scripts. A complementary analysis was also performed by comparing all contigs and singletons in the community metagenomes to NCBI's non-redundant nucleoltide (nt) database (accessed 11/01/2010) using BLASTN (e-value of 1e-5), using only the best hit for classification (Table

2.2). PhymmBL v. 1.0 (10) was also used for taxonomic classification of contigs and singletons in the community metagenomes, and the phylogenetic bins created using this program were subsequently used for all metabolic pathway analyses. Because the accuracy of phylogenetic binning is greatly increased when reference genomes of organisms closely related to microbes sequenced in the metagenomic data are used for classification, we supplemented the PhymmBL dataset with the draft genomes of three bacteria we isolated from the fungus gardens of leaf-cutter ants. Two of these isolates, *Serratia* FGI 94 and *Enterobacter* FGI 35, were cultured from *Atta colombica* fungus garden samples. The third isolate, *Pseudomonas* FGI 182, was isolated from an *Atta cephalotes* fungus garden sample. No eukaryotic or viral sequences were present in the PhymmBL classification database, and phylogenetic binning performed using this approach is presented only for bacteria.

**A1.4: Draft Genomes Used for Phylogenetic Binning**

All cultures were originally isolated on media containing a basal salts solution used in (11) supplemented with either 5% w/v carboxymethylcellulose (EMD Biosciences, 441 Charmany Dr. Madison WI, USA), Acetate (BD Biosciences, Becton Drive, Franklin Lakes, NJ, *USA*), or Pectin (Sigma-Aldrich, 3050 Spruce St., St. Louis, MO, USA). DNA was extracted from all cultures using a Qiagen DNeasy Blood and Tissue Kit (Qiagen Sciences, Germantown, MD, USA). Draft genomes were sequenced at the Department of Energy Joint Genome Institute (DOE-JGI) using 454 pyrosequencing. Average read lengths for FGI94, FGI35, and FGI182 were 383.3 bp, 359 bp, and 423bp, respectively. General information concerning library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. Draft assemblies were created using the program Newbler v. 2.3 with default parameters. For *Serratia sp.* FGI94, 65.3 Mbp of 454 data were used to assemble 95 contigs: For *Enterobacter sp.* FGI35,

50.5 Mbp of 454 data were used to assemble 47 contigs: For *Pseudomonas sp.* FGI182, 108.7 Mb of 454 data were used to assemble 649 contigs, respectively. Only 33 contigs larger than 1 Kb were used in the fragment recruitment analysis of *Enterobacter* FGI 35.

### A1.5: Comparison of *Enterobacter* Populations between Metagenomes

*Enterobacter* populations were compared between metagenomes by aligning all genes on contigs classified as *Enterobacter* FGI 35 to the draft FGI 35 genome. This procedure is based on methods used for fragment recruitment analyses published elsewhere (12). Genes were compared using BLASTN with parameters -q 1, -X 150, and –F, and all hits less than 50 bp long were discarded. For functional annotation of FGI 35, genes were predicted using GeneMark (13) and annotations were assigned using the KEGG and COG databases (see below). The recruitment plot was visualized using the programming language R. Conserved and variable regions of recruitment were determined through manual inspection.

### A1.6: KEGG, CAZy, and COG Analyses

Proteins predicted from phylogenetic bins representing dominant community members were analyzed using the KEGG Automatic Annotation Server (KAAS, http://www.genome.jp/tools/kaas, SDH method, Accessed: 06/15/2010, (14)) to obtain metabolic pathway predictions. The resulting data was then extracted and used to construct a matrix in which each column corresponded to a KEGG Orthology (KO) entry and each row corresponded to a phylogenetic bin. Clustering analysis was performed similar to in (15). Briefly, a similarity matrix of the bins was constructed using custom PERL scripts implementing Spearman's rank correlation, and this was subsequently used to construct a UPGMA tree using Phylip (16). The tree file was visualized using the Interactive Tree of Life (http://itol.embl.de/) (17) web server.

Percent coverage of each metabolic pathway in each bin was also calculated by dividing the total number of KO proteins in each pathway by the total number of non-redundant annotations recovered. This information was visualized using the matrix2png web application (http://bioinformatics.ubc.ca/matrix2png, (18)).

Carbohydrate Active Enzymes (CAZymes (19)) were predicted from all proteins in the three community metagenomes as previously described (1). Protein predictions from the termite hindgut (20) and *Macropus eugenii* (Wallaby) foregut (21) metagenomes were downloaded from IMG/M and processed in the same manner as all fungus garden protein predictions.

All Clusters of Orthologous Groups (COGs, (22)) used in this study were annotated via the IMG/ER pipeline (7) and downloaded directly from IMG/M (08/01/2010). For the three fungus garden metagenomes, the percentage representation of each COG category was calculated. The COG profiles for each of the three fungus garden metagenomes were them combined, as were all other available metagenomes on IMG/M. Fisher's exact test ($p < 0.05$) was used to calculate enrichment of COG categories in the fungus garden metagenomes relative to the dataset comprising COGs of all other metagenomes. The same analysis was then done on individual COGs from each category found to be enriched.

**A1.7: Sample Collection for Metaproteomics**

Fungus garden samples used for metaproteomic analysis were distinct from the samples used for community metagenomic investigations. Fungus garden material for proteomic analyses was collected from a nest of *A. colombica* near Gamboa, Panama, in June 2008. Combined top and bottom strata from this sample were combined and frozen at -20 °C within 20 minutes of collection. Also, fungus garden strata collected from a nest of *A. sexdens* maintained

in the laboratory of Cameron R. Currie (University of Wisconsin-Madison) was collected, pooled and immediately frozen for metaproteomic work. The laboratory-maintained nest was originally collected in Argentina in 2006 and kept in the laboratory on a diet of Oak (*Quercus*) and Maple (*Acer*) leaves.

**A1.8: Proteomic Sample Preparation - (Samples 01 [Lab] & 02 [Field])**

Each garden sample was weighed to the mass of 250 mg and added to the ram end of shredder barocycler pulse tubes (Pressure Biosciences Inc., South Easton, MA) and 100 mM ammonium bicarbonate ($NH_4HCO_3$), pH 8.0 buffer was added to the sample. The sample was shredded for five 1-minute bursts with cooling on ice using the PCT-Shredder Kit (PBI, MA). All samples were barocycled for 10 cycles (20 seconds at 35,000 psi back down to ambient pressure for 10 seconds). All of the material was removed from the pulse tube and a Coomassie Plus (Thermo Scientific, Rockford, IL) assay was used to determine protein concentration. The sample was added to a 15 mL Falcon tube with 8 M urea, 2 M thiourea, 5 mM dithiothreitol (DTT) ((*2S,3S*)-1,4-Bis-sulfanylbutane-2,3-diol) and 1% Chaps detergent (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate) (Sigma-Aldrich, St. Louis, MO). The sample was incubated at 60 °C for 30 minutes to reduce and denature the protein and centrifuged at 5000 x g to remove the debris. The supernatant was retained in a fresh 15 mL Falcon tube and the pellet was washed and spun 3 times with 100 mM $NH_4HCO_3$. All supernatants were pooled together and brought to a 10 times dilution of the original volume with the buffer. Tryptic digestion (Promega, Madison, WI) was performed at a 1:50 (w/w) ratio with the addition of 1 mM $CaCl_2$ to stabilize the trypsin and reduce autolysis. The sample was incubated for 3 hours and cleaned via strong cation exchange (SCX) solid phase extraction (SPE) (Supelco, Bellefonte, PA) and dried to ~200 µl and assayed with Bicinchoninic acid (BCA) (Thermo Scientific,

Rockford, IL) to determine the final peptide concentration. Each global garden peptide solution was analyzed in duplicate by capillary LC-MS/MS with a 100 minute separation on an LTQ Orbitrap.

**A1.9: Off-Line Sample Fractionation (Samples 01 [Lab] and 02 [Field])**

Tryptic peptides were resusupended in 900 mL of 10 mM ammonium formate (pH 3.0)/25% ACN and fractionated by SCX chromatography on a Polysulfoethyl 2.1 mm x 35 mm, 3 mm particle size column (PolyLC, Columbia, MD) that was preceded by a 2.1 mm, 10 mm guard column. The separations were performed using an Agilent 1100 series HPLC system at a flow rate of 200 mL/min, and mobile phases that consisted of 10 mM ammonium formate (pH 3.0)/25% ACN (A), and 500 mM ammonium formate (pH 6.8)/25% ACN (B). After loading sample onto the column, the mobile phase was maintained at 100% A for 10 min. Peptides were then separated by using a gradient from 0 to 50% B over 40 min, followed by a gradient of 50-100% B over 10 min. The mobile phase was then held at 100% B for 10 min. A total of 25 fractions were collected for each separation. Each fraction was dried under vacuum. The fractions were dissolved in 30 mL of 25 mM $NH_4HCO_3$ and 10 mL of each fraction were analyzed by capillary LC-MS/MS with a 100 minute separation on a LTQ ion trap mass spectrometer As each fraction was only analyzed once, 25 analyses were done for the lab sample 01 (fractions 1-25) and 25 analyses were done for the field sample (fractions 1-25)

**A1.20: Alternative Sample Preparation (Samples 05 [Lab], 06 [Field] & 09 [Field])**

Filter Aided Sample Preparation (FASP) (23) was used as an alternative sample preparation method. The same lysis method as previously described was applied with the exception of the lysis buffer which consisted of 4% SDS (w/v), 100 mM DTT in 100 mM

Tris/HCL, pH 7.6. After shredding and barocycling the sample was transferred to a 2 mL centrifuge tube and incubated at 95 °C for 5 minutes for thorough lysis and solubilization. The sample was allowed to cool and spun at 12,000 x g for 5 minutes to pellet debris and clarify the protein solution. Amicon Ultra-15 10K MWCO centrifuge devices (Millipore, Billerica, MA) were used for buffer exchange. The sample was added to the filter unit with 8 M urea, pH 8.5 at 15% of the total volume. The unit was centrifuged at 4000 x g for 40 minutes (until it reached the dead volume of 200 ul). The flow through was discarded and 2 mL of 0.05M iodoacetamide in 8 M urea, pH 8, was added and the sample and mixed at 600 rpm for 5 minutes in the dark. The unit was centrifuged at 4000 x g for 15 minutes to remove the iodoacetamide solution and 10 mL of 8 M urea, pH 8.0 was added. The unit was spun at 4000 x g for 40 minutes and this was repeated 3 times followed by 3 rinses with 100 mM $NH_4HCO_3$, pH 8.0. A BCA assay was used to determine the protein concentration and enough 25 mM $NH_4HCO_3$ was added to the filter unit to cover the filter. The flow-through collection tube was thoroughly cleaned and the sample was tryptically digested in a 1:50 (w/w) ratio with 1 mM CaCl2 and was allowed to incubate overnight at 37°C. The following day the much smaller peptide sample was centrifuged at 4000 x g for 30 minutes and the peptides were collected in the flow through. The filter was rinsed with 25 mM $NH_4HCO_3$, and the flow-through was pooled. The sample was dried to ~50 ul in a Speed vac and a BCA assay was performed for a final protein concentration. Each global garden peptide solution was analyzed in duplicate by capillary LC-MS/MS with a 100 minute separation on an LTQ Orbitrap. These samples were not fractionated.

**A1.21: Mass Spectrometry Analyses**

The unfractionated samples were analyzed by tandem MS (MS/MS) on an LTQ Orbitrap (Thermo Fisher Corporation, San Jose, CA). The fractionated samples were analyzed by tandem

MS on a LTQ (Thermo Fisher Corporation, San Jose, CA). Columns (60 cm long / 75 µm ID) were packed (in-house) with Jupiter 3 µm C18 resin (phenomenex; Torrance, California, USA). Liquid chromatography systems parameters: constant pressure of 10,000 PSI, aqueous to organic (0.1% FA in water - A, and 0.1% FA in 99.9% acetonitrile - B) exponential gradient over 100 minutes, and a 500 nL/min flow rate. MS parameters; capillary voltage 2.2 kV, Collision energy 35 eV, exclusion time 1 min, 10 most intense peaks selected for fragmentation (LTQ), 6 most intense peaks selected for fragmentation (LTQ Orbitrap).

**A1.22: Proteomic Data Analysis and Data Filtering**

Sequest analysis software (24) was used to match the MS/MS fragmentation spectra with peptides from the community metagenome predicted protein databases. These databases comprised all proteins annotated from the community metagenomes by the IMG/M-ER pipeline (see above for details on how these databases were created). The protein datasets for the *Atta colombica* Top, *Atta colombica* Bottom, and *Atta cephalotes* samples are publicly available and can be found on the IMG/M website with the taxon subject IDs 2029527005, 2029527006, 2029527004, respectively. Mass spectra were mapped onto these databases individually. Because we wished to focus on bacterial proteins in fungus gardens in this study, only proteins which mapped to contigs which could be confidently classified as bacterial were considered in downstream analyses. To achieve this, all contigs which contained proteins to which peptides could be mapped in our proteomic analyses were phylogenetically binned using PhymmBL v. 3.0. Contigs and reads with matches to bacteria and genus-level confidence scores of > 75% were considered to be bacterial. Peptides which mapped to proteins of interest were manually annotated, and spectra for these peptides can be found in Figure 2.5.

As stated previously in the literature, a key problem in computational proteomics is distinguishing between correct and false peptide identifications (25). This is even more of a problem in metaproteomic studies where there are confounding factors such as; large theoretical protein databases (based on metagenomic data), large percentage of non-unique peptides (functional redundancy) (26), and the unknown degree of genomic sequence representation in the samples. In order to increase both the amount and confidence of peptide identifications we first used MS-GF scoring to generate a spectra level confidence value (this software is available as an open source from http://www.cs.ucsd.edu/users/ppevzner/software.html.) (25). We filtered the spectra using a MS-GF cutoff of 1E-10 and then calculated our false discovery rate (FDR) utilizing decoy database searches (%FDR=((Reverse Identifications*2)/Total Identifications) *100). To achieve ≤ 5% spectral level FDR we varied Xcorr (Sequest's main scoring value related to peptide confidence) and DelCN (how good the XCorr is relative to the next best match DelCN = $(XCorr_1-XCorr_2)/XCorr_1$) values for spectra (group by peptide charge state and number of tryptic ends). In pure cultures MS-GF often eliminates the need of decoy database searches, but the complexity of these samples lead us to believe additional filtering was necessary. Likely, our spectral FDR overestimates the actual FDR of the data. The reverse database (decoy database) search we employed is the current accepted approach, but was developed for pure cultures (27), where the proteome present in the sample is represented in the genomic database. This is a key assumption to the accuracy of this approach. As applied to microbial communities, metagenomic databases do not completely represent the proteome actually present in the sample. This is especially concerning for a complex microbial community, like that associated with the leaf-cutter ant fungus gardens. Precursor ppm mass errors are also given for all high mass accuracy (LTQ Orbitrap) identifications. Because ppm mass error was not used as a filtering

metric, we did not eliminate systematic errors from parent ion mass measurements with software tools such as DtaRefinery (28). For high mass accuracy data low false discovery rates can be achieved by using ppm error. The data used in the analysis can requested at http://ober-proteomics.pnl.gov/.

**A1:23: Supplementary Discussion**

Metagenomics and metaproteomics provide contrasting and complementary data for describing a microbial community. While metagenomics provides a large quantity of data regarding the taxonomic composition and functional potential of a community, it lacks any information regarding the presence of proteins. Metaproteomics, conversely, provides a relatively small quantity of data that is highly useful in characterizing which proteins are being produced in a community. Together these approaches allow for the broad-scale characterization of a microbial community together with a smaller and more directed investigation of functions being carried out by specific groups.

The metagenomic and metaproteomic analysis of leaf-cutter ant fungus gardens presented here confidently identified bacterial proteins annotated to numerous metabolic pathways. As is typical of proteomic investigations focusing on bacteria, numerous housekeeping genes such as ribosomal proteins, RNA polymerase subunits, translation elongation factors, ATPase subunits, and chaperonins were characterized. In addition to these, our identification of numerous proteins potentially participating in amino acid biosynthesis, carbohydrate degradation and transport, and pantothenate biosynthesis indicates that bacteria, specifically *Enterobacteriaceae* and *Pseudomonadaceae*, may be carrying out these processes in fungus gardens (Table 2.5).

Due to the small number of bacterial peptides identified (359), we chose to focus our

metaproteomic analyses on those metabolic functions for which multiple proteins were identified. These included amino acid biosynthesis, nutrient transport, pantothenate and CoA biosynthesis, sulfur metabolism, and polysaccharide degradation (Figure 2.5). High-quality spectra with low ppm mass errors were identified for peptides corresponding to each of these processes. Moreover, our manual annotation of 38 peptides mapping to proteins with predicted functions in these processes confirms their presence in the samples analyzed. Although the identification of these peptides cannot confirm that microbial groups are participating in biomass degradation or nutrient cycling, it suggests numerous roles for bacteria in fungus gardens which warrant further investigation, especially considering the large and previously unidentified population of *Enterobacter* present in our data.

One challenge of using metaproteomics to target bacterial populations when using whole fungus gardens is that samples invariably contain large quantities of plant, fungal, and arthropod DNA, making the specific investigation of bacterial proteins difficult. Our decision to use whole fungus gardens for metaproteomic analyses, rather than just the bacterial fraction that was used in the construction of the community metagenomes, presented both benefits and challenges to downstream analyses. The benefits included a more accurate "snapshot" of bacterial processes in fungus gardens, as a lengthy differential centrifugation procedure (~30 min) may have altered the physiological state of the bacteria and removed any secreted or lignocellulose-adherent proteins. The challenges associated with this decision included the production of metaproteomic data which likely included a large quantity of mass spectra belonging to plant, fungal, and arthropod proteins in addition to those belonging to bacteria. Indeed, we identified 294 proteins that could not be confidently assigned to bacteria and are likely of eukaryotic origin. Moreover, multiple peptides and mass spectra could be assigned to some of these proteins, indicating that they may

be part of the abundant plant/fungal matrix of which fungus gardens are largely comprised. Thus, although the metaproteomic data presented here likely provides accurate information regarding the presence of bacterial proteins, it is limited in the total number of such proteins identified. A productive avenue of future research would be to use metaproteomics to investigate other microbial constituents of fungus gardens and broaden our understanding of overall community function.

Our metaproteomic analyses were performed on fungus garden samples from both field-collected and laboratory-reared leaf-cutter ant colonies. Of the total of 869 distinct peptides identified from our analyses, 129 were identified from both samples. A number of factors likely influence the degree of overlap. Firstly and most importantly, our recovery of 1186 total peptides grossly underestimates the diversity of proteins in these environments, and precludes any definite conclusions. Secondly, field-collected and laboratory-reared colonies live in drastically different environments and forage on different plant substrates, likely affecting the overall proteomic profile of their microbial communities. Thirdly, the field-collected sample originated from a nest of *A. colombica* while the laboratory-reared sample originated from a colony of *A. sexdens*. Although these two species cultivate the same species of fungus, it is unknown to what degree the diversity of other fungus garden members varies. A promising avenue of future research may be to investigate the overall physiological differences between fungus gardens of different ant species and environmental habitation in more detail.

**References**

1. **Suen, G, Scott, JJ, Aylward, FO, Adams, SM, Tringe, SG, Pinto-Tomas, AA, Foster, CE, Pauly, M, Weimer, PJ, Barry, KW, Goodwin, LA, Bouffard, P, Li, L, Osterberger, J, Harkins, TT, Slater, SC, Donohue, TJ, and Currie, CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6:**e1001129.

2. **Kunin, V, and Hugenholtz, P.** 2010. PyroTagger: A fast, accurate pipeline for analysis of rRNA amplicon pyrosequence data. *The Open Journal*.

3. **Cole, JR, Wang, Q, Cardenas, E, Fish, J, Chai, B, Farris, RJ, Kulam-Syed-Mohideen, AS, McGarrell, DM, Marsh, T, Garrity, GM, and Tiedje, JM.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* **37:**D141-145.

4. **Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, YJ, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, ML, Jarvie, TP, Jirage, KB, Kim, JB, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, McDade, KE, McKenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth, GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A, Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF, and Rothberg, JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-380.

5. **Noguchi, H, Park, J, and Takagi, T.** 2006. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34:**5623-30.

6. **Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:**403-410.

7. **Markowitz, VM, Mavromatis, K, Ivanova, NN, Chen, IM, Chu, K, and Kyrpides, NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:**2271-8.

8. **Markowitz, VM, Ivanova, NN, Szeto, E, Palaniappan, K, Chu, K, Dalevi, D, Chen, IM, Grechkin, Y, Dubchak, I, Anderson, I, Lykidis, A, Mavromatis, K, Hugenholtz, P, and Kyrpides, NC.** 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36:**D534-8.

9. **Angly, FE, Willner, D, Prieto-Davo, A, Edwards, RA, Schmieder, R, Vega-Thurber, R, Antonopoulos, DA, Barott, K, Cottrell, MT, Desnues, C, Dinsdale, EA, Furlan, M, Haynes, M, Henn, MR, Hu, Y, Kirchman, DL, McDole, T, McPherson, JD, Meyer, F, Miller, RM, Mundt, E, Naviaux, RK, Rodriguez-Mueller, B, Stevens, R, Wegley, L, Zhang, L, Zhu, B, and Rohwer, F.** 2009. The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput. Biol.* **5:**e1000593.

10. **Brady, A, and Salzberg, SL.** 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat. Methods* **6:**673-6.

11. **Stevenson, BS, Eichorst, SA, Wertz, JT, Schmidt, TM, and Breznak, JA.** 2004. New strategies for cultivation and detection of previously uncultured microbes. *Appl. Environ.*

*Microbiol.* **70:**4748-55.

12. **Rusch, DB, Halpern, AL, Sutton, G, Heidelberg, KB, Williamson, S, Yooseph, S, Wu, D, Eisen, JA, Hoffman, JM, Remington, K, Beeson, K, Tran, B, Smith, H, Baden-Tillson, H, Stewart, C, Thorpe, J, Freeman, J, Andrews-Pfannkoch, C, Venter, JE, Li, K, Kravitz, S, Heidelberg, JF, Utterback, T, Rogers, YH, Falcon, LI, Souza, V, Bonilla-Rosso, G, Eguiarte, LE, Karl, DM, Sathyendranath, S, Platt, T, Bermingham, E, Gallardo, V, Tamayo-Castillo, G, Ferrari, MR, Strausberg, RL, Nealson, K, Friedman, R, Frazier, M, and Venter, JC.** 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5:**e77.

13. **Lukashin, AV, and Borodovsky, M.** 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26:**1107-15.

14. **Moriya, Y, Itoh, M, Okuda, S, Yoshizawa, AC, and Kanehisa, M.** 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35:**W182-5.

15. **Suen, G, Goldman, BS, and Welch, RD.** 2007. Predicting prokaryotic ecological niches using genome sequence analysis. *PLoS One* **2:**e743.

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

17. **Letunic, I, and Bork, P.** 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:**127-8.

18. **Pavlidis, P, and Noble, WS.** 2003. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19:**295-6.

19. **Cantarel, BL, Coutinho, PM, Rancurel, C, Bernard, T, Lombard, V, and Henrissat, B.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37:**D233-8.

20. **Warnecke, F, Luginbuhl, P, Ivanova, N, Ghassemian, M, Richardson, TH, Stege, JT, Cayouette, M, McHardy, AC, Djordjevic, G, Aboushadi, N, Sorek, R, Tringe, SG, Podar, M, Martin, HG, Kunin, V, Dalevi, D, Madejska, J, Kirton, E, Platt, D, Szeto, E, Salamov, A, Barry, K, Mikhailova, N, Kyrpides, NC, Matson, EG, Ottesen, EA, Zhang, X, Hernandez, M, Murillo, C, Acosta, LG, Rigoutsos, I, Tamayo, G, Green, BD, Chang, C, Rubin, EM, Mathur, EJ, Robertson, DE, Hugenholtz, P, and Leadbetter, JR.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:**560-565.

21. **Pope, PB, Denman, SE, Jones, M, Tringe, SG, Barry, K, Malfatti, SA, McHardy, AC, Cheng, JF, Hugenholtz, P, McSweeney, CS, and Morrison, M.** 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles

different from other herbivores. *Proc. Natl. Acad. Sci. USA* **107:**14793-8.

22. **Tatusov, RL, Natale, DA, Garkavtsev, IV, Tatusova, TA, Shankavaram, UT, Rao, BS, Kiryutin, B, Galperin, MY, Fedorova, ND, and Koonin, EV.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:**22-8.

23. **Wisniewski, JR, Zougman, A, Nagaraj, N, and Mann, M.** 2009. Universal sample preparation method for proteome analysis. *Nat. Methods* **6:**359-62.

24. **Eng, J, McCormack, A, and Yates, J.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5:**976-989.

25. **Kim, S, Gupta, N, and Pevzner, PA.** 2008. Spectral Probabilities and Generating Functions of Tandem Mass Spectra: A Strike against Decoy Databases. *Journal of Proteome Research* **7:**3354-3363.

26. **Burnum, KE, Callister, SJ, Nicora, CD, Purvine, SO, Hugenholtz, P, Warnecke, F, Scheffrahn, RH, Smith, RD, and Lipton, MS.** 2011. Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *ISME J* **5:**161-164.

27. **Peng, J, Elias, JE, Thoreen, CC, Licklider, LJ, and Gygi, SP.** 2002. Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC-MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *Journal of Proteome Research* **2:**43-50.

28. **Petyuk, VA, Mayampurath, AM, Monroe, ME, Polpitiya, AD, Purvine, SO, Anderson, GA, Camp, DG, 2nd, and Smith, RD.** 2010. DtaRefinery, a software tool for elimination of systematic errors from parent ion mass measurements in tandem mass spectra data sets. *Mol. Cell Proteomics* **9:**486-96.

## Appendix 2: Supplementary Information for Chapter 3

**Previously Published as Supplementary Material**: Frank O Aylward, Kristin E Burnum-Johnson, Susannah G Tringe, Clotilde Teiling, Daniel M Tremmel, Joseph A Moeller, Jarrod J Scott, Kerrie W Barry, Paul D Piehowski, Carrie D Nicora, Stephanie A Malfatti, Matthew E Monroe, Samuel O Purvine, Lynne A Goodwin, Richard D Smith, George M Weinstock, Nicole M Gerardo, Garret Suen, Mary S Lipton, Cameron R Currie (2013). *Leucoagaricus gongylophorus* produces a diversity of enzymes for recalcitrant plant polymer degradation in leaf-cutter ant fungus gardens. *Applied and Environmental Microbiology*, 79 (12), 3770-3778.

**Author contributions are the same as in Chapter 3.**

### A2.1: Protein Prediction and Annotation of the Draft *L. gongylophorus* Genome

In this study, we used Augustus (1) and GeneMark-ES (2) to predict proteins in the draft *L. gongylophorus* genome. These programs have been shown to effectively predict coding sequences in eukaryotic genomes, but further sequencing efforts and complementary transcriptomic analyses in the future will be necessary to provide a more precise estimate of the complete coding potential of this fungus. Initially 5,418 proteins were predicted by Augustus and 60,427 by GeneMark-ES. The number of proteins predicted by Augustus is likely an underestimate of that actually encoded in the draft genome, while the number of proteins predicted by GeneMark-ES is an overestimate likely due to the large number of small proteins this program predicted (< 50 aa) as well as proteins predicted on short contigs (< 1,000 bp). To remove spurious predictions we restricted our GeneMark-ES analyses only to contigs > 1,000 bp, and all proteins < 50 aa were removed. This left a total of 11,604 GeneMark-ES protein predictions, which were compared to the Augustus predictions using BLASTP to obtain a total of 12,132 non-redundant protein predictions.

To test the validity of the 12,132 high-quality protein predictions, we compared them to several databases to identify homologous proteins in other organisms. Using RPS-BLAST we

compared these proteins to the KOG (3) and Pfam (4) databases, and using BLASTP we compared them to the NCBI non-redundant proteins database (nr) (5), the UniProtKB-Swissprot manually-curated database (6), and a custom database of proteins encoded by five Agaricales fungi (*Coprinopsis cinerea* Okayama (#130), *Gymnopus luxurians*, *Hypholoma sublateritium*, *Moniliophthora perniciosa* FA553, and *Schizophyllum commune* H4-8) currently available on the Integrated Microbial Genomes database (IMG, (7)) (e-value 1e-5 for all). Of the 12,132 protein predictions, the majority of these proteins had homologs in at least one of the databases (8,413, 69.3%), suggesting they represent actual coding sequences.

To optimize chances of identifying proteins present in the fungus garden samples in our metaproteomic experiments, we used the initial Augustus and GeneMark-ES predictions (5,418 and 60,427 amino acid sequences, respectively) for our spectral mapping analyses. The 12,132 proteins described above are likely a better estimate of the actual number of proteins encoded by *L. gongylophorus*, but due to the utility of proteomics for confirming gene predictions, we decided to use as many putative proteins as possible for our spectral mapping analyses. After spectral mapping of all of our metaproteomic datasets to the full Augustus and GeneMark-ES protein prediction datasets, we confirmed a total of 4,567 predictions (31 from the Augustus dataset only, 1,161 from the GeneMark-ES dataset only, and 3,375 from both). See below for a detailed description of our spectral mapping procedures.

### A2.2: Proteomic Sample Preparation

Six samples corresponding to the top, middle, and bottom of *Atta cephalotes* and *Acromyrmex echinatior* gardens were weighed to the mass of 150 mg and added to an ice cold mortar and pestle with liquid nitrogen. The sample was ground for ~2 minutes and then 2 mL of

water was added to the sample with continuous grinding until the mixture was thawed. The sample was pipetted into 0.6 mL centrifuge tubes with 0.1 mm zirconia/silica beads and bead-beat for 3 minutes. The tube was centrifuged out at 4 °C at 10,000 g for 5 minutes and the supernatant removed. The collected supernatant was then centrifuged at 4000 g for 5 minutes to remove debris.

**A2.3: Proteomic sample preparation part 1- Rapigest**

Rapigest SF Surfactant (Waters, Milford, MA) was added to the protein pellet at a concentration of 0.2% and a volume of 100 µl. The sample was vortexed into solution and boiled for 5 minutes to completely solubilize. The sample was cooled and assayed with Bicinchoninic acid (BCA) (Thermo Scientific, Rockford, IL) to determine the protein concentration and then reduced with 5 mM dithiothreitol (DTT) (Sigma-Aldrich, St. Louis, MO) at 60 °C for 30 minutes. The sample was then tryptically digested (50:1 protein: trypsin (w/w)) (Promega, Madison, WI) at 37 °C overnight. The following day the sample was acidified to pH ~2 with 0.5% high purity trifluoroacetic acid (TFA) (Sigma-Aldrich, St. Louis, MO) and then incubated at 37 °C for 45 minutes in order to precipitate the Rapigest surfactant. The sample was then centrifuged at 15,000 g for 10 minutes to pellet the surfactant. The supernatant was removed without disturbing the pellet and neutralized using ammonium hydroxide (Sigma-Aldrich, St. Louis, MO). Another BCA assay was done to determine the final peptide concentration. These peptides were added in equal mass to the peptides from the following FASP proteomic sample preparation part 2. The rest of the peptides were used for offline high pH RP C-18 fractionation described below.

**A2.4: Proteomic sample preparation part 2- FASP (Filter-Aided Sample Preparation)**

Filter Aided Sample Preparation (FASP (8)) was used as an alternative sample preparation method. Each separate garden sample was weighed to the mass of 150mg and added to an ice cold mortar and pestle with liquid nitrogen. The sample was ground for ~2 minutes and added to a centrifuge tube. Then 1mL of buffer (100 mM DTT in 100 mM Tris/HCL, pH 7.6) was added and the sample was ground using an OMNI handheld homogenizer with disposable probes (OMNI International, Kennesaw, GA) on ice. Sodium dodecyl sulfate (SDS) (Sigma-Aldrich, St. Louis, MO) was added to the sample to a concentration of 2%, vortexed and incubated at 95C for 5 minutes. All samples were then barocycled (Pressure Biosciences Inc., South Easton, MA) for 10 cycles (20 seconds at 35,000 psi back down to ambient pressure for 10 seconds). The sample was clarified by centrifuging at 14,000 g for 5 minutes. Amicon Ultra-15 30K MWCO centrifuge devices (Millipore, Billerica, MA) were used for buffer exchange. The sample was added to the filter unit with 8 M urea, pH 8.5 at 15% of the total volume. The unit was centrifuged at 4000 g for 40 minutes (until it reached the dead volume of 200 ul). Another 10 mL of 8 M urea, pH 8.0 was added. The unit was spun at 4000 g for 40 minutes and this was repeated 3 times followed by 3 rinses with 100 mM $NH_4HCO_3$, pH 8.0. A BCA assay was used to determine the protein concentration and enough 25 mM $NH_4HCO_3$ was added to the filter unit to cover the filter. The flow-through collection tube was thoroughly cleaned and the sample was tryptically digested in a 1:50 (w/w) ratio with 1mM CaCl2 and was allowed to incubate overnight at 37 °C. The following day the much smaller peptide sample was centrifuged at 4000 g for 30 minutes and the peptides were collected in the flow through. The filter was rinsed with 25 mM $NH_4HCO_3$, and the flow-through was pooled. The sample was dried to ~100 µl in a

Speed vac and a BCA assay was performed for a final protein concentration. An equal mass of peptide from the FASP procedure was added to the Rapigest peptides and analyzed in triplicate by capillary LC-MS/MS with a 100 minute formic acid separation on a Velos Orbitrap. The rest of the peptides were used for offline high pH RP C-18 fractionation described below.

**A2.5: High pH RP C-18 Fractionation**

Samples were diluted to a volume of 900 uL with 10 mm ammonium formate buffer (pH 10.0), and resolved on a XBridge C18, 250x4.6 mm, 5 µM with 4.6x20 mm guard column (Waters, Milford, MA). Separations were performed at 0.5 mL/min using an Agilent 1100 series HPLC system (Agilent Technologies, Santa Clara, CA) with mobile phases (A) 10 mM Ammonium Formate, pH 10.0 and (B) 10 mM Ammonium Formate, pH 10.0/acetonitrile (10:90). The gradient was adjusted from at 100% A to 95% A over the first 10 min, 95% A to 65% A over minutes 10 to 70, 65% A to 30% A over minutes 70 to 85, maintained at 30% A over minutes 85 to 95, re-equilibrated with 100% A over minutes 95 to 105, and held at 100% A until minute 120. Fractions were collected every 1.25 minutes (96 fractions over the entire gradient) and every 24th fraction was combined for a total of 24 samples (each with n=4 fractions pooled). All fractions were half way dried under vacuum and like sample fractions from the Rapigest procedure and FASP procedure were combined. The fractions were then completely dried down and 35 µL of 25 mM ammonium bicarbonate was added to each fraction for storage at -20 °C until LC-MS/MS analysis.

**A2.6: Mass Spectrometry Analyses**

The unfractionated samples were analyzed in triplicate by tandem MS (MS/MS) on a LTQ Orbitrap Velos (Thermo Fisher Corporation, San Jose, CA). The fractionated samples were

analyzed by tandem MS on a LTQ Orbitrap (Thermo Fisher Corporation, San Jose, CA). Columns (60 cm long / 75 µm ID) were packed (in-house) with Jupiter 3 µm C18 resin (phenomenex; Torrance, California, USA). Liquid chromatography systems parameters; constant pressure of 10,000 PSI, aqueous to organic (0.1% FA in water - A, and 0.1% FA in 99.9% acetonitrile - B) exponential gradient over 100 minutes, and a 500 nL/min flow rate (2). MS parameters; capillary voltage 2.2 kV, Collision energy 35 eV, exclusion time 1 min, 10 most intense peaks selected for fragmentation, high mass accuracy MS, low mass accuracy MS/MS.

**A2.7: Proteomic Data Analysis and Data Filtering**

As previously published (9) Sequest analysis software (10) was used to match the MS/MS fragmentation spectra with peptides predicted from genomic and metagenomic datasets. The 12 protein prediction datasets used in this study were generated from the *L. gongylophorus* draft genome (both Augustus and full GeneMark-ES predictions), metagenomes constructed from the gardens of *At. cephalotes, Ac. echinatior,* and *At. colombica* (garden top and bottom strata, see (9) for details on these metagenomes), and complete or draft genomes of *Enterobacter* FGI 35, *Pseudomonas* FGI 182, *Serratia* FGI 94, *Cronobacter* FGI 157, *Klebsiella variicola* AT-22, and *Pantoea* at-9b. In order to increase both the amount and confidence of peptide identifications we first used MS-GF scoring to generate a spectra level confidence value (this software is available as an open source from http://www.cs.ucsd.edu/users/ppevzner /software.html.) (11). We filtered the spectra using a MS-GF cutoff of 1E-10 and then calculated our false discovery rate (FDR) utilizing decoy database searches (%FDR=((Reverse Identifications*2)/Total Identifications)*100). To achieve ≤ 5% spectral level FDR we varied Xcorr (Sequest's main scoring value related to peptide confidence) and DelCN (how good the

XCorr is relative to the next best match DelCN=($XCorr_1$-$XCorr_2$)/$XCorr_1$) values for spectra (group by peptide charge state and number of tryptic ends). Ultimately, only spectra matching with < 10 ppm mass error were included in subsequent analyses. Confident spectra identifications along with their associated MS-GF, Xcorr, and DelCN values can be found online at http://omics.pnl.gov/view/publication_1059.html. Mass spectra mapping to fungal lignocellulases were examined manually if they represented the only unique peptide matching to that protein.

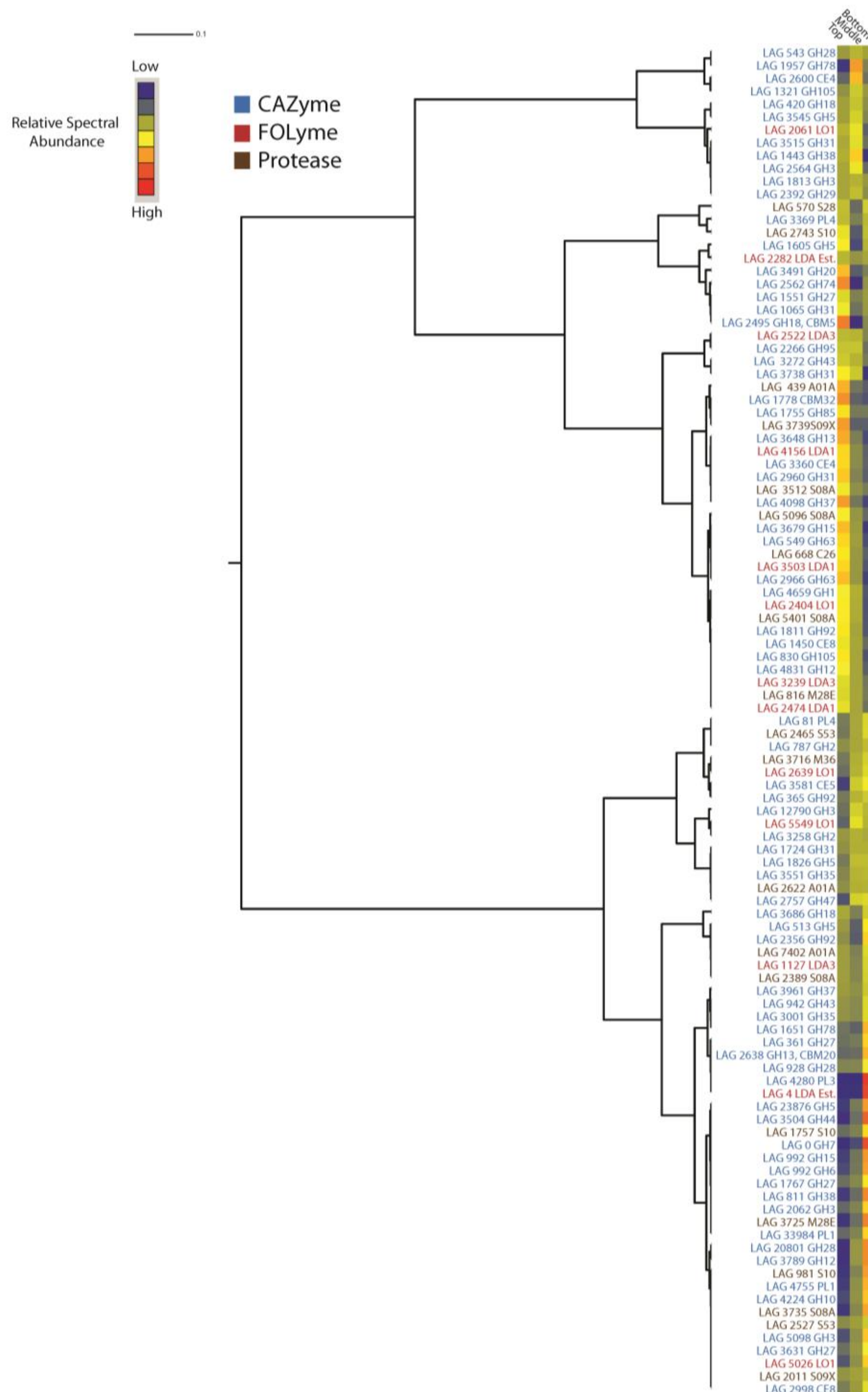



**Figure A2.1.** Clustering of all lignocellulases identified in the *Ac. echinatior* metaprotoemic samples with at least 10 mass spectra mapping. Individual proteins are color-coded by enzyme type. Rows have been normalized to unity. Samples were clustered using a Pearson's r.

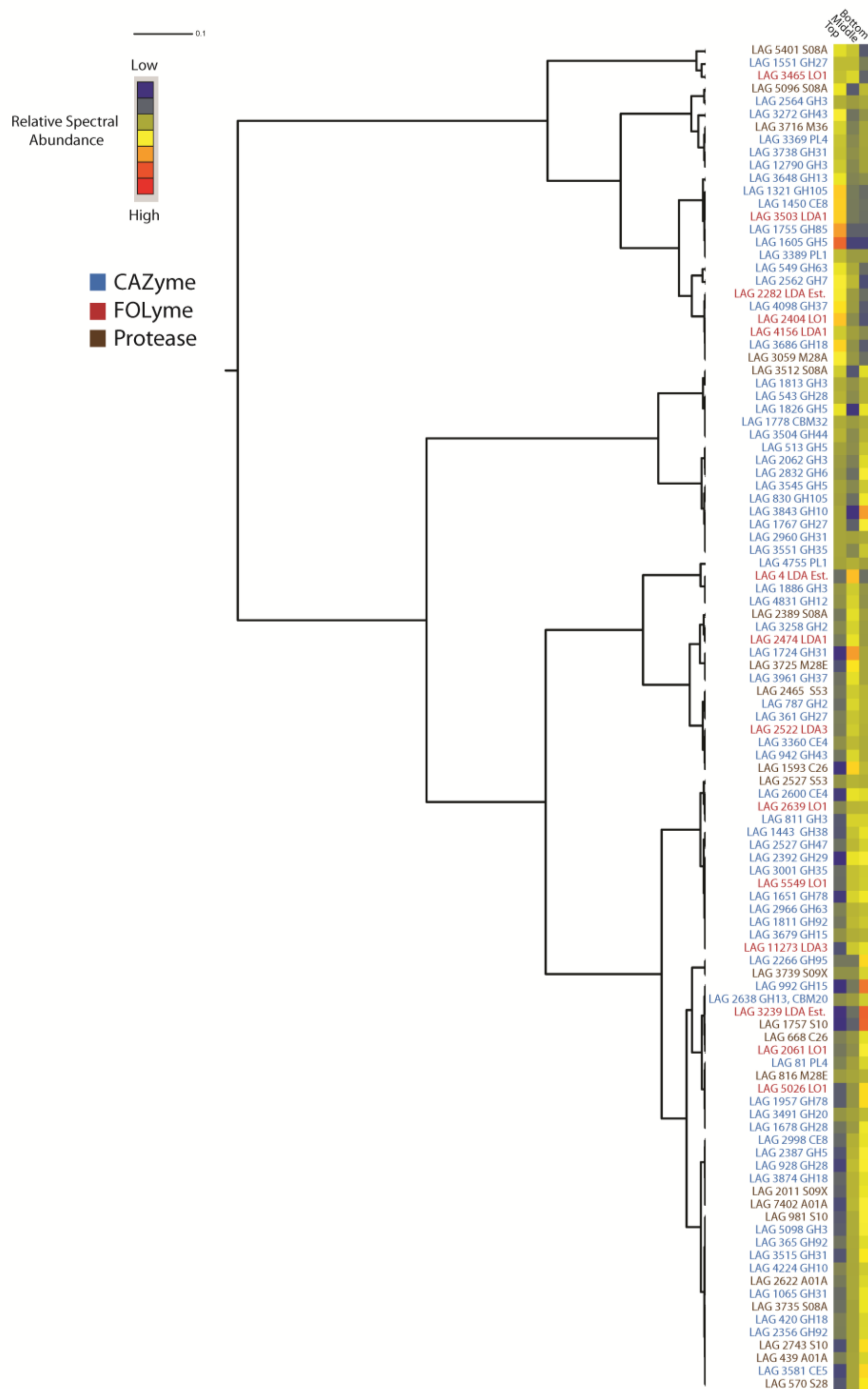



**Figure A2.2.** Clustering of all lignocellulases identified in the *At. cephalotes* metaprotoemic samples with at least 10 mass spectra mapping. Individual proteins are color-coded by enzyme type. Rows have been normalized to unity. Samples were clustered using a Pearson's r.

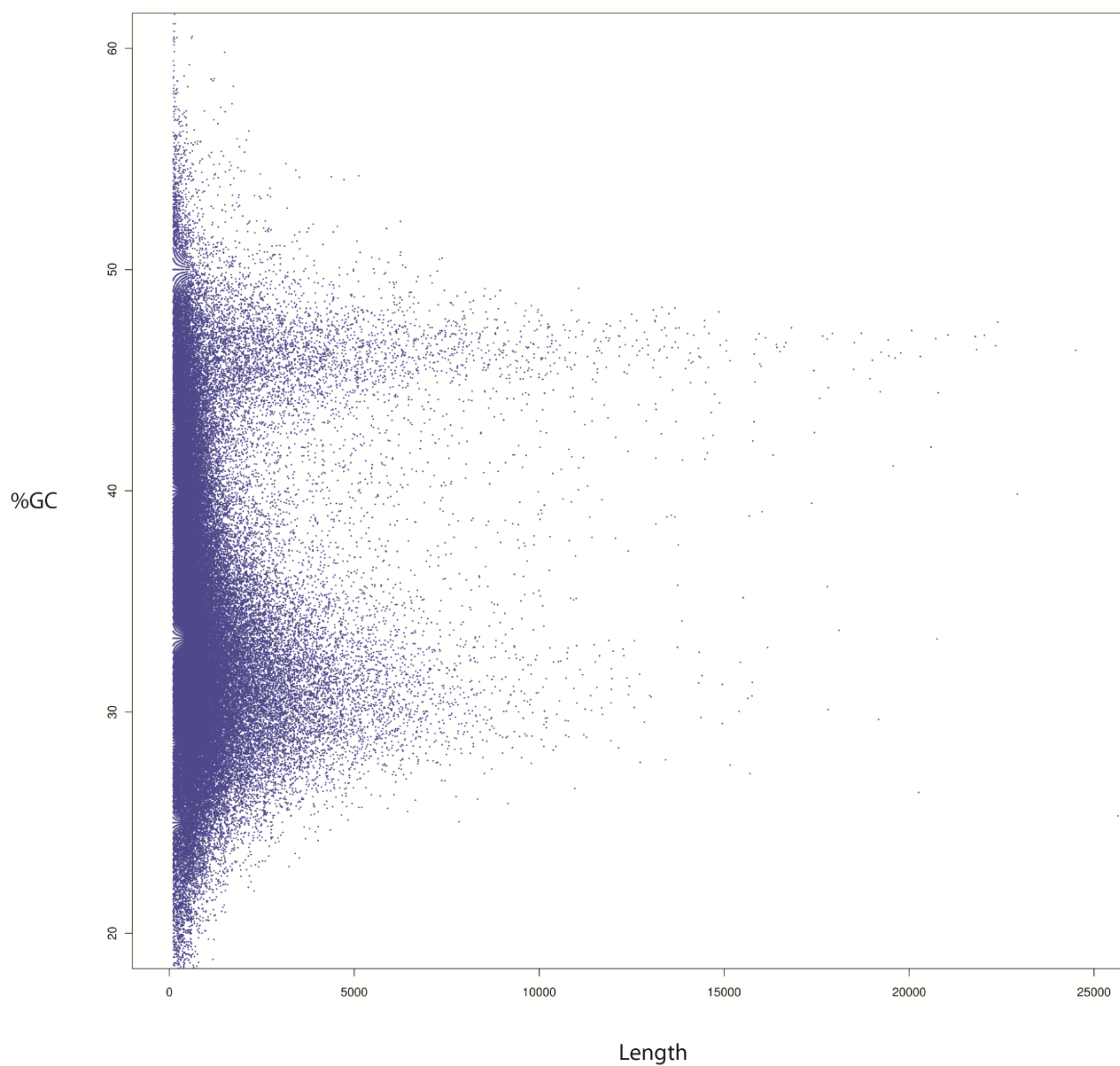



**Figure A2.3.** Graph of % GC content and nucleotide length for contigs in the *L. gongylophorus* assembly. Contigs larger than ~250,000 bp are not shown.

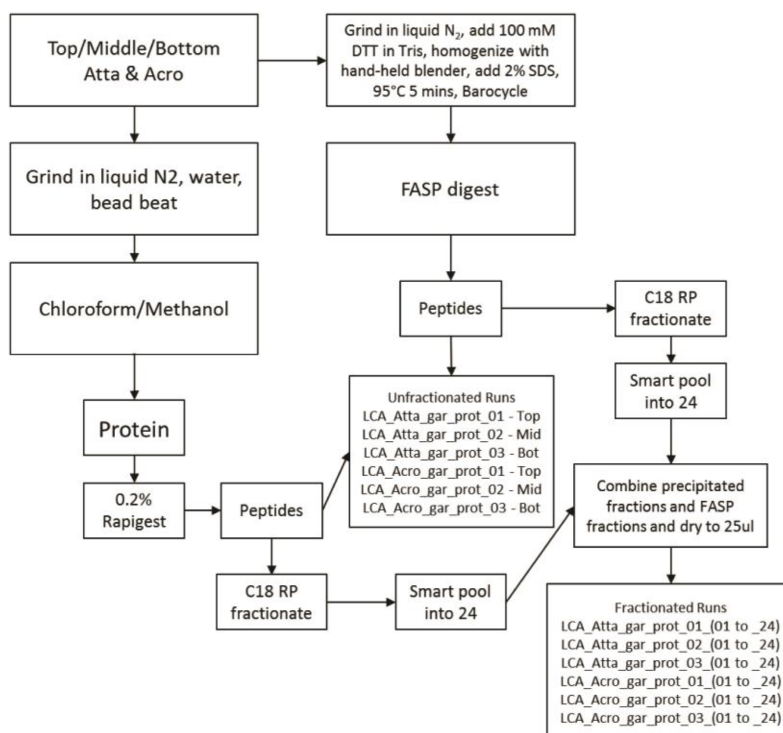



**Figure A2.4.** Flow chart outlining the processing of the metaproteomic samples.

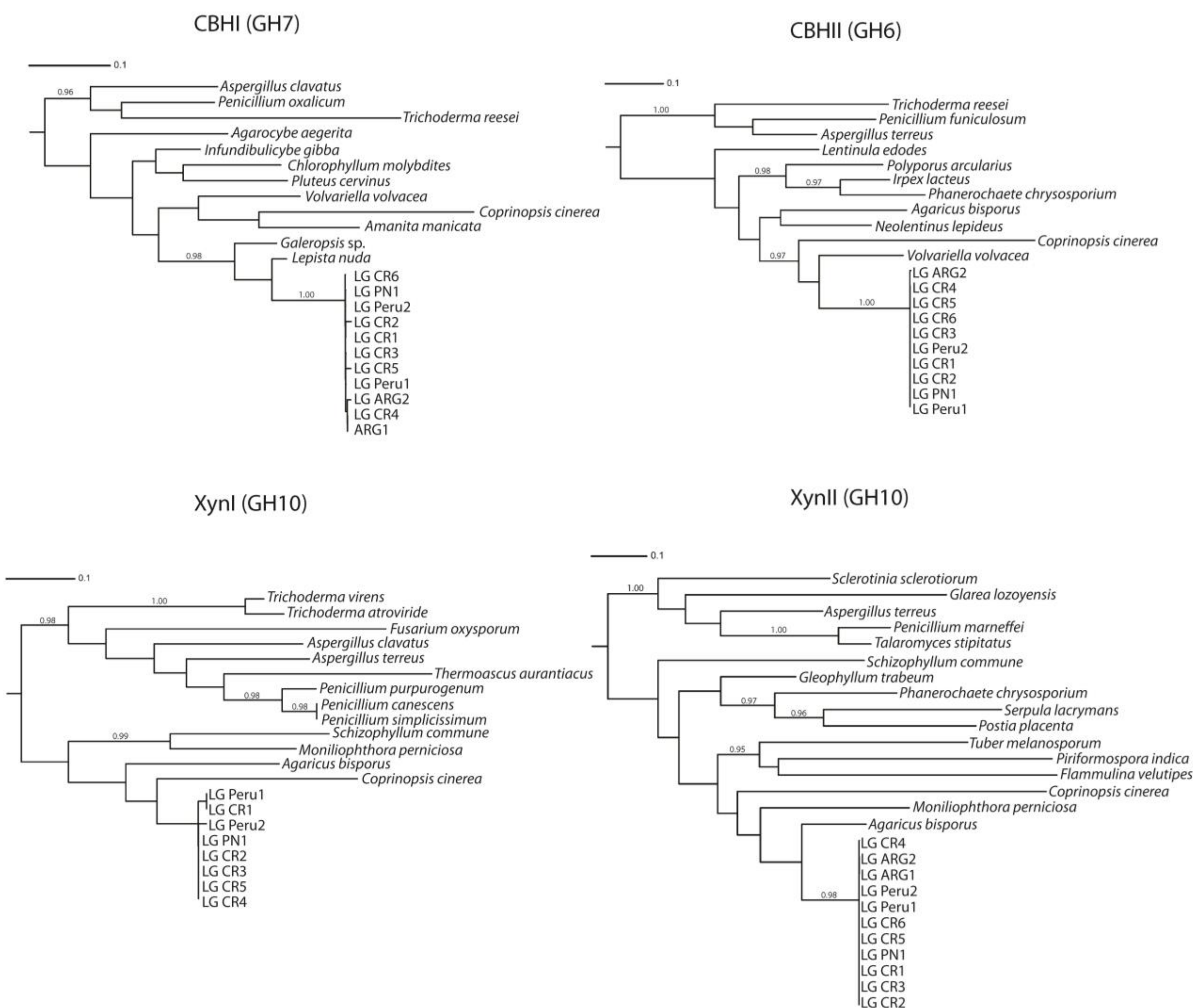



**Figure A2.5**. Maximum-likelihood phylogenetic trees created with FastTree from amino acid alignments of CAZymes amplified and sequenced from *L. gongylophorus* isolates. Aside from the *L. gongylophorus* sequences, proteins from select fungi have also been included. SH-like local support values between 0-1 were estimated using FastTree. Only support values > 0.95 are shown. The scale bar indicates expected amino acid substitutions per site. The trees were rooted manually at the ancestor of the Ascomycota and Basidiomycota.

**Table A2.1**. Annotation statistics for the 12,132 non-redundant protein predictions made from the draft *Leucoagaricus gongylophorus* genome.

| Database Used | Proteins Matching (% of Total) |
|---|---|
| KOG Database | 4,170 (34.4%) |
| Pfam Database | 4,333 (35.7%) |
| UniProtKB/Swiss-Prot Database | 4,930 (40.6%) |
| NCBI nr Database | 8,274 68.2%) |
| Custom IMG Agaricales Dataset | 7,780 (64.1%) |
| Matches to at least one database | 8,413 (69.3%) |

**Table A2.2**. List of CAZyme modules identified in the *L. gongylophorus* genome that were verified by spectral mapping of metaproteomic data.

| CAZy Family | Annotation | Number Identified |
|---|---|---|
| CBM 13 | Carbohydrate binding module | 2 |
| CBM20 | Carbohydrate binding module | 1 |
| CBM27 | Carbohydrate binding module | 1 |
| CBM32 | Carbohydrate binding module | 1 |
| CBM5 | Carbohydrate binding module | 1 |
| CBM57 | Carbohydrate binding module | 1 |
| CE15 | Glucuronoyl methylesterase | 1 |
| CE4 | Acetyl xylan esterase; chitin deacetylase | 4 |
| CE5 | Acetyl xylan esterase; cutinase | 1 |
| CE8 | Pectin methylesterase | 3 |
| GH1 | β-glucosidase; β-galactosidase; β-mannosidase | 2 |
| GH10 | Xylanase | 2 |
| GH105 | Rhamnogalacturonyl hydrolase | 3 |
| GH11 | Xylanase | 1 |
| GH115 | Xylan α-1,2-glucuronidase | 1 |
| GH12 | Endoglucanase; xyloglucan hydrolase | 2 |
| GH13 | α-amylase; pullulanase | 2 |
| GH15 | glucoamylase; glucodextranase | 2 |
| GH16 | Galactosidase; licheninase; endo-1,4-β-galactosidase | 1 |
| GH17 | glucan endo-1,3-β-glucosidase; licheninase | 1 |
| GH18 | chitinase; endo-β-N-acetylglucosaminidase | 5 |
| GH2 | β-galactosidase; β-mannosidase | 2 |
| GH20 | β-hexosaminidase; lacto-N-biosidase | 2 |
| GH27 | α-galactosidase; α-N-acetylgalactosaminidase | 4 |
| GH28 | polygalacturonase | 4 |
| GH29 | α-L-fucosidase | 1 |
| GH3 | β-glucosidase; xylan 1,4-β-xylosidase | 8 |
| GH31 | α-glucosidase; α-1,3-glucosidase | 5 |
| GH35 | β-galactosidase; exo-β-glucosaminidase | 3 |
| GH37 | α,α-trehalase | 2 |
| GH38 | α-mannosidase | 1 |
| GH43 | β-xylosidase; β-1,3-xylosidase | 4 |
| GH44 | endoglucanase; xyloglucanase | 1 |
| GH47 | α-mannosidase | 4 |
| GH5 | chitosanase; β-mannosidase | 6 |
| GH6 | endoglucanase; cellobiohydrolase | 1 |
| GH63 | α-glucosidase; α-1,3-glucosidase | 2 |
| GH7 | endo-β-1,4-glucanase; cellobiohydrolase | 1 |
| GH74 | endoglucanase; oligoxyloglucan cellobiohydrolase | 1 |
| GH78 | α-L-rhamnosidase | 2 |
| GH85 | endo-β-N-acetylglucosaminidase | 1 |
| GH9 | endoglucanase; cellobiohydrolase | 1 |
| GH92 | Mannosidase | 3 |
| GH95 | α-1,2-L-fucosidase | 1 |
| PL1 | Pectate lyase | 2 |
| PL3 | Pectate lyase | 1 |
| PL4 | rhamnogalacturonan lyase | 2 |
| PL8 | hyaluronate lyase; chondroitin AC lyase | 1 |
| Total | | 103 |

**Table A2.3.** List of FOLymes identified in the *L. gongylophorus* genome that were verified by spectral mapping of metaproteomic data.

| Enzyme | Annotation | Number Identified |
|---|---|---|
| LDA Esterase | Auxiliary esterase | 2 |
| LDA1 | Aryl-alcohol oxidase | 4 |
| LDA3 | Glyoxal oxidase | 5 |
| LO1 | Laccase | 9 |
| Total | | 20 |

**Table A2.4.** List of MEROPS proteases with signal peptides identified in the *L. gongylophorus* genome confirmed through spectral mapping of metaproteomic data.

| MEROPS Family | Annotation | Number Identified |
|---|---|---|
| A01A | Aspartyl protease | 4 |
| C13 | Cysteine peptidase | 1 |
| C26 | Cysteine peptidase | 2 |
| M28A | Metalloprotease | 1 |
| M28E | Metalloprotease | 2 |
| M36 | Metalloprotease | 2 |
| S08A | Serine Protease | 5 |
| S08B | Serine Protease | 1 |
| S09X | Serine Protease | 2 |
| S10 | Serine Protease | 3 |
| S28 | Serine Protease | 1 |
| S53 | Serine Protease | 2 |
| Total | | 26 |

**Table A2.5.** Primers used for the amplification of CAZyme-encoding genes in *L. gongylophorus* isolates.

| Name of Primer (Forward/Reverse) | CAZy Amplicon | Forward primer | Reverse Primer | Reference |
|---|---|---|---|---|
| fungcbhIF/fungcbhIR | GH7 (cbhI) | 5′-ACCAAYTGCTAYACIRGYAA-3′ | 5’-GCYTCCCAIATRTCCATC-3′ | (12) |
| Cbh2F/Cbh2R | GH6 (cbh II) | 5'-ATGGATACCAAAGCCAGCAC-3’ | 5'-GATCTCCCAGGTTTGGTTGA-3’ | This study |
| GH10AF/GH10AR | GH10 (xyn1) | 5'-TGAACTTTGGCATTGTTGGA-3’ | 5'-GCTGCTGACCCCAACACTAT-3’ | This study |
| GH10BF/GH10BR | GH10 (xynII) | 5'-ATTGTCAGACCCTGGTGTCC-3’ | 5'-CATCGTCTGGTGAAAGTGGA -3’ | This study |

**A2.8 References**


1. **Stanke, M, Schoffmann, O, Morgenstern, B, and Waack, S.** 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7:**62.

2. **Borodovsky, M, and Lomsadze, A.** 2011. Current protocols in bioinformatics, vol. Unit 4.6, eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. Wiley Online Library.

3. **Tatusov, RL, Fedorova, ND, Jackson, JD, Jacobs, AR, Kiryutin, B, Koonin, EV, Krylov, DM, Mazumder, R, Mekhedov, SL, Nikolskaya, AN, Rao, BS, Smirnov, S, Sverdlov, AV, Vasudevan, S, Wolf, YI, Yin, JJ, and Natale, DA.** 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:**41.

4. **Finn, RD, Tate, J, Mistry, J, Coggill, PC, Sammut, SJ, Hotz, HR, Ceric, G, Forslund, K, Eddy, SR, Sonnhammer, EL, and Bateman, A.** 2008. The Pfam protein families database. *Nucleic Acids Res.* **36:**D281-8.

5. **Pruitt, KD, Tatusova, T, and Maglott, DR.** 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35:**D61-5.

6. **Apweiler, R, Bairoch, A, Wu, CH, Barker, WC, Boeckmann, B, Ferro, S, Gasteiger, E, Huang, H, Lopez, R, Magrane, M, Martin, MJ, Natale, DA, O'Donovan, C, Redaschi, N, and Yeh, LS.** 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32:**D115-9.

7. **Markowitz, VM, Chen, IM, Palaniappan, K, Chu, K, Szeto, E, Grechkin, Y, Ratner, A, Jacob, B, Huang, J, Williams, P, Huntemann, M, Anderson, I, Mavromatis, K, Ivanova, NN, and Kyrpides, NC.** 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res.* **40:**D115-22.

8. **Wisniewski, JR, Zougman, A, Nagaraj, N, and Mann, M.** 2009. Universal sample preparation method for proteome analysis. *Nat. Methods* **6:**359-62.

9. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **6:**1688–1701.

10. **Eng, J, McCormack, A, and Yates, J.** 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5:**976-989.

11. **Kim, S, Gupta, N, and Pevzner, PA.** 2008. Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* **7:**3354-63.

12. **Edwards, IP, Upchurch, RA, and Zak, DR.** 2008. Isolation of fungal cellobiohydrolase I genes from sporocarps and forest soils by PCR. *Appl. Environ. Microbiol.* **74:**3481-9.

# Appendix 3: Mountain Pine Beetles Colonizing Historical and Naïve Host Trees are Associated with a Bacterial Community Highly Enriched in Genes Contributing to Terpene Metabolism

**Previously Published:** Aaron S. Adams*, Frank O. Aylward*, Sandye M. Adams, Nadir Erbilgin, Brian H. Aukema, Cameron R. Currie, Garret Suen, and Kenneth F. Raffa (2013). Mountain Pine Beetles Colonizing Historical and Naïve Host Trees are Associated with a Bacterial Community Highly Enriched in Genes Contributing to Terpene Metabolism. *Applied and Environmental Microbiology*, 79 (11), 3468-3475.

*Authors contributed equally.

**Author Contributions**: For this chapter I performed all computational experiments, and I conceived of analyses with Garret Suen and Ken Raffa. Other authors collected the samples, generated the metagenomic data, provided helpful advice for analyses, and helped with writing.

## A3.1 Abstract

The mountain pine beetle, *Dendroctonus ponderosae*, is a subcortical herbivore native to western North America that can kill healthy conifers by overcoming host tree defenses, which consist largely of high terpene concentrations. The mechanisms by which these beetles contend with toxic compounds are not well understood. Here, we explore a component of the hypothesis that beetle-associated bacterial symbionts contribute to the ability of *D. ponderosae* to overcome tree defenses by assisting with terpene detoxification. Such symbionts may facilitate host tree transitions during range expansions currently being driven by climate change. For example, this insect has recently breached the historical geophysical barrier of the Canadian Rocky Mountains, providing access to näive tree hosts and unprecedented connectivity to eastern forests. We use culture-independent techniques to describe the bacterial community associated with *D. ponderosae* and their galleries from their historical host, *Pinus contorta*, and their more recent host, hybrid *P. contorta* - *Pinus banksiana*. We show that these communities are enriched with genes involved in terpene degradation compared with other plant biomass-processing microbial communities. These pine beetle microbial communities are dominated by members of the genera

*Pseudomonas, Rahnella, Serratia*, and *Burkholderia* and the majority of genes involved in terpene and monoterpene degradation belong to these genera. Our work provides the first metagenome of bacterial communities associated with a bark beetle, and is consistent with a potential microbial contribution to detoxification of tree defenses needed to survive the subcortical environment.

**A3.2 Introduction**

In addition to the well-established roles of pairwise symbioses, the importance of multipartite associations and microbial communities to multicellular organisms is becoming increasingly recognized (1-6). Despite this, our understanding of how symbiotic relationships contribute to large-scale processes, such as ecosystem dynamics and the structure and functioning of biomes, remains poorly understood. Insects serve as a particularly useful model for exploring these cross-scale interactions because of their widespread and diverse associations with microbiota, roles in driving ecosystem processes, influences on human socioeconomic values, rapid evolutionary adaptations, and shifting responses to anthropogenic forces such as climate change, species invasions, and habitat alteration.

Bark beetles (Curculionidae: Scolytinae) have the ability to overcome host tree defenses, and thus colonize and kill healthy conifers. Several species undergo intermittent landscape-scale outbreaks, which appear to be increasing in frequency, magnitude, and interspecific confluence as a result of a warming climate and habitat conversions. For example, bark beetles caused substantial mortality across 47 million acres of conifers in western North America from 1997-2007, and the ongoing mountain pine beetle (*Dendroctonus ponderosae* Hopkins) outbreak is predicted to deplete 1 trillion cubic meters of pine in British Columbia, Canada alone by 2014 (7).

The historical range of *D. ponderosae* extends from northern Mexico to southern British Columbia, and inland to western North Dakota in the United States and the Rocky Mountains in Canada (8). Its preferred host is lodgepole pine, *Pinus contorta* (Douglas ex. Loud.), which occurs throughout this range. As conditions have warmed, *D. ponderosae* has expanded to

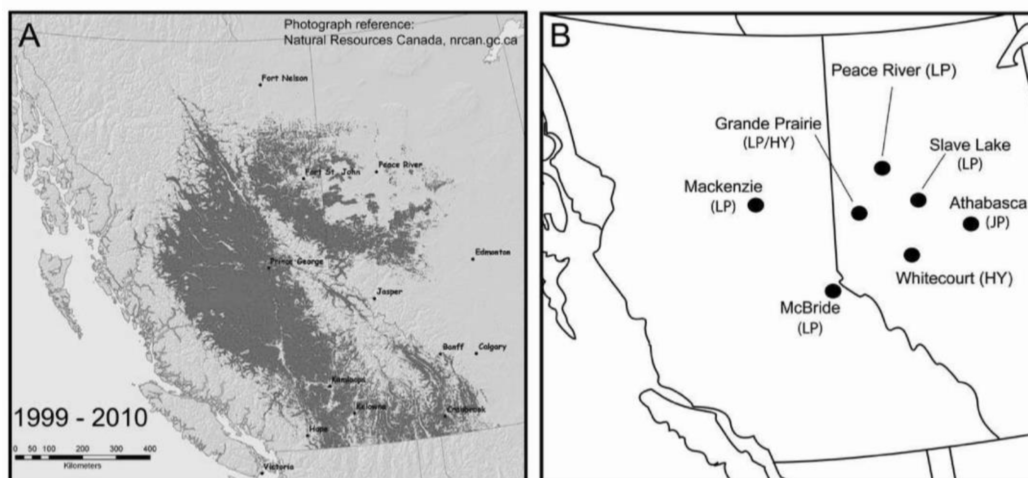



**Figure A3.1.** Outbreak and spread of *D. ponderosae* in western Canada from 1999-2010 (A), and sampling locations of beetles and galleries from lodgepole (LP) and hybrid (HY) pines (B).

higher latitudes (9-11), and breached the geophysical barrier of the Canadian Rocky Mountains (12) (Figure A3.1). It has now colonized the hybrid crossings of *P. contorta* and jack pine, *P. banksiana* Lamb., and also the pure *P. banksiana* in this region (13, 14). These trees are contiguous with *P. banksiana* forests throughout all of southern Canada east of the Rocky Mountains, connecting with eastern white pine, *Pinus strobus L.,* and red pine, *Pinus resinosa* Sol. Ex Aiton, further to the east (15). These beetles therefore potentially threaten pines across North America.

An important feature of bark beetles is their reliance on symbiotic microbes, which likely assist with host colonization and utilization. Symbiotic relationships with fungi have been

studied for several decades (4, 16, 17) and are known to contribute to larval nutrition. Additionally, some bacterial symbionts have been implicated in defense against antagonistic fungi (18-20) and in nutrient acquisition (21). Moreover, beetle-associated microbes have demonstrated community structures that reflect both host beetle and geographic zones (22).

Both bacterial-beetle and bacterial-fungal-beetle relationships are mediated by host tree chemistry (18, 23). Pines synthesize terpenoids that are toxic at high concentrations to a broad range of insects, including bark beetles and their symbiotic fungi (24). These compounds are present in constitutive resin and phloem, and are synthesized and translocated in response to the early stages of beetle-microbial attack. Monoterpenes can kill or repel bark beetles, and both monoterpenes and diterpene acids can inhibit fungal symbionts (25-27). Terpenoid-based defenses of healthy trees confront beetles with a significant barrier during periods when their populations are low, and thus colonization is restricted to highly stressed hosts. However, several species, including *D. ponderosae*, can exhaust tree defenses through pheromone-mediated mass attack when population densities are high (7, 28).

An important question emerging from the current climate-driven range expansion of these beetles is how they will perform in new tree species and whether they will persist in a relatively endemic state, or engage in self-driving outbreaks (29-31). Here, we focus on the potential role that *D. ponderosae*-associated bacterial communities may play during the colonization of both lodgepole and hybrid pines. We first sampled beetles, galleries, and trees to broadly identify similarities and differences between the communities associated with these environments using Denaturing Gradient Gel Electrophoresis (DGGE). Then, using a community metagenomic approach, we provide a detailed analysis of the bacterial communities associated with *D. ponderosae* and their galleries in both native and hybrid tree hosts. Finally, we explore

the hypothesis that bacteria play a role in the detoxification of host tree defenses by specifically analyzing genes involved in terpene degradation encoded by these bacteria.

**Table A3.1**. Site location and sample sizes of the collection of *D. ponderosae*, galleries, and phloem of unattacked trees.

| Site Location | Host | Sample Type | Sample Size | Method of Community Analysis |
|---|---|---|---|---|
| Mackenzie, British Columbia | *P. contorta* | *D. ponderosae*<br>Gallery<br>Phloem | 26<br>26<br>7 | DGGE |
| McBride, British Columbia | *P. contorta* | Whole beetles<br>Gallery | 300<br>150 | Community Metagenomics |
| Grand Prairie, Peace River, & Slave Lake, Alberta | *P. contorta* | Whole beetles<br>Gallery<br>Phloem | 52<br>15<br>8 | DGGE |
| Whitecourt, Alberta | *P. contorta- P. banksiana* | Whole beetles<br>Gallery<br>Phloem | 41<br>13<br>24 | DGGE |
| Grand Prairie, Alberta | *P. contorta- P. banksiana* | *D. ponderosae*<br>Gallery | 300<br>150 | Community Metagenomics |
| Athabasca, Alberta | *P. banksiana* | Whole beetles<br>Gallery<br>Phloem | 53<br>8<br>18 | DGGE |

## A3.3 Materials and Methods

### A3.3.1 Sample Collection

We sampled *D. ponderosae* adults, their galleries, and unattacked tree phloem from four sites in 2010 across a gradient ranging from historical to naïve tree species of *P. contorta* (Mackenzie, British Columbia and Grande Prairie, Peace River, and Slave Lake, Alberta), hybrid *P. contorta - P. banksiana* (Whitecourt, Alberta), and *P. banksiana* (Athabasca, Alberta) (Table A3.1, Figure A3.1). For DGGE analysis, 291 beetle, gallery, and unattacked tree samples were obtained from multiple sites in Alberta and British Columbia, and for the community metagenomes, sampling consisted of approximately 300 adult beetles and 150 galleries from 20 trees from pines located at each of two sites in July of 2009. One site was located near McBride, British Columbia in a *P. contorta* stand, and the other was near Grande Prairie, Alberta in a mixed *P. contorta* and hybrid *P. contorta - P. banksiana* stand (32) (Table A3.1, Figure A3.1). At the time of sampling, *D. ponderosae* were not observed colonizing *P. banksiana*.

For both analyses, *D. ponderosae* adults and 0.5 cm x 2 cm sections of phloem that included egg galleries and frass were removed from recently attacked trees and placed on ice before being transported to the laboratory at the University of Wisconsin-Madison. Phloem tissue from trees with no visible signs of bark beetle attack or disease was also sampled for DGGE analysis, with a maximum of three beetles and galleries and two phloem samples from a single tree, and no samples from the same gallery. Samples for the community metagenomes were collected in the same way and pooled after collection for a total of four samples representing beetles and galleries from *P. contorta* and hybrid *P. contorta-P. banksiana* stands.

**Table A3.2.** Summary of sequencing statistics for the four *D. ponderosae*-associated metagenomes.

| Statistic | ***D. ponderosae*** **(Hybrid Pine)** | ***D. ponderosae*** **Gallery (Hybrid Pine)** | ***D. ponderosae*** **(Lodgepole Pine)** | ***D. ponderosae*** **Gallery (Lodgepole Pine)** |
|---|---|---|---|---|
| Size of assembly (Mbp) | 29 | 30.2 | 61.3 | 32.7 |
| # of protein-coding genes | 42,426 | 53,181 | 86,776 | 66,202 |
| # of contigs | 9,161 | 9,616 | 20,178 | 10,665 |
| # of singleton reads | 24,220 | 47,597 | 56,588 | 60,200 |
| N50 contig size (bp) | 1,986 | 606 | 1,535 | 482 |
| Largest contig (bp) | 142,661 | 59,558 | 140,467 | 47,283 |

### A3.3.2 Community Metagenome Sequencing and Analysis

Total bacterial genomic DNA was extracted from all four pooled samples. Eukaryotic material was removed through a series of washes, as previously described (33). Briefly, materials were ground with mortar and pestle and washed in 1X Phosphate Buffered Saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH2PO4) containing 0.1% Tween. Each sample was shaken for 15 minutes at 150 rpm and centrifuged for 5 minutes at 500 rpm. This resulted in a 3-layer mixture containing (from the top) bacterial cells, fungal material, and insect material. The bacterial layer was removed and subjected to this process two more times. The final bacterial suspensions were centrifuged for 30 minutes at 4,100 rpm, re-suspended in 1X Phosphate Buffer Solution (PBS) containing 0.1% Tween and filtered through a 100 um filter.

Total DNA was then extracted using a Qiagen DNeasy Plant Maxi Kit (Qiagen Sciences, Germantown, MD, USA).

Community metagenomes were subsequently generated from the extracted DNA using Roche 454 Titanium pyrosequencing (34). These reads were assembled using Newbler v. 2.1 with default parameters, and phylogenetic binning of all contigs and singleton reads of the resulting assemblies was performed using BLASTN (35) and PhymmBL (36). All contigs and singletons were first compared to a reference dataset containing all completely sequenced bacterial and archael genomes available in NCBI as of 01/01/2012. All contigs having BLASTN hits with e-values < 1e-10 were classified according to their best hit, while all other contigs were subsequently classified using PhymmBL. Contigs with no BLASTN matches that were classified by PhymmBL with confidence scores < 50 were considered unclassified. Proteins were predicted from these metagenomes using the IMG-ER pipeline (37).

Predicted proteins putatively involved in terpene degradation were identified from each metagenome using either the KEGG Automatic Annotation Server (KAAS) (38) or BLASTP (35) against a custom dataset of proteins belonging to the *dit* diterpene acid degradation pathway. Specifically, genes encoding enzymes putatively involved in monoterpene degradation, as defined in the KEGG Limonene and Pinene Degradation Pathway (ko00903), were annotated and retained. For the diterpene acid degradation pathway, predicted proteins in the metagenomes were compared to the proteins comprising the *dit* gene cluster, as described for *Pseudomonas abietaniphila* BKME-9 (39), using BLASTP (e-value 1e-5). To identify if terpene-degrading genes in the beetle-associated communities were enriched compared to other plant biomass-degrading communities, we performed the same KAAS- and BLASTP-based annotations on publicly-available metagenomes associated with the termite hindgut (40), wallaby foregut (41),

cow rumen (42), panda gut (43), leaf-cutter ant fungus garden (44), switchgrass compost (45), and poplar biomass bioreactor (46). Using the raw number of proteins annotated in the KEGG Limonene and Pinene Degradation pathway and the *dit* operon, and the total number of proteins encoded in these metagenomes, we performed Fisher's Exact Test to identify potential enrichment of these genes in the beetle-associated metagenomes ($P < 0.05$). Fisher's Exact Test takes into account the total number of proteins in the datasets being compared, and is therefore appropriate for identifying relative changes in coding potential. The BLASTN- and PhymmBL-based phylogenetic binning results were then used to identify the taxonomic origin of the contigs and singletons encoding these proteins.

### A3.3.3 Data Availability

Raw pyrosequencing data for the community metagenomes have been deposited in the National Center for Biotechnological Information's (NCBI) Short Read Archive under accession numbers SRA4088237, SRA4088238, SRA4088239, and SRA4088241. Assembled community metagenomes are on the Joint Genome Institute's Integrated Microbial Genomes/Microbiomes (IMG/M) database (47) under project IDs 2032320008, 2032320009, 2029527007, 2035918003. Sequences of all DGGE bands obtained in this study are available in GenBank under accession numbers JF810915- JF810926.

## A3.4 Results

### A3.4.1 Microbial Community Analysis

Principle component analysis of the DGGE data revealed no distinct clustering among beetle, gallery, or tree samples from either location or tree species. We sequenced bands corresponding to the 12 most commonly identified Operational Taxonomic Units (OTUs) in our

DGGE analyses and found that 10 were most similar to Gammaproteobacteria, one to Betaproteobacteria, and one to Actinobacteria.

**A3.4.2 Community Metagenome Sequencing and Phylogenetic Binning.**

The four metagenomes generated using shotgun 454 pyrosequencing comprised 27.1 – 58.8 Mbp of sequence after assembly (Table A3.2). These were constructed from (a) whole *D. ponderosae* beetles from infested *P. contorta*; (b) gallery material from infested *P. contorta*; (c) whole *D. ponderosae* beetles from infested hybrid *P. contorta*-*P. banksiana*; and (d) gallery material from infested hybrid *P. contorta*-*P. banksiana*. A phylogenetic binning analysis revealed similar taxonomic patterns across all four community metagenomes. Broadly, Gammaproteobacteria comprised the majority of all sequences in each sample, although Betaproteobacteria, Alphaproteobacteria, Firmicutes, and Actinobacteria were also represented. Analysis at the genus-level identified abundant sequences matching to *Pseudomonas, Rahnella, Serratia, Erwinia, Stenotrophomonas*, and *Pantoea*.

**A3.4.3 Analysis of Genes Putatively Involved in Terpene Degradation**

We performed a metabolic reconstruction analysis using the Kyoto Encylcopedia of Genes and Genomes (KEGG) database (48) (Table A3.3). An analysis of proteins involved in the Limonene and Pinene Degradation pathway yielded between 90 - 198 proteins in each community metagenome (Table A3.3). This comprises 0.17 - 0.27% of the predicted proteins in these datasets. Of the 20 enzymes comprising this pathway, 5 were found to be enriched in the pine beetle metagenomes when compared to metagenomes of other plant-biomass associated microbial communities (Table A3.3, Fisher's exact test, $P < 0.05$). These included an aldehyde dehydrogenase (EC:1.2.1.3), an oxidoreductase (EC:1.14.13.-), an enoyl-CoA hydratase

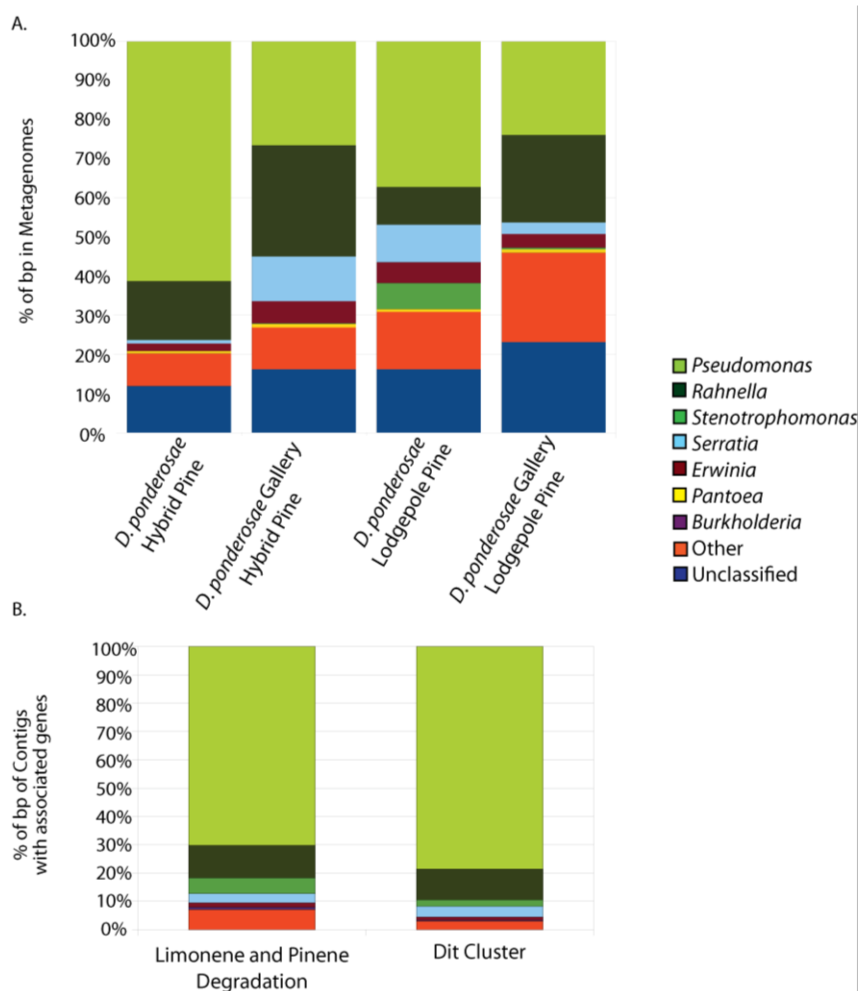



**Figure A3.2.** Results of genus-level phylogenetic binning of the four beetle-associated community metagenome assemblies (a). Phylogenetic binning of only contigs containing genes annotated in the KEGG Limonene and Pinene Degradation Pathway or *dit* cluster of *Pseudomonas abietaniphila* BKME-9 (b). Only those genera with greater than 1% representation are shown.

(EC:4.2.1.17), and two 3-hydroxyacyl-CoA epimerases (EC:1.1.1.35, 4.2.1.17, 5.1.2.3, 5.3.3.8). All enzymes involved in the conversion of α-pinene to 3-Isopropylbut-3-enoic acid, myrtenic acid, or pinocarvone, as described in KEGG, were detected. The beetle metagenome from hybrid pine contained the highest proportion of potential terpene degradation enzymes (113 proteins,

0.27% of proteins), while the gallery metagenome from hybrid pine contained the smallest (90 proteins, 0.17% of proteins).

We also investigated genes involved in diterpene degradation using the well-characterized *dit* gene cluster found in the diterpene degrading bacterium *Pseudomonas abientaniphila* BKME-9 (39, 49). Of the 20 proteins annotated as belonging to the *dit* gene cluster, 17 - 19 were present in each of the community metagenomes sequenced in our study (Figure A3.3a). Of these, 12 - 16 were found to be enriched compared to metagenomes of other plant biomass-associated microbial communities (Figure A3.3b, Fisher's exact test, $P < 0.05$). Between 459 - 920 proteins encoded by *dit* genes were identified in our metagenomes, which comprise 0.54 - 1.3% of the total proteins in the beetle-associated community metagenomes. Homologs of DitA1, an aromatic ring hydroxylating dioxygenase, were identified in all metagenomes (9 - 15 proteins). Moreover, homologs to the predicted dehydrogenases DitB, DitG, and DitI were among the most abundant proteins identified using this approach.

**A3.4.4 Terpene and Diterpene Taxonomic Classification**

We characterized the putative taxonomic origin of those genes involved in pinene or limonene degradation or belonging to the *dit* gene cluster in our metagenomes. Consistent with the general phylogenetic composition of the metagenomes, we found that the majority of these genes were classified as belonging to bacteria in the genera *Pseudomonas*, *Rahnella*, *Serratia*, and *Stenotophomonas* (Figures A3.2b). In all cases, sequences from *Pseudomonas* and *Rahnella* comprised over 60% of the genes identified as belonging to these pathways.

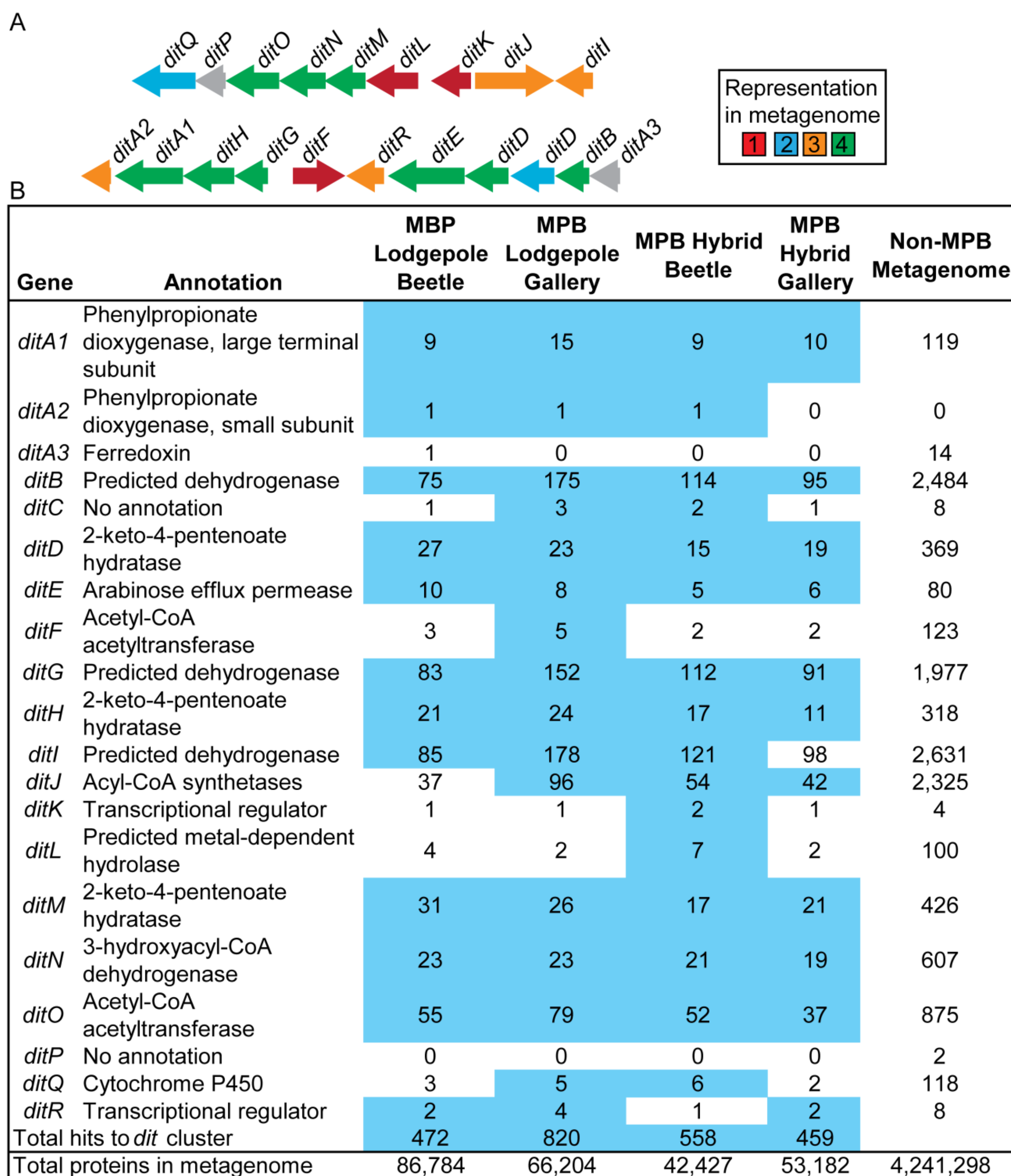
**Table A3.3.** Identification of genes involved in the limonene and pinene degradation pathways within four mountain pine beetle metagenomes and their comparison to other plant biomass degrading metagenomes. Genes enriched relative to other plant biomass degrading metagenomes are shaded (Fisher's exact test $P < 0.05$).

| Annotation | EC | KO | Beetle (Hybrid Pine) | Gallery (Hybrid Pine) | Beetle (Lodgepole Pine) | Gallery (Lodgepole Pine) | Non-MPB Combined |
|---|---|---|---|---|---|---|---|
| Oxidoreductases | 1.1.-.- | K00120 | 0 | 0 | 0 | 0 | 62 |
| Aldehyde dehydrogenase | 1.2.1.3 | K00128 | 43 | 37 | 91 | 80 | 1,485 |
| Oxidoreductases | 1.2.1.- | K00155 | 0 | 0 | 1 | 1 | 23 |
| Oxidoreductases | 1.14.13.- | K00492 | 12 | 7 | 17 | 13 | 96 |
| Oxidoreductases | 1.14.-.- | K00517 | 4 | 2 | 6 | 5 | 378 |
| Acyltransferases | 2.3.1.- | K00680 | 1 | 0 | 3 | 0 | 271 |
| Hydrolase | 3.1.2.- | K01076 | 0 | 0 | 0 | 0 | 25 |
| Enoyl-CoA hydratase | 4.2.1.17 | K01692 | 32 | 19 | 51 | 34 | 768 |
| Carbon-oxygen lyases | 4.2.1.- | K01726 | 0 | 0 | 0 | 0 | 0 |
| 3-hydroxyacyl-CoA dehydrogenase | 1.1.1.35; 4.2.1.17 | K01782 | 10 | 16 | 15 | 15 | 182 |
| 3-hydroxyacyl-CoA dehydrogenase | 1.1.1.35; 4.2.1.17 | K01825 | 10 | 8 | 14 | 18 | 56 |
| Ligase | 6.2.1.- | K01913 | 0 | 0 | 0 | 0 | 10 |
| (S)-limonene 6-monooxygenase | 1.14.13.48 | K07381 | 0 | 0 | 0 | 0 | 0 |
| (S)-limonene 7-monooxygenase | 1.14.13.49 | K07382 | 0 | 0 | 0 | 0 | 0 |
| Limonene-1 2-epoxide hydrolase. | 3.3.2.8 | K10533 | 0 | 0 | 0 | 0 | 0 |
| Trans-carveol dehydrogenase. | 1.1.1.275 | K12466 | 0 | 0 | 0 | 0 | 0 |
| Trans-carveol dehydrogenase | 1.1.1.243 | K14730 | 0 | 0 | 0 | 0 | 0 |
| Monoterpene hydrolase | 3.1.1.83 | K14731 | 0 | 0 | 0 | 0 | 0 |
| Limonene hydroxylase | 1.14.13.48; 1.14.13.49 | K14732 | 0 | 0 | 0 | 0 | 0 |
| Limonene 1,2-monooxygenase | 1.14.13.107 | K14733 | 1 | 1 | 0 | 0 | 49 |
| Total # of proteins | | | 42,427 | 53,182 | 86,784 | 66,204 | 4,241,298 |

**A3.5 Discussion**

This study provides the first community metagenomic analysis of a bark beetle. The results of both our community metagenomic and DGGE analyses indicate that Gammaproteobacteria are prevalent in both *D. ponderosae* beetles and their galleries from both lodgepole and hybrid lodgepole-jack pines. In particular, Gammaproteobacteria belonging to the genera *Pseudomonas, Stenotrophomonas, Erwinia*, and *Serratia* were particularly abundant in these metagenomes (Figure A3.2a), indicating these groups are consistently associated with *D. ponderosae* or their host trees. Our DGGE-based analysis did not resolve differences between these environments, indicating that bacterial communities may be broadly similar among all of them. Moreover, no distinct differences were observed in the composition of metagenomes from the beetle and gallery samples from Alberta and British Columbia. Taken together, this suggests that a relatively consistent bacterial community is associated with *D. ponderosae* and its microenvironment, and that the recent expansion of this insect's range will not be impeded by a lack of appropriate bacterial communities.

Conifers produce monoterpenes and diterpenes that are toxic to bark beetles and their fungal symbionts, both constitutively and in response to attack. We identified numerous bacterial genes associated with degradation of these compounds, including well-represented KEGG pathways for limonene and pinene degradation in each of the four metagenomes (Table A3.3). Moreover, we identified numerous genes homologous to the *dit* gene cluster of *Pseudomonas abietophila* BKME-9 (Figure A3.3a), which is known to be involved in diterpene degradation. We found a significantly higher proportion of these genes in our beetle metagenomes than those




| Gene | Annotation | MBP Lodgepole Beetle | MPB Lodgepole Gallery | MPB Hybrid Beetle | MPB Hybrid Gallery | Non-MPB Metagenome |
|---|---|---|---|---|---|---|
| *ditA1* | Phenylpropionate dioxygenase, large terminal subunit | 9 | 15 | 9 | 10 | 119 |
| *ditA2* | Phenylpropionate dioxygenase, small subunit | 1 | 1 | 1 | 0 | 0 |
| *ditA3* | Ferredoxin | 1 | 0 | 0 | 0 | 14 |
| *ditB* | Predicted dehydrogenase | 75 | 175 | 114 | 95 | 2,484 |
| *ditC* | No annotation | 1 | 3 | 2 | 1 | 8 |
| *ditD* | 2-keto-4-pentenoate hydratase | 27 | 23 | 15 | 19 | 369 |
| *ditE* | Arabinose efflux permease | 10 | 8 | 5 | 6 | 80 |
| *ditF* | Acetyl-CoA acetyltransferase | 3 | 5 | 2 | 2 | 123 |
| *ditG* | Predicted dehydrogenase | 83 | 152 | 112 | 91 | 1,977 |
| *ditH* | 2-keto-4-pentenoate hydratase | 21 | 24 | 17 | 11 | 318 |
| *ditI* | Predicted dehydrogenase | 85 | 178 | 121 | 98 | 2,631 |
| *ditJ* | Acyl-CoA synthetases | 37 | 96 | 54 | 42 | 2,325 |
| *ditK* | Transcriptional regulator | 1 | 1 | 2 | 1 | 4 |
| *ditL* | Predicted metal-dependent hydrolase | 4 | 2 | 7 | 2 | 100 |
| *ditM* | 2-keto-4-pentenoate hydratase | 31 | 26 | 17 | 21 | 426 |
| *ditN* | 3-hydroxyacyl-CoA dehydrogenase | 23 | 23 | 21 | 19 | 607 |
| *ditO* | Acetyl-CoA acetyltransferase | 55 | 79 | 52 | 37 | 875 |
| *ditP* | No annotation | 0 | 0 | 0 | 0 | 2 |
| *ditQ* | Cytochrome P450 | 3 | 5 | 6 | 2 | 118 |
| *ditR* | Transcriptional regulator | 2 | 4 | 1 | 2 | 8 |
| Total hits to *dit* cluster | | 472 | 820 | 558 | 459 | |
| Total proteins in metagenome | | 86,784 | 66,204 | 42,427 | 53,182 | 4,241,298 |

**Figure A3.3.** Representation of genes involved in diterpene degradation in mountain pine beetle- associated metagenomes. A) In the diterpene degradation gene cluster *dit*, each gene is colored according to its representation within the four mountain pine beetle metagenomes. B) Total copy numbers for each gene is shown, with those in blue enriched (Fisher's exact test, $P < 0.05$) relative to other plant biomass-degrading metagenomes.

from other plant biomass-associated microbial communities (Figure A3.3b). The biomass-degrading communities used for comparison originate from a diversity of different environments where plant toxins would be likely encountered, suggesting that bacteria associated with pine beetles are particularly well adapted to metabolize the aromatic plant toxins in their environment.

The majority of genes involved in terpene and diterpene degradation belong to bacteria in the genera *Pseudomonas* and *Rahnella* (Figures A3.2b). The genus *Pseudomonas* contains numerous species that degrade a wide range of aromatic compounds, including plant toxins, xenobiotics, and pollutants (39, 50-52). The biodegradative capacities of *Rahnella* isolates have been less intensively studied, but numerous *Enterobacteriaciae* closely related to this genus are known to degrade a diversity of aromatic compounds (53, 54). Our finding that these genera are both associated with *D. ponderosae* and possess numerous genes putatively involved in terpenoid degradation suggests they may contribute to *D. ponderosae's* ability to attack live conifers. Future transcriptomic and culture-based work confirming the ability of these bacteria to degrade plant toxins is required.

These results suggest several testable models to explain associations of bacterial communities with *D. ponderosae* and their galleries. One hypothesis is that *D. ponderosae* vector terpene-degrading microbiota between trees. Bark beetles are known to consistently vector both fungi (16) and several genera of bacteria, including *Streptomyces* (19, 55), thereby explaining the similar composition of all metagenomes analyzed here. This would also be consistent with the high representation of *Pseudomonas* and *Rahnella* genes associated with terpene degradation in all four *D. ponderosae*–associated metagenomes we described. A second non-exclusive model is that these communities are associated primarily with host trees rather than the beetles themselves. This is likewise supported by our DGGE results showing similarities among bacteria

in attacked and unattacked trees. Beetle colonization of a tree might induce proliferation of specific bacteria, such as *Pseudomonas* or *Rahnella,* that can exploit terpenoids and other carbon sources present in resin. Thus, even if not vectored by the beetles, colonization attempts and subsequent tree responses may create an environment that promotes the growth of terpenoid metabolizing bacteria. According to this model, resident microbial populations may influence bark beetles, with trees harboring fewer terpene-degrading bacteria posing more resistance to colonization.

This work provides insight into host colonization and range expansion of *D. ponderosae* by characterizing the microbiome associated with these beetles and host conifers. A combination of methods suggests that a relatively consistent bacterial community is associated with these beetles in lodgepole and hybrid lodgepole-jack pines. Our results identify bacteria of the genera *Pseudomonas* and *Rahnella* that may directly or indirectly contribute to the ability of beetles to overcome tree defenses. Future studies confirming and quantifying the ability of bacteria to degrade tree defenses are needed to understand influences they may have on bark beetle biology.

### A3.6 References

1. **Douglas, AE.** 2010. The symbiotic habit. Princeton University Press, Princeton, New Jersey, USA.

2. **Moran, NA.** 2006. Symbiosis. *Curr. Biol.* **16:**R866.

3. **Currie, CR.** 2001. A community of ants, fungi, and bacteria: a multilateral approach to studying symbiosis. *Annu. Rev. Microbiol.* **55:**357-380.

4. **Kleipzig, K, Adams, AS, Handelsman, J, Raffa, KF.** 2009. Symbioses, A key driver of insect physiologyical processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environ. Entomol.* **36:**67-77.

5. **Zilber-Rosenberg, I, and Rosenberg, E.** 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol. Rev.* **32:**723-35.

6. **Frago, E, Dicke, M, and Godfray, HC.** 2012. Insect symbionts as hidden players in insect-plant interactions. *Trends Ecol. Evol.* **27:**705-11.

7. **Raffa, KF, Aukema, BH, Bentz, BJ, Carroll, AL, Hicke, JA, Turner, MG, and Romme, WH.** 2008. Cross-scale drivers of natural disturbances prone to anthropogenic amplification: the dynamics of biome-wide bark beetle eruptions. *BioScience* **58:**501-517.

8. **Amman, G, and Cole, W.** 1983. Mountain pine beetle dynamics in lodgepole pine forests, part II: Population dynamics. *USDA For. Serv. Gen. Tech. Rep INT-145***:**59.

9. **Hicke, J, Logan, J, Powell, J, and Ojima, D.** 2006. Changing temperatures influence suitability for modeled mountain pine beetle (*Dendroctonus ponderosae*) outbreaks in the western United States. *J. Geophys. Res.* **111**.

10. **Bentz, B, Jaques, R, Fettig, C, Hansen, E, Hayes, J, Hicke, J, Kelsey, R, Negron, J, and Seybold, S.** 2010. Climate change and barke beetles of the western United States and Canada: direct and indirect effects. *BioScience* **60:**602-613.

11. **Sambaraju, K, Carrol, A, Zhu, J, Stahl, K, Moore, R, and Aukema, B.** 2012. Climate change could alter the distribution of mountain pine beetle outbreaks in western Canada. *Ecography* **35:**111-223.

12. **de la Giroday, H-MC, Carroll, AL, and Aukema, BH.** 2012. Breach of the northern Rocky Mountain geoclimatic barrier: initiation of range expansion by the mountain pine beetle. *J. Biogeogr.* **39:**1112-1123.

13. **Cullingham, CI, Cooke, JE, Dang, S, Davis, CS, Cooke, BJ, and Coltman, DW.** 2011. Mountain pine beetle host-range expansion threatens the boreal forest. *Mol. Ecol.* **20:**2157-71.

14. **Lusebrink, I, Evenden, ML, Blanchet, FG, Cooke, JE, and Erbilgin, N.** 2011. Effect of water stress and fungal inoculation on monoterpene emission from an historical and a new pine host of the mountain pine beetle. *J. Chem. Ecol.* **37:**1013-26.

15. **Safranyik, L, Carroll, AL, Régnière, J, Langor, DW, Riel, WG, Shore, TL, Peter, B, Cooke, BJ, Nealis, VG, and Taylor, SW.** 2010. Potential for range expansion of mountain pine beetle into the boreal forest of North America. *Can. Entomol.* **142:**415-442.

16. **Paine, TD, Raffa, KF, and Harrington, TC.** 1997. Interactions among scolytid bark beetles, their associated fungi, and host conifers. *Annu. Rev. Entomol.* **42:**179-206.

17. **Six, DL, and Wingfield, MJ.** 2011. The role of phytopathogenicity in bark beetle-fungus symbioses: a challenge to the classic paradigm. *Annu. Rev. Entomol.* **56:**255-72.

18. **Adams, AS, Currie, CR, Cardoza, YJ, Klepzig, KD, and Raffa, KF.** 2009. Effects of symbiotic bacteria and tree chemistry on the growth and reproduction of bark beetle fungal symbionts. *Can. J. Forest Res.* **39:**1133-1147.

19. **Scott, JJ, Oh, D-C, Yuceer, MC, Klepzig, KD, Clardy, J, and Currie, CR.** 2008. Bacterial Protection of Beetle-Fungus Mutualism. *Science* **322:**63.

20. **Cardoza, YJ, Klepzig, KD, and Raffa, KF.** 2006. Bacteria in oral secretions of an endophytic insect inhibit antagonistic fungi. *Ecol. Entomol.* **31:**636-645.

21. **Morales-Jimenez, J, Zuniga, G, Villa-Tanaca, L, and Hernandez-Rodriguez, C.** 2009. Bacterial community and nitrogen fixation in the red turpentine beetle, *Dendroctonus valens* LeConte (Coleoptera: Curculionidae: Scolytinae). *Microb. Ecol.* **58:**879-91.

22. **Adams, AS, Adams, SM, Currie, CR, Gillette, NE, and Raffa, KF.** 2010. Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae). *Environ. Entomol.* **39:**406-14.

23. **Adams, AS, Boone, CK, Bohlmann, J, and Raffa, KF.** 2011. Responses of bark beetle-associated bacteria to host monoterpenes and their relationship to insect life histories. *J. Chem. Ecol.* **37:**808-17.

24. **Keeling, C, and Bohlmann, J.** 2006. Diterpene resin acids in conifers. *Phytochemistry* **67:**2415-2423.

25. **Raffa, K, and Smalley, E.** 1995. Interaction of pre-attack and induced monoterpene concentrations in host conifer defense against bark beetle-fungal complexes. *Oecologia* **102:**285-295.

26. **Wallin, K, and Raffa, KF.** 2000. Influences of host chemicals and internal physiology on the multiple steps of postlanding host acceptance behavior of *Ips pini* (Coleoptera:Scolytidae). *Environ. Entomol.* **29:**442-453.

27. **Kopper, B, Illman, B, Kersten, P, Klepzig, K, and Raffa, K.** 2005. Effects of diterpene acids on components of a conifer bark beetle-fungal interaction: tolerance by *Ips pini* and sensitivity by its associate *Ophiostoma ips*. *Environ. Entomol.* **34:**486-493.

28. **Boone, C, Aukema, B, Bohlmann, J, Carroll, A, and Raffa, K.** 2011. Efficacy of tree defense physiology varies with bark beetle population density: a basis for positive feedback in eruptive species. *Can. J. For. Res.* **41:**1174-1188.

29. **Raffa, KR, Powell, EN, and Townsend, PA.** 2012. Temperature-driven range expansion of an irruptive insect heightened by weakly coevolved plant defenses. *Proc. Natl. Acad. Sci. USA* **110:**2193-8.

30. **Clark, EL, Carroll, AL, and Huber, DPW.** 2010. Differences in the constitutive terpene profile of lodgepole pine across a geographical range in British Columbia, and correlation with historical attack by mountain pine beetle. *Can. Entomol.* **142:**557-573.

31. **Cudmore, TJ, Björklund, N, Carroll, AL, and Lindgren, BS.** 2010. Climate change and range expansion of an aggressive bark beetle: evidence of higher beetle reproduction in naïve host tree populations. *J. Appl. Ecol.* **47:**1036–1043.

32. **Bleiker, K, and Carroll, A.** 2011. Rating introgression between lodgepole and jack pine at the individual tree level using morphological traits. *North J. Appl. For.* **28:**138-145.

33. **Holben, WE, Noto, K, Sumino, T, and Suwa, Y.** 1998. Molecular analysis of bacterial communities in a three-compartment granular activated sludge system indicates community-level control by incompatible nitrification processes. *Appl. Environ. Microbiol.* **64:**2528-32.

34. **Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, YJ, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, ML, Jarvie, TP, Jirage, KB, Kim, JB, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, McDade, KE, McKenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth, GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A, Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF, and Rothberg, JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-80.

35. **Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:**403-410.

36. **Brady, A, and Salzberg, SL.** 2009. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat. Methods* **6:**673-6.

37. **Markowitz, VM, Mavromatis, K, Ivanova, NN, Chen, IM, Chu, K, and Kyrpides, NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:**2271-8.

38. **Moriya, Y, Itoh, M, Okuda, S, Yoshizawa, AC, and Kanehisa, M.** 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35:**W182-5.

39. **Martin, VJ, Yu, Z, and Mohn, WW.** 1999. Recent advances in understanding resin acid biodegradation: microbial diversity and metabolism. *Arch. Microbiol.* **172:**131-8.

40. **Warnecke, F, Luginbuhl, P, Ivanova, N, Ghassemian, M, Richardson, TH, Stege, JT, Cayouette, M, McHardy, AC, Djordjevic, G, Aboushadi, N, Sorek, R, Tringe, SG, Podar, M, Martin, HG, Kunin, V, Dalevi, D, Madejska, J, Kirton, E, Platt, D, Szeto, E, Salamov, A, Barry, K, Mikhailova, N, Kyrpides, NC, Matson, EG, Ottesen, EA, Zhang, X, Hernandez, M, Murillo, C, Acosta, LG, Rigoutsos, I, Tamayo, G, Green, BD, Chang, C, Rubin, EM, Mathur, EJ, Robertson, DE, Hugenholtz, P, and Leadbetter, JR.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:**560-565.

41. **Pope, PB, Denman, SE, Jones, M, Tringe, SG, Barry, K, Malfatti, SA, McHardy, AC, Cheng, JF, Hugenholtz, P, McSweeney, CS, and Morrison, M.** 2010. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. *Proc. Natl. Acad. Sci. USA* **107:**14793-8.

42. **Hess, M, Sczyrba, A, Egan, R, Kim, TW, Chokhawala, H, Schroth, G, Luo, S, Clark, DS, Chen, F, Zhang, T, Mackie, RI, Pennacchio, LA, Tringe, SG, Visel, A, Woyke, T, Wang, Z, and Rubin, EM.** 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331:**463-467.

43. **Zhu, L, Wu, Q, Dai, J, Zhang, S, and Wei, F.** 2011. Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc. Natl. Acad. Sci. USA* **108:**17714-9.

44. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **9:**1688-701.

45. **Allgaier, M, Reddy, A, Park, JI, Ivanova, N, D'Haeseleer, P, Lowry, S, Sapra, R, Hazen, TC, Simmons, BA, VanderGheynst, JS, and Hugenholtz, P.** 2009. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* **5:**e8812.

46. **van der Lelie, D, Taghavi, S, McCorkle, SM, Li, LL, Malfatti, SA, Monteleone, D, Donohoe, BS, Ding, SY, Adney, WS, Himmel, ME, and Tringe, SG.** The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS One* **7:**e36740.

47. **Markowitz, VM, Ivanova, NN, Szeto, E, Palaniappan, K, Chu, K, Dalevi, D, Chen, IM, Grechkin, Y, Dubchak, I, Anderson, I, Lykidis, A, Mavromatis, K, Hugenholtz, P, and Kyrpides, NC.** 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36:**D534-8.

48. **Kanehisa, M, Araki, M, Goto, S, Hattori, M, Hirakawa, M, Itoh, M, Katayama, T, Kawashima, S, Okuda, S, Tokimatsu, T, and Yamanishi, Y.** 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36:**D480-4.

49. **Mohn, WW, Wilson, AE, Bicho, P, and Moore, ER.** 1999. Physiological and phylogenetic diversity of bacteria growing on resin acids. *Syst. Appl. Microbiol.* **22:**68-78.

50. **Kanaly, RA, and Harayama, S.** 2000. Biodegradation of high-molecular-weight polycyclic aromatic hydrocarbons by bacteria. *J. Bacteriol.* **182:**2059-67.

51. **Whyte, LG, Bourbonniere, L, and Greer, CW.** 1997. Biodegradation of petroleum hydrocarbons by psychrotrophic *Pseudomonas* strains possessing both alkane (alk) and naphthalene (nah) catabolic pathways. *Appl. Environ. Microbiol.* **63:**3719-23.

52. **Foght, JM, and Westlake, DW.** 1988. Degradation of polycyclic aromatic hydrocarbons and aromatic heterocycles by a *Pseudomonas* species. *Can. J. Microbiol.* **34:**1135-41.

53. **Sarma, PM, Bhattacharya, D, Krishnan, S, and Lal, B.** 2004. Degradation of polycyclic aromatic hydrocarbons by a newly discovered enteric bacterium, *Leclercia adecarboxylata*. *Appl. Environ. Microbiol.* **70:**3163-6.

54. **Wongsa, P, Tanaka, M, Ueno, A, Hasanuzzaman, M, Yumoto, I, and Okuyama, H.** 2004. Isolation and characterization of novel strains of *Pseudomonas aeruginosa* and *Serratia marcescens* possessing high efficiency to degrade gasoline, kerosene, diesel oil, and lubricating oil. *Curr. Microbiol.* **49:**415-22.

55. **Hulcr, J, Adams, AS, Raffa, K, Hofstetter, RW, Klepzig, KD, and Currie, CR.** 2011. Presence and diversity of *Streptomyces* in *Dendroctonus* and sympatric bark beetle galleries across North America. *Microb. Ecol.* **61:**759-68.

# Appendix 4: Complete Genome of *Enterobacteriaceae* Bacterium Strain FGI 57, a Strain Associated with Leaf-cutter Ant Fungus Gardens

**Previously Published:** Frank O Aylward, Daniel M Tremmel, David C Bruce, Patrick Chain, Amy Chen, Karen W Davenport, Chris Detter, Cliff S Han, James Han, Marcel Huntemann, Natalia N Ivanova, Nikos C Kyrpides, Victor Markowitz, Kostas Mavrommatis, Matt Nolan, Ioanna Pagani, Amrita Pati, Sam Pitluck, Shweta Deshpande, Lynne Goodwin, Tanja Woyke, Cameron R Currie (2013). Complete Genome of *Enterobacteriaceae* sp. Strain FGI 57, a Strain Associated with Leaf-Cutter Ant Fungus Gardens. *Genome Announcements*, 1(1).

**Author Contributions**: For this chapter I performed all genomic analyses. The genome was sequenced by the Joint Genome Institute. I wrote the paper with Cameron R. Currie. Daniel M. Tremmel assisted with DNA extraction.

## A4.1 Abstract

*Enterobacteriaceae* bacterium strain FGI 57 was isolated from a fungus garden of the leaf-cutter ant *Atta colombica*. Analysis of its single 4.76 Mbp chromosome will shed light on community dynamics and plant biomass degradation in ant fungus gardens.

## A4.2 Genome Announcement

*Enterobacteriaceae* bacterium strain FGI 57 was isolated in 2009 from a fungus garden of the leaf-cutter ant *Atta colombica* on Pipeline Road in Panama. Leaf-cutter ants are dominant herbivores in the Neotropics that derive nutrition from plant biomass by culturing specialized fungus gardens on foliar material (1). Rather than consuming the plant material directly, they use it as manure to cultivate specialized fungus gardens composed primarily of the obligate fungal symbiont *Leucoagaricus gongylophorus* (2). Recent work has shown that bacteria within the family *Enterobactereaceae* and genus *Pseudomonas* also inhabit fungus gardens (3-5), and metaproteomic investigations of these environments have identified several Strain FGI 57 proteins (3), suggesting that this bacterium is an active member of the fungus garden community.

The genome of *Enterobacteriaceae* bacterium Strain FGI 57 was generated at the DOE Joint genome Institute (JGI) using a combination of Illumina (6) and 454 technologies (7). For this genome we constructed and sequenced an Illumina GAii shotgun library (46,058,996 reads) totaling 4,605.9 Mb together with both a 454 Titanium standard library (110,558 reads) and a paired end 454 library (269,554 reads, average insert size of 7 kb) totaling 110.8 Mb. Assemblies were generated using a combination of Newbler, version 2.3-PreRelease-6/30/2009, Vevet, version 1.0.13 (8), and parallel phrap, version SPS - 4.24 (High Performance Software, LLC), and Consed was used in the finishing process (9-11). Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher (12), or by sequencing cloned bridging PCR fragments. Gaps between contigs were closed by editing in Consed, and by PCR and Bubble PCR (J-F Cheng, unpublished data) primer walks. A total of 38 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/.

The complete genome of *Enterobacteriaceae* bacterium strain FGI 57 comprises a single circular chromosome of 4.76 Mpb with an average of 228.5-fold coverage and a G+C content of 54.0%. Annotation of the finished chromosome was performed using the IMG-ER pipeline (13). A total of 9 copies 16S rDNA, 86 tRNAs, and 4,548 protein-coding genes were identified in this way. This bacterium was originally classified in the genus *Cronobacter*, but BLAST-based comparison (14) of its 16S rDNA genes with other *Enterobacteriaceae* indicate that *Raoultella planticola, Kluyvera georgiana,* and *Citrobacter freundii* all have high nucleic acid identity (all

98%). Phylogenetic analysis of the 16S rDNA genes using FastTree (15) failed to resolve the placement of this bacterium, suggesting that Strain FGI 57 may represent a novel group within the *Enterobacteriaceae*. The complete genome of this bacterium will facilitate future investigations of the symbiotic microbial community residing in leaf-cutter ant fungus gardens.

### A4.3 Nucleotide Sequence Accession Numbers

The complete genome sequence of *Enterobacteriaceae* bacterium strain FGI 57 has been deposited at DDBJ/EMBL/GenBank under the accession number CP003938.

### A4.4 References

1. **Hölldobler, B, and Wilson, EO.** 2010. The leafcutter ants: civilization by instinct. W. W. Norton and Company, Inc., New York, NY.

2. **Weber, NA.** 1966. Fungus-growing ants. *Science* **153:**587-604.

3. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **6:**1688–1701.

4. **Aylward, FO, Currie, CR, and Suen, G.** 2012. The Evolutionary Innovation of Nutritional Symbioses in Leaf-cutter Ants. *Insects* **3:**41-61.

5. **Suen, G, Scott, JJ, Aylward, FO, Adams, SM, Tringe, SG, Pinto-Tomas, AA, Foster, CE, Pauly, M, Weimer, PJ, Barry, KW, Goodwin, LA, Bouffard, P, Li, L, Osterberger, J, Harkins, TT, Slater, SC, Donohue, TJ, and Currie, CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6:**e1001129.

6. **Bennett, S.** 2004. Solexa Ltd. *Pharmacogenomics* **5:**433-8.

7. **Margulies, M, Egholm, M, Altman, WE, Attiya, S, Bader, JS, Bemben, LA, Berka, J, Braverman, MS, Chen, YJ, Chen, Z, Dewell, SB, Du, L, Fierro, JM, Gomes, XV, Godwin, BC, He, W, Helgesen, S, Ho, CH, Irzyk, GP, Jando, SC, Alenquer, ML, Jarvie, TP, Jirage, KB, Kim, JB, Knight, JR, Lanza, JR, Leamon, JH, Lefkowitz, SM, Lei, M, Li, J, Lohman, KL, Lu, H, Makhijani, VB, McDade, KE, McKenna, MP, Myers, EW, Nickerson, E, Nobile, JR, Plant, R, Puc, BP, Ronan, MT, Roth, GT, Sarkis, GJ, Simons, JF, Simpson, JW, Srinivasan, M, Tartaro, KR, Tomasz, A,**

**Vogt, KA, Volkmer, GA, Wang, SH, Wang, Y, Weiner, MP, Yu, P, Begley, RF, and Rothberg, JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:**376-80.

8. **Zerbino, DR, and Birney, E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18:**821-9.

9. **Ewing, B, Hillier, L, Wendl, MC, and Green, P.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:**175-85.

10. **Gordon, D, Abajian, C, and Green, P.** 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8:**195-202.

11. **Ewing, B, and Green, P.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8:**186-94.

12. **Han, CS, and Chain, P.** 2006. Finishing repeat regions automatically with Dupfinisher. *In* Arabnia. H. R. and V. H. (ed.), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology. CSREA Press, Las Vegas, NV.

13. **Markowitz, VM, Mavromatis, K, Ivanova, NN, Chen, IM, Chu, K, and Kyrpides, NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:**2271-8.

14. **Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:**403-410.

15. **Price, MN, Dehal, PS, and Arkin, AP.** 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* **5:**e9490.

# Appendix 5: Complete Genome of *Serratia* sp. Strain FGI 94, a Strain Associated with Leaf-cutter Ant Fungus Gardens

**Previously Published:** Frank O Aylward, Daniel M Tremmel, Gabriel J Starrett, David C Bruce, Patrick Chain, Amy Chen, Karen W Davenport, Chris Detter, Cliff S Han, James Han, Marcel Huntemann, Natalia N Ivanova, Nikos C Kyrpides, Victor Markowitz, Kostas Mavrommatis, Matt Nolan, Ioanna Pagani, Amrita Pati, Sam Pitluck, Hazuki Teshima, Shweta Deshpande, Lynne Goodwin, Tanja Woyke, Cameron R Currie (2013). Complete Genome of Serratia sp. Strain FGI 94, a Strain Associated with Leaf-Cutter Ant Fungus Gardens. Genome Announcements, 1(2).

**Author Contributions**: For this chapter I performed all genomic analyses. The genome was sequenced by the Joint Genome Institute. I wrote the paper with Cameron R. Currie. Daniel M. Tremmel and Gabriel J. Starrett assisted with DNA extraction.

## A5.1 Abstract

***Serratia* sp. Strain FGI 94 was isolated from a fungus garden of the leaf-cutter ant *Atta colombica*. Analysis of its 4.86 Mbp chromosome will help advance our knowledge of symbiotic interactions and plant biomass degradation in this ancient ant-fungus mutualism.**

## A5.2 Genome Announcement

The genus *Serratia* is a subgroup within the *Enterobacteriaceae* comprising isolates known to inhabit a variety of aquatic, terrestrial, and host-associated habitats (1). *Serratia* species are commonly associated with animals and plants, and some strains have been shown to cause nosocomial infections in humans (1, 2). Members of this genus are prevalent symbionts of insects, and both beneficial (3) and pathogenic (4) associations have been described.

*Serratia* sp. strain FGI 94 was isolated in 2009 from a fungus garden of the leaf-cutter ant *Atta colombica* near Pipeline Road, Panama. Leaf-cutter ants are dominant herbivores in the Neotropics that use the fresh foliar biomass they harvest to culture specialized fungus-bacteria gardens for food (5). Although these gardens are composed primarily of the obligate fungal

symbiont *Leucoagaricus gongylophorus* (6), metagenomic studies have identified numerous lineages of *Enterobactereaceae* and *Pseudomonas* that also inhabit fungus gardens (7-9). Moreover, nitrogen fixation has been shown to take place in these gardens, and bacteria of the genera *Klebsiella* and *Pantoea* have been implicated in the enrichment of the carbon-rich forage of the ants using this process (10).

The complete genome of *Serratia* sp. Strain FGI 94 was sequenced at the DOE Joint Genome Institute (JGI) using Illumina technology (11). Details of library construction, sequencing, and assembly can be found on the JGI website (http://www.jgi.doe.gov/). Two paired-end libraries with average insert sizes of 230 bp (17,979,030 reads, standard paired-end) and 7,839 bp (19,386,708 reads, CLIP paired-end (12)) were constructed using an Illumina HiSeq 2000, and initial assemblies were generated using a combination of Velvet, version 1.1.05 (13) and Allpaths, version r38445 (14). Consensus sequences were then computationally shredded into fake reads and integrated with a subset of the CLIP paired-end reads using parallel phrap, version 4.24 (High Performance Software, LLC). Possible mis-assemblies were corrected with manual editing in Consed (15-17). Gap closure was accomplished using repeat resolution software (Wei Gu, unpublished). Additionally, 21 PCR PacBio consensus sequences (18) were completed to close gaps and to raise the quality of the final sequence. The genome of this bacterium comprises a single circular chromosome of 4.86 Mpb with an average of 954-fold coverage and a G+C content of 58.9%.

Annotation of the finished chromosome was performed using the IMG-ER pipeline (19). A total of 7 copies 16S rDNA, 83 tRNAs, and 4,434 protein-coding genes were identified in this way. Comparison of the 16S rDNA genes with the NCBI 16S RNA database using BLASTN (20) revealed the invasive pathogen *S. rubidaea* strain JCM1240 to have the highest nucleic acid

identity (99%), with the next highest matches belonging to *S. marcescens* subsp. *sakuensis* (98%) and *S. nematodiphila* strain DZ0503SBS1 (98%).

### A5.3 Nucleotide Sequence Accession Numbers

The complete genome sequence of *Serratia* sp. Strain FGI94 has been deposited at DDBJ/EMBL/GenBank under the accession number CP003942.

### A5.4 References

1. **Grimont, F, and Grimont, PAD.** 2006. The Genus *Serratia*. *In* Martin Dworkin, Stanley Falkow, Eugene Rosenberg, Karl-Heinz Schleifer, and E. Stackebrandt (ed.), The Prokaryotes: A Handbook on the Biology of Bacteria, vol. 3. Springer Science, New York, NY.

2. **Mahlen, SD.** 2011. *Serratia* infections: from military experiments to current practice. *Clin. Microbiol. Rev.* **24:**755-91.

3. **Burke, GR, and Moran, NA.** 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* **3:**195-208.

4. **Vallet-Gely, I, Lemaitre, B, and Boccard, F.** 2008. Bacterial strategies to overcome insect defences. *Nat. Rev. Microbiol.* **6:**302-13.

5. **Hölldobler, B, and Wilson, EO.** 2010. The leafcutter ants: civilization by instinct. W. W. Norton and Company, Inc., New York, NY.

6. **Weber, NA.** 1966. Fungus-growing ants. *Science* **153:**587-604.

7. **Aylward, FO, Burnum, KE, Scott, JJ, Suen, G, Tringe, SG, Adams, SM, Barry, KW, Nicora, CD, Piehowski, PD, Purvine, SO, Starrett, GJ, Goodwin, LA, Smith, RD, Lipton, MS, and Currie, CR.** 2012. Metagenomic and metaproteomic insights into leaf-cutter ant fungus gardens. *ISME J.* **6:**1688–1701.

8. **Suen, G, Scott, JJ, Aylward, FO, Adams, SM, Tringe, SG, Pinto-Tomas, AA, Foster, CE, Pauly, M, Weimer, PJ, Barry, KW, Goodwin, LA, Bouffard, P, Li, L, Osterberger, J, Harkins, TT, Slater, SC, Donohue, TJ, and Currie, CR.** 2010. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet.* **6:**e1001129.

9. **Aylward, FO, Currie, CR, and Suen, G.** 2012. The Evolutionary Innovation of Nutritional Symbioses in Leaf-cutter Ants. *Insects* **3:**41-61.

10. **Pinto-Tomas, AA, Anderson, MA, Suen, G, Stevenson, DM, Chu, FS, Cleland, WW, Weimer, PJ, and Currie, CR.** 2009. Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* **326:**1120-1123.

11. **Bennett, S.** 2004. Solexa Ltd. *Pharmacogenomics* **5:**433-8.

12. **Peng, Z, Zhao, Z, Nath, N, Froula, JL, Clum, A, Zhang, T, Cheng, JF, Copeland, AC, Pennacchio, LA, and Chen, F.** 2012. Generation of long insert pairs using a Cre-LoxP Inverse PCR approach. *PLoS One* **7:**e29437.

13. **Zerbino, DR, and Birney, E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18:**821-9.

14. **Butler, J, MacCallum, I, Kleber, M, Shlyakhter, IA, Belmonte, MK, Lander, ES, Nusbaum, C, and Jaffe, DB.** 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. *Genome Res.* **18:**810-20.

15. **Ewing, B, and Green, P.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8:**186-94.

16. **Ewing, B, Hillier, L, Wendl, MC, and Green, P.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:**175-85.

17. **Gordon, D, Abajian, C, and Green, P.** 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8:**195-202.

18. **Zhang, X, Davenport, KW, Gu, W, Daligault, HE, Munk, AC, Tashima, H, Reitenga, K, Green, LD, and Han, CS.** 2012. Improving genome assemblies by sequencing PCR products with PacBio. *Biotechniques* **53:**61-2.

19. **Markowitz, VM, Mavromatis, K, Ivanova, NN, Chen, IM, Chu, K, and Kyrpides, NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:**2271-8.

20. **Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ.** 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:**403-410.