

**Ecology and Evolution of Microbial Lignocellulose Degradation in  
Insect-Associated Niches**

By

Gina Rachel Lewin

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

Doctor of Philosophy  
(Microbiology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 06/10/2016

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

**Brain G. Fox**, Professor, Biochemistry

**Paul J. Weimer**, Bacteriology, USDA Dairy Forage Research Center

## **Dissertation Abstract**

Ecology and Evolution of Microbial Lignocellulose Degradation in

Insect-Associated Niches

By Gina Rachel Lewin

Under the supervision of Professor Cameron R. Currie

at the University of Wisconsin-Madison

All life on Earth is fundamentally shaped through interactions with other organisms. However, most of our knowledge of microbes comes from studying model bacteria in the lab in isolation. To fully understand the evolution and ecology of traits in bacterial populations, it is necessary to study these organisms with a focus on their interactions with other lineages. Complex communities of microbes in soil and in association with herbivores degrade cellulose and other polymers within plant biomass, a key step in the global carbon cycle. In this dissertation, I analyze the evolutionary influence and ecological influence interactions on cellulose degradation. In Chapter 1, I review the ecology and evolution of the bacterial phylum Actinobacteria, including its role in plant biomass degradation, and I discuss the potential to harness the metabolic capabilities of Actinobacteria for the production of bioproducts. In Chapter 2, I present data showing two phylogenetically related strains of *Streptomyces* isolated from herbivorous insects share a conserved mechanism for cellulose degradation. I expand on this project in Chapter 3 by demonstrating that across the genus *Streptomyces*, cellulolytic ability is restricted to select clades of host-associated strains. Data support that cellulolytic ability evolved

through the selective retention of horizontally transferred genes and an expansion of the cellulase regulatory circuit, potentially driven by selective pressure from herbivorous hosts. In Chapter 4, I demonstrate that the ability of bacterial communities from leaf-cutter ant refuse dumps to degrade cellulose differs between ant colonies and corresponds with differences in taxonomic diversity. These results indicate that interactions contribute to the ability of a community to degrade cellulose. To further investigate this hypothesis, in Chapter 5, I enriched bacterial communities from leaf-cutter ant refuse dumps on cellulose for up to 60 transfers. In these cultures, an increased ability to degrade cellulose over time was reflected by corresponding shifts in community structure. Additionally, increased competition for the sugars released by cellulose breakdown correlated with a slower and more variable rate of cellulose degradation. In sum, this work demonstrates the importance of community interactions in shaping both the evolution and the ecology of cellulose degradation.

## **Acknowledgements**

Reflecting back on my path towards completing my dissertation, it is impossible to grasp how many people have contributed to my success in getting my PhD through their encouragements, support, and trust in my ability to succeed. I know that I am hugely shaped by the wonderful people I have been lucky enough to interact with.

Over the last six years, I feel extremely fortunate to have the support of so many wonderful people in the Currie lab. I am grateful for having a mentor who has supported me, guided me, and given me the freedom to explore. Cameron, thank you for allowing me to do research that I care about deeply and that I find so engaging and for always challenging me to improve and do my best. I have learned so much from you in the past six years that I will carry with me through the rest of my career. Also, I have had the honor to work with labmates who I am excited to see every morning, who have taught me so much, and who always are willing to help – in and out of the laboratory. Especially to Adam, thank you for your constant guidance, encouragement, and collaboration. Additionally, I have gotten to be a mentor to a large group of undergraduates, and their hard work has been critical for the success and productivity of much of my research.

Beyond the Currie-lab, I have very much enjoyed working with a set of wonderful collaborators in the Fox-lab, who have contributed substantially to many of my projects and have taken our research in exciting directions. I also have been lucky to get to work with Adrian Pinto-Tomas, his lab members at the University of Costa Rica, and the staff at La Selva Biological Research Station. Much of my work could not have happened without their support. Finally, I have had a



committee filled with extremely helpful and intelligent scientists. Paul Weimer, Brian Fox, Trina McMahon, and Heidi Goodrich-Blair, thank you for all your guidance and encouragement.

However, I likely would not have ended up in a microbiology graduate program in Wisconsin if not for a large group of wonderful teachers and mentors. Especially Professors Jonathan Wright and EJ Crane, thank you for letting me into your labs, for advising me constantly, and encouraging my interest in research. Also, I am indebted to Brian Pflieger, Rebecca Lennen, and John Greenler for their support during my summer REU in Madison as an undergraduate. I learned so much that summer and that experience was instrumental in guiding me towards pursuing a PhD.

Outside of the lab, I also have an amazing group of friends and family. Thank you to my parents for your constant support. It means so much to know that you are both always there for me. Thank you for encouraging me to explore, to be my best, and to live life to the fullest. Gabe, I am always so grateful for our time together. Also, none of this would be as meaningful without friends to share it with – thanks for all the walks, good times, and for always having my back. And lastly, thank you Eric for believing in me, loving me, and keeping me laughing.

## Table of Contents

|                                                                                                                                                                         |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Dissertation Abstract .....                                                                                                                                             | i         |
| Acknowledgements .....                                                                                                                                                  | iii       |
| List of Figures.....                                                                                                                                                    | xi        |
| List of Tables and Datasets .....                                                                                                                                       | xiv       |
| <b>Chapter 1: Evolution and Ecology of Actinobacteria and their Bioenergy Applications .....</b>                                                                        | <b>1</b>  |
| <b>1.1 Abstract.....</b>                                                                                                                                                | <b>1</b>  |
| <b>1.2 Introduction.....</b>                                                                                                                                            | <b>1</b>  |
| <b>1.3 Evolution and Genomics of Actinobacteria.....</b>                                                                                                                | <b>3</b>  |
| <i>Sidebar 1.1. Coevolution of Actinobacteria in the fungus-growing ant system.....</i>                                                                                 | <i>7</i>  |
| <b>1.4 Ecology of Actinobacteria .....</b>                                                                                                                              | <b>7</b>  |
| <b>1.4.1 Role of Free-Living Actinobacteria in Carbon Cycling .....</b>                                                                                                 | <b>8</b>  |
| <b>1.4.2 Role of Actinobacteria in Host Associations.....</b>                                                                                                           | <b>10</b> |
| <b>1.5 Actinobacteria in Bioenergy.....</b>                                                                                                                             | <b>12</b> |
| <b>1.5.1 Actinobacteria as a source of enzymes in biofuels .....</b>                                                                                                    | <b>13</b> |
| <i>Sidebar 1.2. Integrating actinobacterial enzymes in industrial cellulosic biofuel production.....</i>                                                                | <i>14</i> |
| <b>1.5.2 Actinobacteria, advanced biofuels, and biocommodities.....</b>                                                                                                 | <b>19</b> |
| <b>1.6 Conclusions.....</b>                                                                                                                                             | <b>21</b> |
| <b>1.7 References .....</b>                                                                                                                                             | <b>22</b> |
| <b>Chapter 2: Cellulolytic <i>Streptomyces</i> strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose .....</b> | <b>40</b> |
| <b>2.1 Abstract.....</b>                                                                                                                                                | <b>40</b> |
| <b>2.2 Introduction.....</b>                                                                                                                                            | <b>41</b> |
| <b>2.3 Materials and Methods.....</b>                                                                                                                                   | <b>44</b> |
| <b>2.3.1 Isolation of <i>Streptomyces</i>.....</b>                                                                                                                      | <b>44</b> |
| <b>2.3.2 Culturing of organisms.....</b>                                                                                                                                | <b>45</b> |
| <b>2.3.3 Quantification of residual cellulose. ....</b>                                                                                                                 | <b>45</b> |
| <b>2.3.4 Genomic sequencing and phylogenetic analysis.....</b>                                                                                                          | <b>46</b> |
| <b>2.3.5 Preparation of secretomes. ....</b>                                                                                                                            | <b>47</b> |
| <b>2.3.6 Enzyme activity measurements.....</b>                                                                                                                          | <b>47</b> |
| <b>2.3.7 Extracellular proteomics.....</b>                                                                                                                              | <b>48</b> |
| <b>2.3.8 RNA extraction and preparation .....</b>                                                                                                                       | <b>49</b> |
| <b>2.3.9 RNA sequencing and analysis.....</b>                                                                                                                           | <b>50</b> |

|                                                                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.3.10 Nucleotide sequence accession numbers. ....                                                                                                                      | 51 |
| 2.4 Results .....                                                                                                                                                       | 51 |
| 2.4.1 Cellulolytic activity screen. ....                                                                                                                                | 51 |
| 2.4.2 Biochemical analysis of secreted enzymes. ....                                                                                                                    | 51 |
| 2.4.3 Phylogenetic and genomic analyses. ....                                                                                                                           | 52 |
| 2.4.4 Secretome analysis on pure polysaccharides. ....                                                                                                                  | 55 |
| 2.4.5 RNA sequencing.....                                                                                                                                               | 58 |
| 2.4.6 Differential expression analysis. ....                                                                                                                            | 60 |
| 2.5 Discussion.....                                                                                                                                                     | 62 |
| 2.6 References .....                                                                                                                                                    | 65 |
| Chapter 3: Evolution of high cellulolytic activity in symbiotic <i>Streptomyces</i> through<br>selection of expanded gene content and coordinated gene expression ..... | 72 |
| 3.1 Abstract.....                                                                                                                                                       | 72 |
| 3.2 Introduction.....                                                                                                                                                   | 73 |
| 3.3 Results and Discussion.....                                                                                                                                         | 75 |
| 3.3.1 Genus-wide cellulolytic ability.....                                                                                                                              | 75 |
| 3.3.2 Polysaccharide-degrading activity of <i>Streptomyces</i> secretomes.....                                                                                          | 77 |
| 3.3.3 Comparative genomic analysis of plant biomass degrading genes. ....                                                                                               | 78 |
| 3.3.4 Comparative gene expression analysis.....                                                                                                                         | 81 |
| 3.3.5 Coordinated expression of CAZy genes controlled by the <i>CebR</i> transcription<br>regulator. ....                                                               | 83 |
| 3.3.6 Analysis of Important Enzyme Families.....                                                                                                                        | 85 |
| 3.3.7 Evolution of CAZy genes in cellulolytic strains.....                                                                                                              | 88 |
| 3.4 Materials and Methods.....                                                                                                                                          | 90 |
| 3.4.1 Isolation of <i>Streptomyces</i> .....                                                                                                                            | 90 |
| 3.4.2 Culturing of organisms. ....                                                                                                                                      | 91 |
| 3.4.3 Quantification of residual cellulose. ....                                                                                                                        | 91 |
| 3.4.4 Analysis of secreted enzymes. ....                                                                                                                                | 91 |
| 3.4.5 Genomic sequencing and phylogenetic analysis.....                                                                                                                 | 92 |
| 3.4.6 RNA analysis. ....                                                                                                                                                | 94 |
| 3.4.7 <i>CebR</i> analysis. ....                                                                                                                                        | 94 |
| 3.4.8 Lateral Gene Transfer Analysis. ....                                                                                                                              | 95 |
| 3.4.9 <i>Streptomyces</i> sp. <i>SirexAA-E</i> Genetics. ....                                                                                                           | 96 |
| 3.4.10 Accession numbers.....                                                                                                                                           | 98 |
| 3.5 References .....                                                                                                                                                    | 98 |

|                                                                                                                                                                                  |            |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Chapter 4: Cellulose-Enriched Microbial Communities from Leaf-Cutter Ant (<i>Atta colombica</i>) Refuse Dumps Vary in Taxonomic Composition and Degradation Ability .....</b> | <b>109</b> |
| 4.1 Abstract.....                                                                                                                                                                | 109        |
| 4.2 Introduction.....                                                                                                                                                            | 110        |
| 4.3 Materials and Methods.....                                                                                                                                                   | 114        |
| 4.3.1 Sample Collection .....                                                                                                                                                    | 114        |
| 4.3.2 Electron Microscopy .....                                                                                                                                                  | 115        |
| 4.3.3 Overview of Enrichment Design.....                                                                                                                                         | 115        |
| 4.3.4 DNA Extraction.....                                                                                                                                                        | 117        |
| 4.3.5 16S rRNA Gene Amplicon Sequencing .....                                                                                                                                    | 118        |
| 4.3.6 Amplicon Sequence Processing.....                                                                                                                                          | 118        |
| 4.3.7 Statistical Analyses.....                                                                                                                                                  | 119        |
| 4.3.8 Isolation of dominant community members .....                                                                                                                              | 120        |
| 4.3.9 <i>Acidovorax</i> Taxonomy and Growth Assay and Identification .....                                                                                                       | 122        |
| 4.4 Results .....                                                                                                                                                                | 123        |
| 4.4.1 Electron Microscopy of Leaf-Cutter Ant Refuse Dumps .....                                                                                                                  | 123        |
| 4.4.2 Differences in Cellulolytic Ability across Refuse Dumps.....                                                                                                               | 123        |
| 4.4.3 Amplicon Sequencing Analysis .....                                                                                                                                         | 126        |
| 4.4.4 Alpha Diversity of Enriched Communities .....                                                                                                                              | 130        |
| 4.4.5 Beta Diversity of Enriched Communities.....                                                                                                                                | 131        |
| 4.4.6 Correlation Analysis .....                                                                                                                                                 | 132        |
| 4.4.7 <i>Acidovorax</i> .....                                                                                                                                                    | 133        |
| 4.5 Discussion.....                                                                                                                                                              | 134        |
| 4.6 References .....                                                                                                                                                             | 140        |
| <b>Chapter 5: Long-term cellulose enrichment selects for highly cellulolytic taxa and competition for public goods.....</b>                                                      | <b>150</b> |
| 5.1 Abstract.....                                                                                                                                                                | 150        |
| 5.2 Introduction.....                                                                                                                                                            | 151        |
| 5.3 Results .....                                                                                                                                                                | 154        |
| 5.3.1 Cellulose degradation by enrichment lines.....                                                                                                                             | 154        |
| 5.3.2 Taxonomic shifts in community structure across 60 transfers .....                                                                                                          | 155        |
| 5.3.3 Taxonomic shifts in community structure during one transfer .....                                                                                                          | 159        |
| 5.3.4 Functional analysis of cellulose degradation .....                                                                                                                         | 161        |
| 5.3.5 Indications of competition within communities.....                                                                                                                         | 163        |
| 5.4 Discussion.....                                                                                                                                                              | 167        |

|                                                                                                               |     |
|---------------------------------------------------------------------------------------------------------------|-----|
| 5.4.1 Cellulose degradation in enriched communities.....                                                      | 167 |
| 5.4.2 Changes in taxonomic diversity over time .....                                                          | 169 |
| 5.4.3 Interactions within enriched communities.....                                                           | 172 |
| 5.5 Conclusion .....                                                                                          | 173 |
| 5.6 Materials and Methods.....                                                                                | 174 |
| 5.6.1 Refuse dump collection.....                                                                             | 174 |
| 5.6.2 Enrichment procedure .....                                                                              | 174 |
| 5.6.3 Quantification of cellulose degradation.....                                                            | 175 |
| 5.6.4 Seven-day analysis of enrichment line 2B at transfer 70.....                                            | 176 |
| 5.6.5 Collection of cultures for sequencing analyses .....                                                    | 176 |
| 5.6.6 DNA extraction and sequencing.....                                                                      | 177 |
| 5.6.7 RNA extraction and sequencing.....                                                                      | 178 |
| 5.6.8 16S rRNA gene sequencing analysis.....                                                                  | 178 |
| 5.6.9 Analysis of OTU succession over time .....                                                              | 179 |
| 5.6.10 Co-occurrence analysis among enrichment lines across transfers.....                                    | 179 |
| 5.6.11 Phylogenetic analysis of <i>Cellvibrio</i> populations.....                                            | 180 |
| 5.6.12 Metagenome assembly .....                                                                              | 180 |
| 5.6.13 Composite genome assembly and taxonomic analysis.....                                                  | 181 |
| 5.6.14 Classification of 16S rRNA genes in metagenomic data.....                                              | 181 |
| 5.6.15 Metatranscriptome mapping.....                                                                         | 182 |
| 5.6.16 Annotation of metagenomes and composite genomes .....                                                  | 182 |
| 5.7 References.....                                                                                           | 183 |
| Chapter 6: Summary and Future Outlook.....                                                                    | 195 |
| Appendix 1: Supplementary Material for Chapter 1 .....                                                        | 198 |
| Appendix 2: Supplementary Material for Chapter 2 .....                                                        | 204 |
| Appendix 3: Supplementary Material for Chapter 3 .....                                                        | 210 |
| Appendix 4: Supplementary Material for Chapter 4 .....                                                        | 238 |
| Appendix 5: Supplementary Material for Chapter 5. ....                                                        | 248 |
| Appendix 6: Aerobic deconstruction of cellulosic biomass by an insect-associated<br><i>Streptomyces</i> ..... | 254 |
| A6.1 Abstract.....                                                                                            | 254 |
| A6.2 Introduction.....                                                                                        | 255 |
| A6.3 Results .....                                                                                            | 257 |
| A6.3.1 Genomic characteristics of ActE .....                                                                  | 257 |
| A6.3.2 Secreted proteins from ActE during growth on pure polysaccharides.....                                 | 259 |

|                                                                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <i>A6.3.3 Reactions of the ActE secretomes</i> .....                                                                                                                  | 259 |
| <i>A6.3.4 Secretome analysis on pure polysaccharides and plant biomass</i> .....                                                                                      | 263 |
| <i>A6.3.5 Gene expression analysis during growth on purified polysaccharides and plant biomass</i> .....                                                              | 267 |
| <b>A6.4 Discussion</b> .....                                                                                                                                          | 271 |
| <b>A6.5 Methods</b> .....                                                                                                                                             | 276 |
| <i>A6.5.1 Genome analysis</i> .....                                                                                                                                   | 276 |
| <i>A6.5.2 Biomass substrates</i> .....                                                                                                                                | 276 |
| <i>A6.5.3 Growth of organisms</i> .....                                                                                                                               | 277 |
| <i>A6.5.4 Preparation of secretomes</i> .....                                                                                                                         | 277 |
| <i>A6.5.5 Enzyme activity measurements</i> .....                                                                                                                      | 277 |
| <i>A6.5.6 Extracellular protein profiles</i> .....                                                                                                                    | 278 |
| <i>A6.5.7 RNA microarrays</i> .....                                                                                                                                   | 279 |
| <b>A6.6 References</b> .....                                                                                                                                          | 279 |
| <b>A6.7 Supplementary Material</b> .....                                                                                                                              | 285 |
| <b>Appendix 7: Microbial community analysis of leaf-cutter ant refuse dumps reveals patterns of community succession and strata-specific coexisting modules</b> ..... | 297 |
| <b>A7.1 Abstract</b> .....                                                                                                                                            | 297 |
| <b>A7.2 Introduction</b> .....                                                                                                                                        | 298 |
| <b>A7.3 Results</b> .....                                                                                                                                             | 302 |
| <i>A7.3.1 Taxonomic description of leaf-cutter ant colonies</i> .....                                                                                                 | 304 |
| <i>A7.3.2 Alpha and beta diversity estimates</i> .....                                                                                                                | 304 |
| <i>A7.3.3 Co-occurrence analysis</i> .....                                                                                                                            | 306 |
| <i>A7.3.4 Network topology</i> .....                                                                                                                                  | 308 |
| <i>A7.3.5 Markov clustering and coexisting modules</i> .....                                                                                                          | 309 |
| <i>A7.3.6 Putative core community members of refuse dumps</i> .....                                                                                                   | 311 |
| <b>A7.4 Discussion</b> .....                                                                                                                                          | 312 |
| <b>A7.5 Materials and Methods</b> .....                                                                                                                               | 316 |
| <i>A7.5.1 Site description, sampling, and study overview</i> .....                                                                                                    | 316 |
| <i>A7.5.2 Estimates of community diversity</i> .....                                                                                                                  | 317 |
| <i>A7.5.3 Co-occurrence analysis</i> .....                                                                                                                            | 318 |
| <i>A7.5.4 Network analysis of coexisting OTUs</i> .....                                                                                                               | 319 |
| <i>A7.5.5 Identification of coexisting OTU clusters</i> .....                                                                                                         | 319 |
| <b>A7.6 References</b> .....                                                                                                                                          | 320 |
| <b>A7.7 Supporting Material</b> .....                                                                                                                                 | 329 |

|                                                                                     |            |
|-------------------------------------------------------------------------------------|------------|
| <b>A7.7.1 Supplementary Methods.....</b>                                            | <b>330</b> |
| A7.7.1.a DNA extraction, clone library sequencing and bar-coded pyrosequencing .... | 330        |
| A7.7.1.b Sequence processing and taxonomic classification.....                      | 330        |
| A7.7.1.c Detecting differentially abundant OTUs .....                               | 332        |
| A7.7.1.d Comparison of full-length and pyrosequencing datasets .....                | 333        |
| A7.7.1.e Changes in community composition within refuse dumps.....                  | 333        |
| A7.7.1.f Changes in community .....                                                 | 334        |
| <b>A7.7.2 Supplementary Results.....</b>                                            | <b>334</b> |
| A7.7.2.a Network topology.....                                                      | 334        |
| <b>A7.7.3 References.....</b>                                                       | <b>335</b> |

## List of Figures

|                                                                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 1.1. Actinobacterial phylogenomics and relative abundance across ecosystems. ....                                                                                                     | 4   |
| Figure 1.2. Actinobacteria in nature. ....                                                                                                                                                   | 6   |
| Figure 1.3. Enrichment of (a) CAZy domains and (b) biosynthetic gene clusters in select actinobacterial genera .....                                                                         | 16  |
| Figure 1.4. Molecular details of the proposed role of Actinobacteria in biofuels. ....                                                                                                       | 18  |
| Figure 2.1. ....                                                                                                                                                                             | 44  |
| Figure 2.2. Reactions of secretomes from 4 <i>Streptomyces</i> strains with polysaccharide substrates. ....                                                                                  | 54  |
| Figure 2.3. Multilocus phylogenetic tree of the genus <i>Streptomyces</i> . ....                                                                                                             | 55  |
| Figure 2.4. Composition of secretomes from cellulose-grown cultures of SACTE and SDPB6 identified by LC-MS/MS. ....                                                                          | 57  |
| Figure 2.5. Differential expression of genes from SDPB6 (A) and SACTE (B) grown on glucose versus AFEX-pretreated corn stover. ....                                                          | 59  |
| Figure 2.6. Protein similarity network of CAZymes present in SDPB6 and SACTE. ....                                                                                                           | 61  |
| Figure 3.1. Distribution of cellulolytic ability in the genus <i>Streptomyces</i> . ....                                                                                                     | 76  |
| Figure 3.2. Comparative analysis of CAZy genes across the <i>Streptomyces</i> genus. ....                                                                                                    | 79  |
| Figure 3.3. Differential expression and coregulation of biomass degrading genes. ....                                                                                                        | 82  |
| Figure 3.4. Genetic analysis of <i>Streptomyces</i> sp. SirexAA-E. ....                                                                                                                      | 84  |
| Figure 3.5. Multi-domain enzyme expression and evolution. ....                                                                                                                               | 89  |
| Figure 4.1. Leaf-cutter ants concentrate decaying plant matter in refuse dumps. ....                                                                                                         | 111 |
| Figure 4.2. Electron microscopy of leaf-cutter ant refuse dumps. ....                                                                                                                        | 124 |
| Figure 4.3. Comparison of degradation ability across colonies and layers. ....                                                                                                               | 125 |
| Figure 4.4. Morisita-Horn Beta Diversity Clustering of Samples. ....                                                                                                                         | 128 |
| Figure 4.5. Correlation between number of reads and cellulose degradation by the community for OTU1 ( <i>Acidovorax</i> ) and OTU9 ( <i>Ferruginibacter</i> ) across sequenced samples. .... | 130 |
| Figure 4.6. PCoA clustering of Morisita-Horn Diversity Index. ....                                                                                                                           | 133 |
| Figure 5.1. Cellulose enrichment methods and results. ....                                                                                                                                   | 154 |
| Figure 5.2. Diversity changes in enrichment lines over time .....                                                                                                                            | 157 |
| Figure 5.3. OTU patterns and correlations among enrichment lines and across transfers. ....                                                                                                  | 158 |
| Figure 5.4. Decomposition and taxonomic shifts for enrichment line 2B across seven days. ....                                                                                                | 160 |
| Figure 5.5. Relative levels of cellulase (Cel) and $\beta$ -glucosidase (Bg) genes and transcripts. ....                                                                                     | 162 |
| Figure 5.6. Functional annotation and expression level of the 25 most highly expressed genes in select composite genomes. ....                                                               | 166 |
| Figure A2.1. 18S phylogenetic tree of select holometabolous insect phyla. ....                                                                                                               | 205 |
| Figure A2.2. Pairwise comparison of raw number of reads for each biological replicate in each condition. ....                                                                                | 206 |
| Figure A2.3. Sample to sample variance of RNA-Seq data. ....                                                                                                                                 | 207 |
| Figure A3.1. Secreted protein profiles of <i>Streptomyces</i> strains from clades I-III. ....                                                                                                | 211 |
| Figure A3.2. Polysaccharide degrading enzyme activities (DNS activities) of <i>Streptomyces</i> secretomes .....                                                                             | 212 |
| Figure A3.3. Comparative analysis of CAZy genes across the <i>Streptomyces</i> genus. ....                                                                                                   | 213 |
| Figure A3.4. Comparative analysis of <i>Streptomyces</i> phylogeny and the CAZy gene profile. ...                                                                                            | 214 |
| Figure A3.5. Comparative analysis of ABC sugar transporter genes across the <i>Streptomyces</i> genus. ....                                                                                  | 215 |



|                                                                                                                                                                            |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure A3.6. Promoter motif analysis of cellulolytic and non-cellulolytic <i>Streptomyces</i> .                                                                            | 216 |
| Figure A3.7. Genetic engineering of <i>Streptomyces</i> sp. SirexAA-E.                                                                                                     | 217 |
| Figure A3.8. Secreted protein profiles of rescued <i>Streptomyces</i> sp. SirexAA-E deletion strains.                                                                      | 218 |
| Figure A3.9. Evolution of the CebR transcription regulator in cellulolytic and low activity strains.                                                                       | 219 |
| Figure A3.10. Phylogenetic diversity of AA10 genes from strains in clades I-III.                                                                                           | 220 |
| Figure A3.11. Phylogenetic diversity of Cip1 in bacteria and fungi.                                                                                                        | 221 |
| Figure A3.12. Evolution of CAZy gene families in <i>Streptomyces</i> .                                                                                                     | 222 |
| Figure A3.13. Evolution of CBM2, CBM3 and GH6 gene families.                                                                                                               | 223 |
| Figure A4.1.                                                                                                                                                               | 239 |
| Figure A4.2. Representative qualitative samples.                                                                                                                           | 240 |
| Figure A4.3. Relationship between the percentage of cellulose degraded and diversity (Inverse Simpson's Index) across sequenced samples.                                   | 241 |
| Figure A4.4. Pairwise comparisons of OTU abundance between samples with similar diversity.                                                                                 | 242 |
| Figure A4.5. PCoA clustering of the weighted Unifrac metric of similarity among samples.                                                                                   | 243 |
| Figure A4.6. Phylogeny and growth of an <i>Acidovorax</i> isolate.                                                                                                         | 244 |
| Figure A5.1. Replicates of enrichment lines.                                                                                                                               | 249 |
| Figure A5.2. Percentage of cellulose degradation in 10 days.                                                                                                               | 250 |
| Figure A5.3. Taxonomic and functional analysis of <i>Cellvibrio</i> isolates.                                                                                              | 251 |
| Figure A5.4. Cellulase genes and transcript annotation and abundance.                                                                                                      | 252 |
| Figure A6.1. Growth of ActE in minimal medium containing filter paper as the sole carbon source.                                                                           | 258 |
| Figure A6.2: Reactions of ActE secretomes and Spezyme CP.                                                                                                                  | 261 |
| Figure A6.3. Composition of ActE secretomes identified by LC-MS/MS.                                                                                                        | 262 |
| Figure A6.4: Genome-wide changes in expression during growth of ActE on AFEX-treated switchgrass (AFEX-SG) versus glucose.                                                 | 264 |
| Figure A6.5: Expression of ActE CAZy genes on various carbon sources.                                                                                                      | 265 |
| Supplementary Figure A6.1.                                                                                                                                                 | 285 |
| Supplementary Figure A6.2. Fractionation of the ActE cellulose secretome and assays of reactions with different polysaccharides.                                           | 286 |
| Supplementary Figure A6.3. Temperature and pH profiles of the ActE secretome obtained from growth on AFEX-treated corn stover.                                             | 287 |
| Supplementary Figure A6.4. Genome-wide changes in expression during growth of ActE on substrate cellobiose versus glucose visualized as a Cytoscape interaction network.   | 288 |
| Supplementary Figure A6.5. Genome-wide expression changes for growth on the substrate cellulose versus glucose visualized as a Cytoscape interaction network.              | 289 |
| Supplementary Figure A6.6. Genome-wide expression changes for growth on the substrate xylan versus glucose visualized as a Cytoscape interaction network.                  | 290 |
| Supplementary Figure A6.7. Genome-wide expression changes for growth on the substrate switchgrass versus glucose visualized as a Cytoscape interaction network.            | 291 |
| Supplementary Figure A6.8. Genome-wide expression changes for growth on the substrate IL-treated switchgrass versus glucose visualized as a Cytoscape interaction network. | 292 |
| Supplementary Figure A6.9. Genome-wide expression changes for growth on the substrate chitin versus glucose visualized as a Cytoscape interaction network.                 | 293 |

|                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Supplementary Figure A6.10. Expression of 167 predicted CAZy genes in ActE, highlighting group 2 genes. ....                                                              | 294 |
| Supplementary Figure A6.11. Expression of 167 predicted CAZy genes in ActE, highlighting group 3 genes. ....                                                              | 295 |
| Figure A7.1. The leaf-cutter ant refuse system.....                                                                                                                       | 299 |
| Figure A7.2. Clustering analysis and taxonomic distribution of garden and dump samples.....                                                                               | 303 |
| Figure A7.3. Alpha diversity estimates of garden and dump communities.....                                                                                                | 305 |
| Figure A7.4. Co-occurrence networks. ....                                                                                                                                 | 308 |
| Supplementary Figure A7.1. Individual clustering analyses of garden and dump samples. ....                                                                                | 339 |
| Supplementary Figure A7.2. Frequency of co-occurrence association values .....                                                                                            | 340 |
| Supplementary Figure A7.3. Assessment of the number of clones per OTU and node connectivity (degree) for each OTU from the a) Ba/Eu and b) Py co-occurrence networks..... | 341 |
| Supplementary Figure A7.4. Results of MCL analyses of co-occurrence networks. ....                                                                                        | 342 |
| Supplementary Figure A7.5. Taxonomic composition of fungus gardens, refuse dumps, and tropical soils .....                                                                | 343 |
| Supplementary Figure A7.6. Assessment of taxonomic changes through dumps.....                                                                                             | 344 |
| Supplementary Figure A7.7. Degree distribution from co-occurrence networks.....                                                                                           | 345 |

## List of Tables and Datasets

|                                                                                                                                                          |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Table 2.1.</b> Genome Statistics.....                                                                                                                 | 53  |
| <b>Table 2.2.</b> Genomic composition of putative biomass-degrading enzymes and accessory proteins .....                                                 | 53  |
| <b>Table 4.1.</b> Degradation, sequencing depth, and alpha diversity metrics for sequenced samples at a 97% OTU definition.....                          | 127 |
| <b>Table 4.2.</b> Taxonomy of top 30 OTUs across all samples and reports of cellulose degradation. ....                                                  | 129 |
| <b>Table A1.1.</b> MG-RAST datasets and genus distributions used to calculate average abundances in Figure 1.1. ....                                     | 199 |
| <b>Table A1.2.</b> Genomes used for phylogenetic tree construction and molecular clock analysis for Figure 1.1. ....                                     | 200 |
| <b>Table A1.3.</b> CAZy and biosynthetic gene cluster abundance used to calculate data represented in Figure 1.3. ....                                   | 201 |
| <b>Table A1.4.</b> Titters of biofuels and bioproducts produced by Actinobacteria compared with other microbes. <sup>a</sup> .....                       | 201 |
| <b>Table A2.1.</b> RNA-Seq statistics .....                                                                                                              | 208 |
| <b>Dataset A2.1.</b> TIGERFAM protein families used to generate the multilocus phylogenetic tree in Figure 2.3. ....                                     | 208 |
| <b>Dataset A2.2.</b> CAZy annotations for SACTE, SDPB6, SACTG, and SPB74. ....                                                                           | 208 |
| <b>Dataset A2.3.</b> LC-MS/MS of SACTE, S2a protein report, and S2b peptide report. ....                                                                 | 208 |
| <b>Dataset A2.4.</b> LC-MS/MS of SDPB6, S2a protein report, and S2b peptide report. ....                                                                 | 208 |
| <b>Dataset A2.5.</b> Differential expression analysis of SACTE (ActE) RNA-seq data for cultures grown on glucose or AFEX-CS. ....                        | 208 |
| <b>Dataset A2.6.</b> Differential expression analysis of SDPB6 RNA-seq data for cultures grown on glucose or AFEX-CS. ....                               | 209 |
| <b>Dataset A2.7.</b> Differential expression of KEGG and CAZy categories for SDPB6 and SACTE based on RNA-seq differential expression (DE) analysis..... | 209 |
| <b>Table A3.1</b> Qualitative filter-paper degrading activity of <i>Streptomyces</i> .....                                                               | 224 |
| <b>Table A3.2.</b> Genome Accession Numbers .....                                                                                                        | 230 |
| <b>Table A3.3.</b> Summary of genes upregulated $\geq 2$ -fold .....                                                                                     | 233 |
| <b>Table A3.4.</b> Expression and annotation of genes predicted to be directly controlled by the CebR repressor .....                                    | 234 |
| <b>Dataset A3.1.</b> Comparative genomic analysis of all ABC sugar transporters in <i>Streptomyces</i> . ....                                            | 237 |
| <b>Dataset A3.2.</b> RNAseq differential expression data table for SirexAA-E, DpondAA-B6, ATexAB-D23, LaPpAH-95, <i>S. griseus</i> and ScaeMP-e83. ....  | 237 |
| <b>Table A4.1.</b> Qualitative and quantitative data for all enrichment communities. ....                                                                | 245 |
| <b>Table A4.2.</b> Taxonomic identification of isolates. ....                                                                                            | 247 |
| <b>Dataset A4.1.</b> Taxonomic classification of reads and OTUs across samples. ....                                                                     | 247 |
| <b>Dataset A4.2.</b> Correlation analysis of degradation and OTUs with community diversity. ....                                                         | 247 |
| <b>Table A5.1.</b> Composite genome assembly quality and metatranscriptome mapping statistics..                                                          | 253 |
| <b>Table A6.1.</b> Comparison of genomic composition .....                                                                                               | 257 |
| <b>Table A6.2.</b> Analysis of upstream DNA sequence elements in ActE genes upregulated during growth on cellulose. ....                                 | 269 |

|                                                                                                                                                                                                                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>Supplementary Table A6.1.</b> Proteins separated by ion exchange chromatography and identified by mass spectrometry.....                                                                                                                                                                                         | 296 |
| <b>Supplementary Table A6.2.</b> Spectra count of proteins identified on each substrate, where top 95 % spectra covered were highlighted green, light purple, purple, blue, orange, pink, light blue and yellow on glucose, cellobiose, cellulose, xylan switchgrass, AFEX-SG, IL-SG and chitin, respectively. .... | 296 |
| <b>Table A7.1.</b> Results of co-occurrence analysis .....                                                                                                                                                                                                                                                          | 307 |
| <b>Table A7.2.</b> Putative core microbes in leaf-cutter ant refuse dumps .....                                                                                                                                                                                                                                     | 310 |
| <b>Supplementary Table A7.1.</b> Summary of taxonomic distribution (sequences).....                                                                                                                                                                                                                                 | 346 |
| <b>Supplementary Table A7.2.</b> Summary of taxonomic distribution (OTUs). ....                                                                                                                                                                                                                                     | 346 |
| <b>Supplementary Table A7.3.</b> Full results of the Metastats, co-occurrence, and clustering analyses. ....                                                                                                                                                                                                        | 346 |
| <b>Supplementary Table A7.4.</b> Summary of taxon distribution in the co-occurrence analysis ( <i>Ba/Eu</i> ).....                                                                                                                                                                                                  | 347 |
| <b>Supplementary Table A7.5.</b> Summary of taxon distribution in the co-occurrence analysis ( <i>Py</i> ) .....                                                                                                                                                                                                    | 348 |
| <b>Supplementary Table A7.6.</b> Topological parameters of the <i>Ba/Eu</i> and <i>Py</i> co-occurrence networks.....                                                                                                                                                                                               | 349 |
| <b>Supplementary Table A7.7.</b> Total, unique, and shared OTUs from refuse dumps at various definitions for the full-length ( <i>Ba</i> ) and pyrotag ( <i>Py</i> ) datasets. ....                                                                                                                                 | 350 |
| <b>Supplementary Table A7.8.</b> Taxonomic identity of bacterial OTUs shared across all <i>A. colombica</i> three nests from the <i>Ba</i> and <i>Py</i> datasets at 99% OTU cutoff. ....                                                                                                                           | 351 |
| <b>Supplementary Table A7.9.</b> Number (and percentage) of unique OTUs from Supplementary Table A7.8 present as singletons or doubletons for the full-length ( <i>Ba</i> ) and pyrotag ( <i>Py</i> ) 16S datasets at various OTU definitions.....                                                                  | 352 |
| <b>Supplementary Table A7.10.</b> Total OTUs shared between all layers (D0–D4) within each dump from the full-length ( <i>Ba</i> ) and pyrotag ( <i>Py</i> ) datasets at various OTU definitions .....                                                                                                              | 353 |

## **Chapter 1: Evolution and Ecology of Actinobacteria and their Bioenergy Applications**

**Accepted:** Gina R. Lewin, Camila Carlos, Marc G. Chevrette, Heidi A. Horn, Bradon R. McDonald, Robert J. Stankey, Brian G. Fox, and Cameron R. Currie. Evolution and Ecology of Actinobacteria and their Bioenergy Applications. *Annual Reviews of Microbiology*.

**Co-authors contributed in the following ways:** Conceived and designed this work: GRL, BGF, CRC. Analyzed data and prepared Figure 1.1: CC, BRM, GRL; Figure 1.2: GRL; Figure 1.3: CC, MGC, GRL; Figure 1.4: GRL. Wrote the paper: GRL, CRC with contributions from CC, MGC, HAH, BRM, RJS, BGF.

### **1.1 Abstract**

The ancient phylum Actinobacteria is composed of phylogenetically and physiologically diverse bacteria that help Earth's ecosystems function. As free-living organisms and symbionts of herbivorous animals, Actinobacteria contribute to the global carbon cycle through the breakdown of plant biomass. In addition, they mediate community dynamics as producers of small molecules with diverse biological activities. Together, the evolution of high cellulolytic ability and diverse chemistry, shaped by their ecological roles in nature, make Actinobacteria a promising group for the bioenergy industry. Specifically, their enzymes can contribute to industrial-scale breakdown of cellulosic plant biomass into simple sugars that can then be converted into biofuels. Furthermore, harnessing their ability to biosynthesize a range of small molecules has potential for the production of specialty biofuels.

### **1.2 Introduction**

Actinobacteria are ubiquitous and one of the most diverse groups of bacteria in nature. Its

members range from anaerobic, unicellular organisms to aerobic, filamentous, and spore-forming lineages. Alone, the genus *Streptomyces* accounts for almost 5% of the ~16,000 described bacterial species (<http://www.bacterio.net/>). Actinobacteria have made a substantial positive contribution to human health; they are the producers of many compounds that are used as important drugs, including most antibiotics (58). In addition, the phylum includes important pathogens, such as *Mycobacterium tuberculosis*, the causal agent of one of the most devastating diseases in human history, tuberculosis.

Largely because of these biomedical impacts, Actinobacteria have avoided the obscurity relegated to most bacterial phyla, receiving substantial scientific study of their genetics, molecular biology, and biochemistry. In addition to their important influence on human health, Actinobacteria have key ecological roles. Before focusing on antibiotic discovery, Nobel laureate Selman Waksman's work on soil bacteria and their impact on agricultural productivity was among the first to implicate Actinobacteria as important contributors to the process of plant biomass decomposition (58, 146). More recently, Actinobacteria have been revealed as widespread symbionts of eukaryotes, helping herbivores gain access to plant biomass as nutritional mutualists and producing natural products as defensive mutualists (e.g., 18, 33, 38, 65). Studies of Actinobacteria as defensive mutualists have led to the discovery of new antibiotics with potential pharmaceutical applications (97, 106, 117), renewing recognition of the value in understanding the ecology of Actinobacteria for drug discovery (e.g., 12, 122) and hearkening back to Waksman's journey to the discovery of streptomycin.

We suggest, much like in drug discovery, the ecology and evolution of Actinobacteria position

this phylum to contribute to the development of a sustainable bioenergy industry. Specifically, as Actinobacteria are implicated as important decomposers of plant material in nature, their cellulolytic enzymes can be used to more efficiently break down plant biomass into simple sugars, which can then be used for the production of biofuels. In addition, the diverse biosynthetic capacity of Actinobacteria, which evolved to mediate their environmental interactions, could be leveraged to produce a range of bioproducts, including specialty biofuel compounds. Here, we review our current understanding of the ecology and evolution of Actinobacteria. Then, within this framework, we discuss the potential application of Actinobacteria in bioenergy.

### **1.3 Evolution and Genomics of Actinobacteria**

The evolutionary origin of the Actinobacteria is ancient, dating back to ~2,700 million years ago, predating oxygenation of the atmosphere (~2,300 million years ago) (Figure 1.1) (8, 150). 16S rRNA gene sequencing, multilocus phylogenetics, and phylogenomics support a shared common ancestor for Actinobacteria, Cyanobacteria, and Deinococcus (8, 52). Together, these three lineages share a most recent common ancestor with the phylum Firmicutes, with an estimated divergence more than 3,000 million years ago (8). Consistent with evolving in an anoxic atmosphere, the earliest Actinobacteria were obligate anaerobes, were not yet filamentous or able to form spores, and instead had simple rod/coccus morphology (26). The lineage Actinobacteria subsequently diverged into at least six classes and 53 families (52). The ability to grow as filaments and produce spores evolved in the more recently derived lineages but was secondarily lost in a few groups (26).

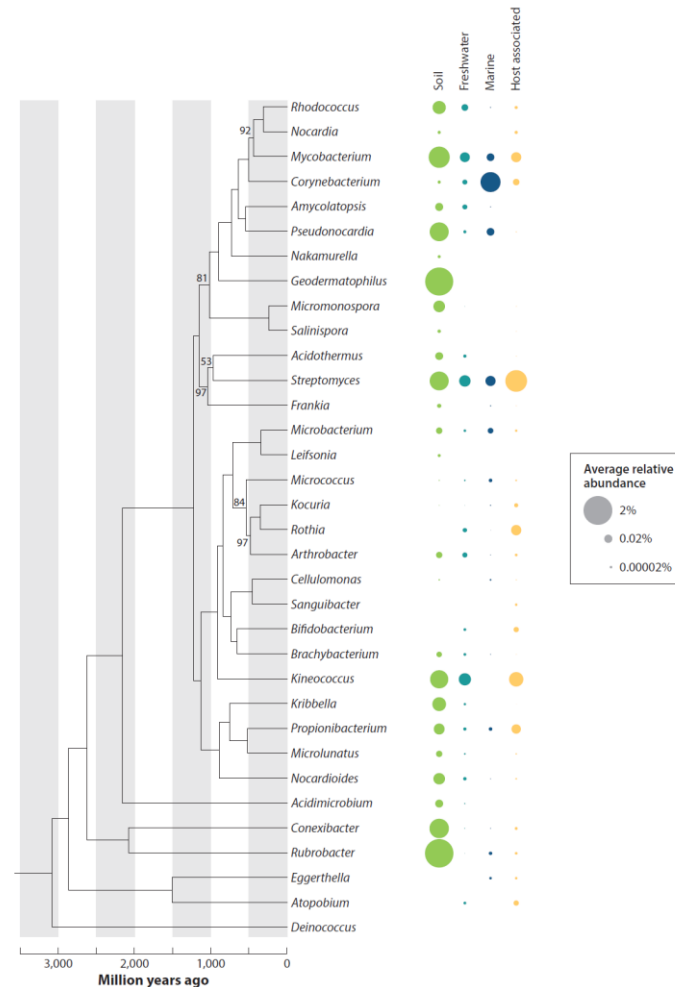


**Figure 1.1. Actinobacterial phylogenomics and relative abundance across ecosystems.**

Actinobacterial phylogenomics and relative abundance across ecosystems. Forty-six publicly available 16S rRNA gene datasets, spanning soil, freshwater, marine, and host-associated ecosystems, were analyzed to determine the average relative percentage of reads assigned to each actinobacterial genus (93) (Table A1.1). Phylogenetic analysis was performed using concatenated alignments from 94 conserved TIGRFAM protein families in representative complete actinobacterial genomes (Table A1.2), with the outgroup *Deinococcus*. Sequences were aligned with MAFFT 7.221, and the phylogeny was generated using RAxML 8.1.24 (67, 124). Molecular clock calculations were performed with RelTime using the origin of life (3,500–3,800 million years ago), the origin of *Cyanobacteria* (2,500–3,500 million years ago), and the *Escherichia Salmonella* divergence time (50–150 million years ago) as calibration points (133). Statistical support for the phylogeny, based on a sequence resampling bootstrap analysis, is shown on nodes with less than 100% support.



In general, the genomes of Actinobacteria are large, relative to other phyla of bacteria, with an average genome size of over 5 megabases (Mb). However, genome sizes vary widely, from less than 1 Mb in some *Tropheryma* species to over 12 Mb in a few *Streptomyces* species. In general, the genomes have a high guanine-plus-cytosine (G+C) content, with the G+C content of some genera, such as *Streptomyces* and *Pseudonocardia*, more than 70%. However, G+C content is less than 50% in *Gardnerella vaginalis* and *Tropheryma whipplei* (2, 6). Of those analyzed, all Actinobacteria have a single chromosome, but the presence of large plasmids is not uncommon. The ancestral state is a circular chromosome, but the genera *Gordonibacter*, *Kineococcus*, *Rhodococcus*, and *Streptomyces* all have linear chromosomes, which appear to have independently evolved in each genus.

Genome content in bacteria is shaped by the interplay between vertical inheritance (i.e., clonal reproduction) with gene loss and acquisition, the latter either involving the genesis of new genes within a lineage or horizontal gene transfer (HGT) from other lineages. Our current understanding of the evolutionary forces shaping gene loss and gene origin in Actinobacteria is limited. Recent genomic analyses of 120 *Streptomyces* genomes indicate that, at least within this genus, HGT is relatively rare and that long-term retention of horizontally transferred genes occurs several orders of magnitude less frequently than the origin of point mutations (B.R. McDonald & C.R. Currie, unpublished information). Nevertheless, HGT enables adaptive divergence of populations through the introduction of novel genetic diversity. For example, virulence operons in *M. tuberculosis* and plant pathogenic *Streptomyces* species appear to have been acquired through HGT (69, 109). Furthermore, HGT has been shown as a major driver of secondary metabolite diversity in the marine actinobacterial genus *Salinispora* (161).

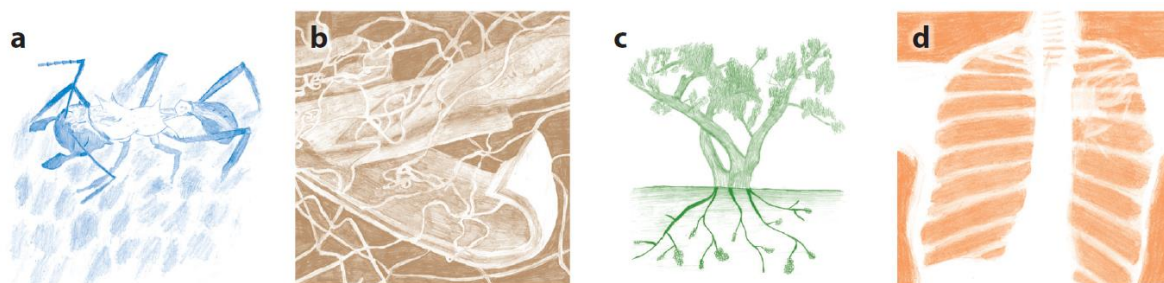


**Figure 1.2. Actinobacteria in nature.** (a) An *Acromyrmex* leaf-cutter ant in her fungus garden covered in white *Pseudonocardia* bacteria. (b) A *Streptomyces* strain growing on decaying plant biomass. (c) *Frankia* nodules on the roots of a Russian olive tree. (d) X-ray showing a healthy lung on the left and *Mycobacterium tuberculosis* growth on the right..

Coevolution is another evolutionary force involved in shaping the biology of Actinobacteria. Coevolution inherently recognizes the importance of species interactions in the evolution and diversity of lineages (134). Thus, given that Actinobacteria typically occur either within complex communities of interacting microbes or as symbionts of eukaryotic hosts (see “1.4 Ecology of Actinobacteria,” below), reciprocal selection is likely to influence the evolution of these lineages. One of the few studied examples of coevolution in Actinobacteria is the *Pseudonocardia*-ant defensive mutualism (see Sidebar 1.1 and Figure 1.2a). In this example, as well as other defensive mutualisms, Actinobacteria provide antimicrobial protection to a host organism, and species interactions are mediated by secondary metabolites that are likely influenced and shaped by coevolution. Further studies are needed, both on how chemical diversification is shaped by coevolutionary forces and on the nutritional interactions between Actinobacteria and their hosts.

### *Sidebar 1.1. Coevolution of Actinobacteria in the fungus-growing ant system*

Some lineages of *Pseudonocardia* are associated with fungus-growing ants (tribe Attini), with the bacterium residing on the ants' exoskeleton (see Figure 1.2a) (38). These *Pseudonocardia* produce antimicrobial compounds that help defend the ants' food source, a fungal cultivar, from specific pathogens (38, 97, 152). The bacterium is vertically inherited (38), and there is a critical window in which the ants must be exposed to the bacterium for acquisition to occur (87). Specialized glands and structures on the ant exoskeleton supply nutrients to the bacterium, indicating evolutionary selection on the ant for hosting *Pseudonocardia* (37, 126). Likewise, molecular phylogenetic studies of symbiotic *Pseudonocardia* reveal congruence with the host ant phylogeny (22, 23). Additional complexity is seen in the population dynamics of *Pseudonocardia*, which are correlated to that of a system-specific pathogen, *Escovopsis* (24, 36, 38). Indeed, studies examining pair-wise interactions between *Pseudonocardia* and *Escovopsis* strains show varying degrees of pathogen inhibition, indicating diversity in secondary metabolites produced by *Pseudonocardia* as well as variation in pathogen resistance (22, 23, 38). These findings support coevolution shaping the dynamics and biology of this system, which likely has generated and influenced the diversity of secondary metabolism of the bacterial symbiont. Similar dynamics are likely common across defensive mutualisms, such as in that of solitary digger wasps (tribe Philanthini), which engage in a defensive mutualism with *Streptomyces* strains to offer antimicrobial protection during larval development (66, 72). Potentially, similar dynamics also occur in nutritional mutualisms.

## **1.4 Ecology of Actinobacteria**

Historically, Actinobacteria were largely viewed as soil bacteria but are now recognized as being cosmopolitan; they are found in virtually all ecosystems, with a distribution covering most of the planet. As free-living microbes, they are abundant in soil environments, especially those with a higher pH (77), and in marine and freshwater ecosystems (Figures 1.1 and 1.2). Actinobacteria are also associated with eukaryotic hosts in diverse niches, such as the exoskeleton of some tropical ants, the lungs and skin of mammals, and the roots and inner tissues of plants (Figures 1.1 and 1.2). Some genera, including *Streptomyces*, *Kineococcus*, and *Mycobacterium*, span

diverse ecosystems from soil to marine to freshwater environments. Others, such as *Atopobium*, *Bifidobacterium*, *Kocuria*, and *Rothia*, are mostly found in host associations (Figure 1.1) (Table A1.1). The diversity of niche utilization in Actinobacteria mirrors their physiological capability to grow in a range of conditions, as supported by culture- and genome-based analyses.

Nevertheless, our understanding of the ecological role of Actinobacteria in many of these systems remains poorly resolved. Here, we explore their ecology, focusing on cases where the functional role of Actinobacteria has been elucidated with the understanding that Actinobacteria are critical for a wide range of ecosystem services and host interactions.

#### ***1.4.1 Role of Free-Living Actinobacteria in Carbon Cycling***

The largest terrestrial source of organic carbon is the network of energy-rich but recalcitrant polymers in the plant cell wall (156). The primary polymer in the plant cell wall is cellulose, an insoluble crystal of  $\beta$ -1,4-linked glucose molecules, which is surrounded by a complex mesh of hemicellulose and lignin (99). Hemicellulose is a heterogeneous polymer composed of five- and six-carbon sugars, such as xylose, arabinose, mannose, and uronic acids. Lignin is an amorphous polymer formed from variable, highly cross-linked phenolic alcohols. In contrast to most other phyla of bacteria, Actinobacteria are enriched in the full suite of enzymes necessary to deconstruct plant biomass (14). This enrichment in enzymes, combined with the abundance of Actinobacteria in soil and the ability of some isolates to grow on plant biomass or its constituent polymers, supports their contribution to plant biomass degradation in soil in conjunction with fungi and other organisms (41, 77, 90). Several recent *in vivo* studies provide further validation (54); for example, in Arctic peat soil, Tveit and colleagues (138) found that most cellulase and hemicellulase RNA transcripts were assigned to Actinobacteria.

Bacteria in the genus *Streptomyces* are among the most abundant organisms in soil (62), and the ability of some isolates to grow on cellulose was first identified ~100 years ago (145).

Subsequently, a variety of *Streptomyces* strains isolated from soil were shown to grow on cellulose (16, 120, 144), hemicellulose (45, 61), and potentially even lignin (34, 163) (Figure 1.2b). However, the majority of strains tested cannot grow on crystalline cellulose or hemicellulose, suggesting that most *Streptomyces* species in the soil only use the breakdown products of plant cell wall polymers (18). Other genera of Actinobacteria have also been implicated in plant biomass degradation in the soil, including *Cellulomonas*, *Nocardia*, and *Micromonospora*, again on the basis of their high prevalence in soil and the ability of some isolates to grow on plant biomass-derived substrates (57, 90, 163).

Free-living Actinobacteria contribute to carbon cycling in other environments enriched in decaying plant material, including leaf litter, compost, and manure (91, 116). In many composting systems, genera of Actinobacteria, including *Thermobifida*, *Saccharomonospora*, and *Thermobispora*, are abundant community members during the thermophilic stage (56, 127, 159). For example, metagenomic analysis of rice straw within composting cow manure showed that Actinobacteria were abundant and encoded more cellulases than any other phylum (149). Similar to compost, leaf litter is also host to a high abundance of Actinobacteria, and metagenomic sequencing identified cellulases from *Streptomyces* and *Thermobispora* genera, as well as from many other Actinobacteria (13, 70, 89, 162). *Micromonospora* strains are common in lake sediment, another environment rich in decaying plant material, and isolates in this genus have been shown to degrade cellulose and hemicellulose (35, 48).

Actinobacteria are also implicated in the cycling of a variety of other carbon sources. This includes chitin, the second most abundant organic carbon source in nature, which is a structural polymer in the cell wall of fungi and in the exoskeletons of invertebrates. As with cellulose, few organisms can degrade chitin, but some Actinobacteria, including most *Streptomyces* and *Arthrobacter* species, have this ability (68, 92). The Ac1 Actinobacteria, dominant organisms in freshwater, are also thought to degrade chitin or uptake degradation by-products of chitin from fungi, crustaceans, and diatoms (9, 46). Additionally, some Actinobacteria are able to degrade hydrocarbons and organic contaminants (5, 85). For example, *Rhodococcus* can degrade diverse natural and synthetic molecules, including chlorinated or aromatic hydrocarbons, nitroaromatics, sulfonated azo dyes, and pesticides (10).

#### **1.4.2 Role of Actinobacteria in Host Associations**

Although Actinobacteria are still generally viewed as free-living bacteria, the recent increased focus on symbiotic associations has revealed their common association with eukaryotic hosts (18, 49, 84). They have broadly been identified in association with animals, both in their guts and on external surfaces. For example, the genera *Bifidobacterium* and *Corynebacterium* are associated with animals ranging from insects to mammals, including humans (137, 143). Additionally, Actinobacteria are associated with plants, including in the rhizosphere and as endophytes (32, 84, 115). Efforts are underway to determine the ecological roles of these bacteria, from aiding hosts with nutrient acquisition to serving as defensive mutualists.

A range of animals and plants rely on Actinobacteria to supplement their diets or help digest

complex food sources. Surveys of insect guts, using 16S rRNA gene sequencing, indicate that Actinobacteria are more abundant in the guts of insects that consume detritus or decaying wood, especially termites (31, 64). Strains that can degrade plant biomass have been isolated from termites and from a range of other insects associated with wood niches, including *Sirex* woodwasps and mountain pine beetles (Figure 1.2b) (1, 18, 98, 151). Other insects, especially some true bugs, rely on Actinobacteria to provide nutritional supplementation (65). For example, *Rhodococcus rhodnii* provides B vitamins to *Rhodnius prolixus*, the insect vector of Chagas disease (20), and likewise, *Coriobacteriaceae* strains provide B vitamins to Pyrrhocoridae (Hemiptera), such as firebugs (110). In the mammalian gut, *Bifidobacterium* species aid in digestion of complex carbohydrates, including plant-derived oligosaccharides and gastric mucin (94, 105). The most clearly defined nutritional association between plants and Actinobacteria is the association between nitrogen-fixing *Frankia* strains and actinorhizal plants (Figure 1.2c) (30), which enables plants to colonize early during primary succession (78, 118, 141).

Actinobacteria also function as defensive mutualists for a number of eukaryotic hosts. A well-studied example of this relationship is in fungus-growing ants (see Sidebar 1.1 and Figure 1.2). Similarly, solitary digger wasps, commonly known as beewolves, inoculate *Streptomyces* symbionts on their larval brood cells to protect cocoons from microbial infection (66), and southern pine beetles harbor a symbiotic *Streptomyces* species to protect their fungal food source (117). Plants are protected by the genera *Streptomyces* and *Micromonospora*, as well as by other Actinobacteria in the rhizosphere, which can help inhibit plant pathogens through the production of secondary metabolites (57, 81, 115). Also, endophytic Actinobacteria suppress pathogens and activate the plant immune system (32, 50). In addition to the nutritional roles discussed above,

*Bifidobacterium* species in the human gut function as defensive mutualists through antimicrobial activity against pathogens and potentially through aiding in immune system development and function (86, 121, 131). Finally, in marine environments, Actinobacteria with diverse secondary metabolism potential are found in association with sponges, but their ecological functions in these environments have not been fully elucidated (119, 153). Although the diversity of Actinobacteria-host defensive relationships has only been explored in a small subset of organisms, it is likely that host defense by Actinobacteria is widespread.

In contrast to these beneficial symbioses, Actinobacteria can also be virulent pathogens of plants and animals. The diverse genus *Mycobacterium* causes tuberculosis (*M. tuberculosis*), leprosy (*M. leprae*), and additional diseases in humans and other mammals (Figure 1.2d) (111). Actinobacteria also cause a range of other diseases in humans, including diphtheria (*Corynebacterium diphtheria*), Whipple's disease (*Tropheryma whipplei*), and bacterial vaginosis (*Gardnerella*). In plants, *Streptomyces scabies* and other plant pathogenic *Streptomyces* strains share a pathogenicity island, allowing them to infect higher plants and cause diseases including, most famously, potato scab (69, 82). Other common plant pathogens fall within the *Microbacteriaceae* family, including *Leifsonia xyli*, which is the cause of ratoon stunting disease in sugarcane, and *Clavibacter michiganensis*, which infects xylem and causes plant wilting in alfalfa, corn, tomato, and potato (21, 47). Finally, leafy gall syndrome, which infects a wide range of dicotyledonous herbaceous plants, is caused by *Rhodococcus fascians* (128, 129).

## 1.5 Actinobacteria in Bioenergy



As of 2013, fossil fuels (petroleum, coal, and natural gas) supplied 80% of the world's energy (60). The negative climate impacts of burning fossil fuels has driven worldwide interest in sustainable, economical production of liquid fuels and other chemicals from renewable sources (40). Use of renewable electric sources, such as solar, wind, and hydroelectric, is expanding, but high-energy density liquid fuels are needed to move ourselves and our goods across continents and oceans (29, 40). Plants and some microbes capture the energy of the sun via photosynthetic carbon dioxide fixation. Efforts to engineer photosynthetic plants, algae, or cyanobacteria to produce fuel compounds are underway and have long-term potential (79, 108, 142, 148). However, near term advances require tapping into stored energy in the plant cell wall. To develop a sustainable biofuel industry, microbes are needed to convert plant biomass into soluble sugars and to transform these sugars into a wide range of desirable compounds (51). Here, we argue, in large part because of their ecology and evolutionary history, Actinobacteria have potential to contribute to both of these challenges.

#### ***1.5.1 Actinobacteria as a source of enzymes in biofuels***

Industrial production of fuels from plant biomass represented less than 5% of energy used for transportation in the United States in 2014 (139). This energy is largely derived from the starch in corn kernels, which is readily converted into ethanol to supplement gasoline (154). However, the use of edible carbohydrates for fuel directly competes with their use as food (135). In contrast, using the abundant, nonedible polysaccharide and lignin fractions of corn plants, other grasses, and trees to make fuels and chemicals is a promising alternative and could remove more carbon from the atmosphere than is released during production and use of the fuel (53). Currently, industrial lignocellulose hydrolysis is carried out using enzyme mixtures from

engineered fungi, such as *Trichoderma reesei* and *Aspergillus niger* (101). However, the enzymatic release of intact clean sugar from plant cell walls remains an important bottleneck to efficient biofuel production (55). As detailed below, the diversity and biochemistry of actinobacterial cellulases, evolved within plant degradation niches, have promise for integration into a successful cellulosic biofuels industry (also see Sidebar 1.2).

*Sidebar 1.2. Integrating actinobacterial enzymes in industrial cellulosic biofuel production*

A number of aspects of the evolution, ecology, and physiology of *Actinobacteria* make them well suited for integration with cellulosic biofuel production. First, Actinobacteria evolved to break down plant biomass under diverse environmental conditions, including under high temperatures, varying pH levels, and anoxic or aerobic states. Similarly, industrial biofuel production requires enzymes that function under diverse conditions, depending on factors such as the biomass composition and pretreatment protocol. The identification of Actinobacteria that have evolved within particular niches can lead to cellulases suitable for specific industrial settings. For example, thermostable enzymes are found in *Thermobifida fusca* and *Thermomonospora curvata* from heated composting environments and in *Acidothermus cellulolyticus* from acidic hot springs (107, 130, 136, 155). Furthermore, given that Actinobacteria freely secrete their enzymes and have evolved in some niches in close association with fungi, actinobacterial cellulases are promising candidates to complement fungal enzymes for plant biomass deconstruction. Finally, Actinobacteria are attractive for industrial use because of their history as successful workhorse organisms in industrial processes (42), the ability to engineer strains that constitutively secrete cellulases at high levels (18, 88), and the resistance of their cell walls to processes, such as milling, that may be important in sustainable cellulosic biofuel production (100).

The ecological roles of Actinobacteria in plant biomass degradation, in soil, in compost, and in association with herbivorous animals have selected for the evolution of high cellulolytic abilities (18). Indeed, a number of model cellulolytic bacteria occur in this phylum, such as *Cellulomonas* spp. and *Thermobifida fusca*, whose study has helped advance our understanding of the

enzymology of cellulose degradation (28, 59, 155). More broadly, the cellulolytic abilities of this phylum are supported by genomic and proteomic analyses, which indicate that actinobacterial genomes have an amazing diversity of enzyme fold families and organizational architectures used to hydrolyze plant biomass compared to most other phyla of bacteria (14). In particular, of the genera analyzed for this review, the genomes of *Streptomyces* and *Micromonospora*, some strains of *Mycobacterium*, and one strain of *Propionibacterium*, are particularly enriched in carbohydrate-active enzyme (CAZy) genes (Figure 1.3a) (80). Within these enriched genera, the diversity of CAZy classes varies: Eighty-six glycoside hydrolase (GH) classes are found across 42 *Streptomyces* genomes, whereas only 35 GH classes are found across 149 *Mycobacterium* genomes (Table A1.3).

Here, we focus on the properties of *Streptomyces* cellulases. The cellulolytic system of *Streptomyces* was defined in *S. reticuli* by Schrempf and colleagues (112–114, 144, 147), whose investigations included secreted cellulolytic enzymes, carbohydrate binding proteins, sugar transporters, and gene regulation. Recent studies have expanded on this work by studying host-associated *Streptomyces*, such as *Streptomyces* sp. SirexAA-E, isolated from a *Sirex* woodwasp, which secretes enzymes with comparable specific activity to an enzyme cocktail from *Trichoderma reesei* (18, 132).

As in many Actinobacteria, cellulolytic *Streptomyces* species rely on a simple, freely secreted catalytic assembly to efficiently deconstruct cellulose, including endoglucanases (GH5, GH9, and GH12) and reducing- and nonreducing-end exoglucanases (GH48 and GH6, respectively). The cellobiose produced by these enzymes is cleaved into glucose by  $\beta$ -glucosidases (GH1 or

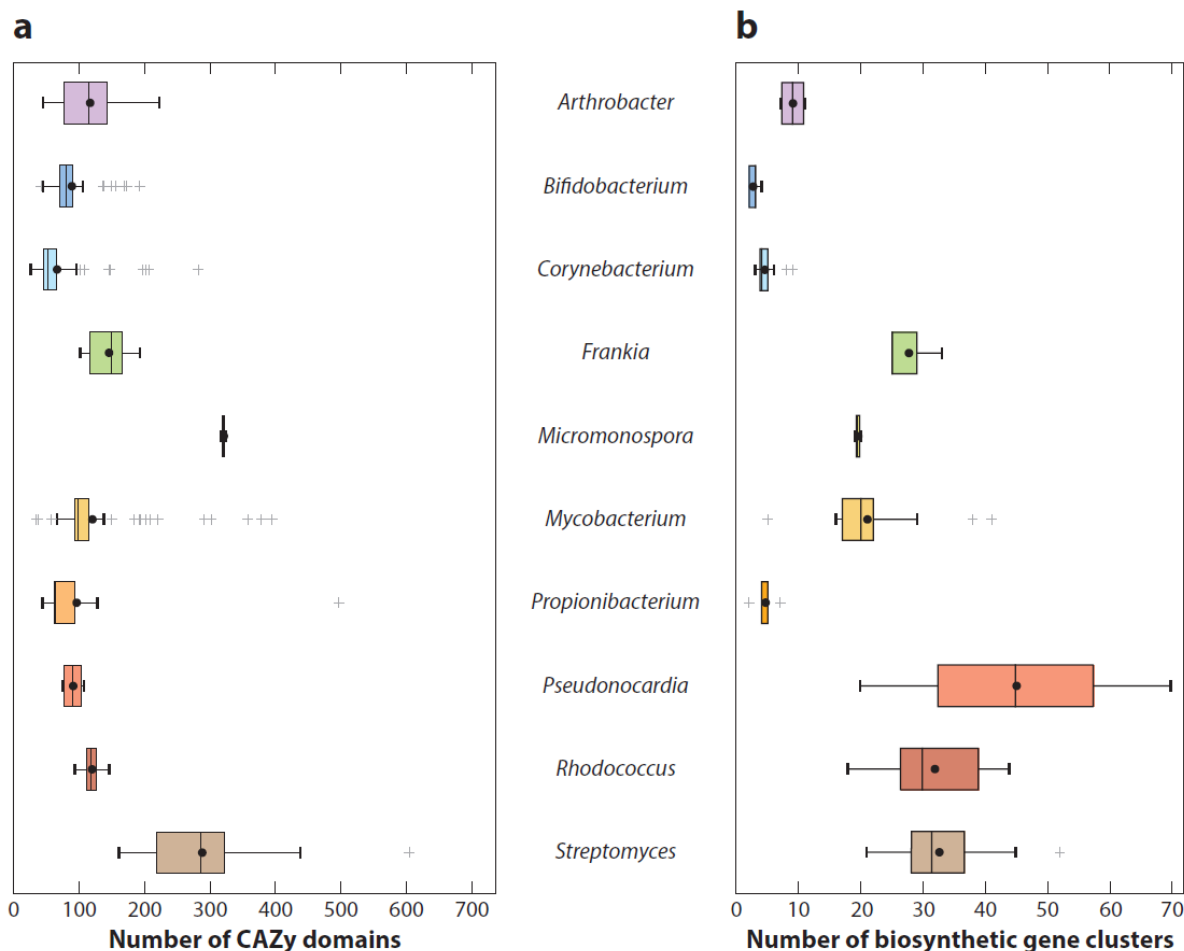


**Figure 1.3. Enrichment of (a) CAZy domains and (b) biosynthetic gene clusters in select actinobacterial genera** (Table A1.3). Each median value is indicated by the line within the box, each mean value is indicated by the black dot, and outliers are indicated with gray plus signs. CAZy domains were identified using the CAZy database (80). Putative biosynthetic gene clusters were identified through genome searches for conserved elements of polyketide, terpene, and nonribosomal peptide biosynthesis enzymes. Hits <30 kb apart were called the same cluster. The number of clusters may be overestimated in draft genomes.

GH3) in the cytoplasm (Figure 1.4a) (17, 18, 132). In addition, the dominant carbohydrate-binding module in these organisms is carbohydrate-binding module family 2. Parallel minimal catalytic assemblies deconstruct other plant cell wall polysaccharides, e.g., xylan (GH10 and GH11), mannan (GH5), and chitin (GH18, GH19, and GH46) (132). Carbohydrate esterases

(CEs) that remove sugar modifications and break linkages between hemicellulose and lignin include families CE4 and CE7 in *Streptomyces* sp. SirexAA-E (132).

Cellulose-oxidizing lytic polysaccharide monooxygenases from the auxiliary activity (AA) family AA10 enzymes primarily are found in Actinobacteria (19). These enzymes likely contribute to their high cellulolytic ability of Actinobacteria and provide an alternative to hydrolytic endoglucanases (Figure 1.4a) (63, 140). Interestingly, although *Streptomyces* are enriched in AA10 enzymes, *Streptomyces* do not appear to secrete an electron donor (e.g., cellobiose dehydrogenase or exogenous donors, such as ascorbate, gallate, or quinol), which is required for lytic polysaccharide monooxygenase function (17, 18, 132). Possibly instead, *Streptomyces* species depend on community association with fungal partners to provide this electron donor (104).

Nonreducing-end exoglucanase (GH6) genes are also expanded in Actinobacteria relative to many other groups of bacteria and fungi (14). Genetic and biochemical analyses have indicated that GH6 enzymes are necessary but not sufficient for cellulose degradation (7, 18, 157). Evolutionary analyses indicate that cellulolytic *Streptomyces* strains evolved to degrade cellulose by maintaining GH6 enzymes, which are highly expressed and secreted during growth on plant biomass (17, 18, 132).

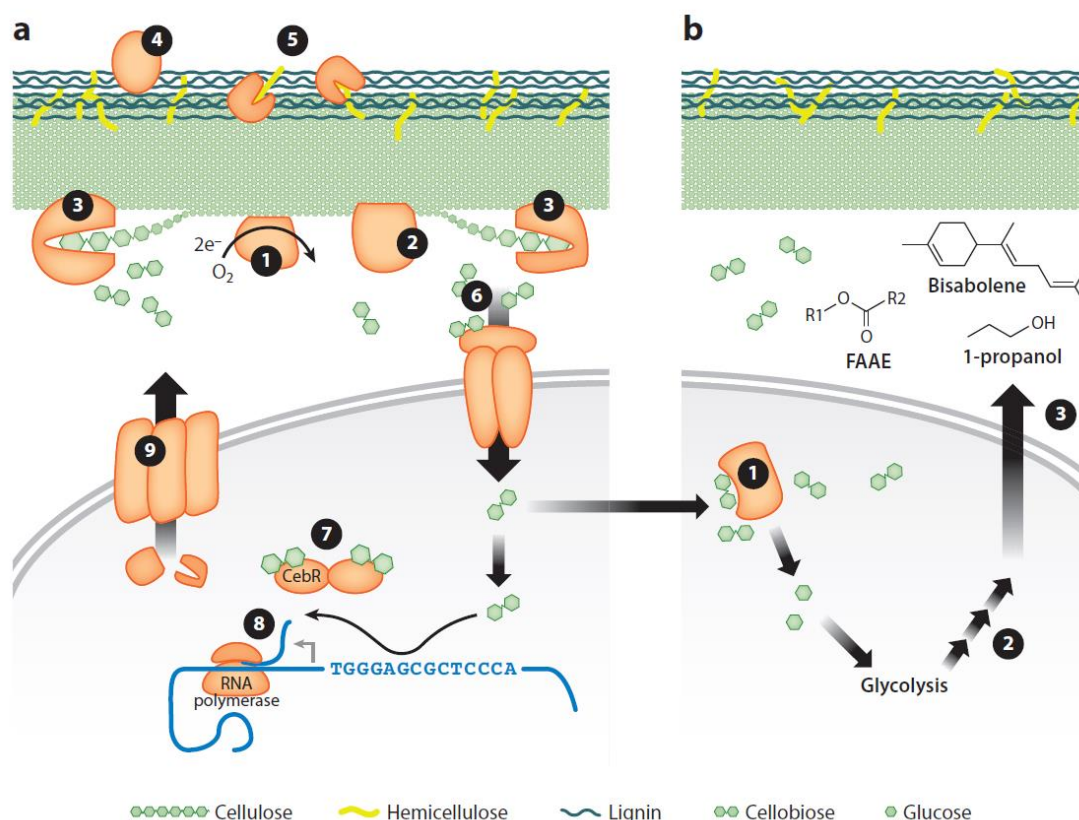


**Figure 1.4. Molecular details of the proposed role of Actinobacteria in biofuels.** Molecular details of the proposed role of Actinobacteria in biofuels. **(a)** Use of Actinobacteria for the deconstruction of plant biomass. (1) Lytic polysaccharide monooxygenases and (2) endoglucanases create internal cuts in the cellulose structure, creating free ends. (3) Reducing- and nonreducing-end exoglucanases break cellulose chains into glucose and cellobiose. Additionally, (4) peroxidases may break apart lignin, and (5) hemicellulases hydrolyze xylan, mannan, and other hemicelluloses into constituent sugars. (6) Cellobiose is imported into the cell, where it (7) binds to the CebR repressor, releasing the repressor from the DNA and (8) allowing RNA polymerase to transcribe cellulase genes. (9) These cellulases are secreted from the cell using the Tat or Sec pathways. **(b)** Use of Actinobacteria for the production of biofuels. (1) Cellobiose also is broken by  $\beta$ -glucosidases into glucose, which enters central metabolism. (2) These sugars are processed in the cell and (3) can be used for production of secreted fuel compounds, such as bisabolene, 1-propanol, or various FFAE species where R1 is a small alkyl group and R2 is an aliphatic hydrocarbon. Currently, there are limited examples of cells that can both degrade plant biomass and produce fuel compounds, but we propose this scheme as one possible future direction in actinobacterial biofuels research.

### **1.5.2 Actinobacteria, advanced biofuels, and biocommodities**

The enzymatic capabilities described above allow Actinobacteria to deconstruct plant biomass into simple sugars that can be converted into biofuels and other bioproducts. Here, we focus on how Actinobacteria, having evolved to produce diverse chemicals, may contribute to the production of these bioproducts.

Although bioethanol from corn and cane sugars and biodiesel from plant oils have already seen widespread adoption, increasing their use as transportation fuels faces substantial hurdles, including storage, engine infrastructure, hygroscopy, and land-use concerns (3, 4, 25, 102). New bioproducts, including microbially produced specialty biofuels, more closely match the properties of traditional fuels and thus are highly desirable (4). Specialty biofuels include higher-alcohol biofuels, fatty acid alkyl esters (FAAEs or biodiesel), triacylglycerols (TAGs) that can serve as fuel precursors, and a diverse family of energy-dense isoprenoid compounds that can be used as diesel or jet-fuel replacements (11, 27, 39, 73, 102). Also, bio-based production of fuel additives and commodity chemicals, especially those currently obtained from petroleum, would help support a bioenergy industry. Actinobacteria are a particularly promising group to contribute to the development of these classes of chemicals or their precursors, especially those developed through alternative biosynthesis routes, given their well-known capacity to produce and secrete a diversity of compounds, the ability to engineer their biosynthetic pathways, and their long history of use in industrial processes (Figure 1.4b).

Several promising studies support the potential utility of Actinobacteria for biocommodity production (**Table A1.4**). For example, *Corynebacterium glutamicum* has been targeted to

produce specialty biofuels because of its innate isobutanol tolerance and ability to produce large amounts of glutamate and various 2-keto acids, which are precursors to branched-chain alcohols. Genetic engineering and optimization of *C. glutamicum* led to strains that could produce isobutanol at similar yields as in optimized *Escherichia coli* strains (15, 123, 158). The oleaginous *Rhodococcus opacus* was shown to accumulate high yields of high-energy density TAG (74). Additionally, native long-chain FFAE (biodiesel) production has been demonstrated in a *Streptomyces* strain isolated from sheep feces (83), and an isoprenoid biosynthesis pathway in *Streptomyces venezuelae* was combined with a cyclase gene from *Abies grandis* (grand fir) to produce bisabolene, another alternative to diesel fuel (103). 1-Propanol, an industrially relevant solvent with good fuel properties, was produced in *T. fusca* at a titer of 0.5 g/L after the addition of an aldehyde/alcohol dehydrogenase, approaching the yields possible in *E. coli* (44). Transcriptomic and metabolomic analyses suggested that additional pathway engineering of the intermediates would result in increased yields (43).

These studies represent a few of the many possible options to engineer Actinobacteria to produce bioproducts. The evolution of diverse natural product pathways in Actinobacteria, especially in genera such as *Pseudonocardia*, *Rhodococcus*, and *Streptomyces* (Figure 1.3b), make these organisms attractive as sources for discovery of biosynthetic enzymes and as hosts for pathway engineering and heterologous expression (103). The use of terpenoid biosynthetic machinery for bisabolene production in *Streptomyces* represents just one possible adaptation of natural product pathways for biofuel production. Polyketide natural products represent another source of industrially relevant molecules from Actinobacteria (160). Potentially, rational engineering of modular polyketide synthases can lead to molecules with extremely diverse structural and functional characteristics (71, 102).



Combining the natural abilities of Actinobacteria both to efficiently degrade plant biomass and to produce diverse chemicals make them well suited to contribute to the long-term goal of deploying consolidated bioprocessing (CBP) (Figure 1.4). The potential of Actinobacteria to contribute to CBP was illustrated by the FAAE-producing *Streptomyces* species, which grew readily on cellulose and xylan and produced higher biodiesel titers on complex carbohydrates than CBP-engineered *E. coli* (83, 125). The TAG-accumulating *R. opacus* was engineered with *Streptomyces* xylose metabolism genes (76) and subjected to directed evolution for high production of TAG with lignocellulosic biomass as a carbon source (75). Use of CBP approaches with Actinobacteria may also be possible for commodity chemicals. For example, the strain of *T. fusca* engineered for 1-propanol production produced higher titers when grown on a complex biomass, including cellulose and switchgrass, than on refined sugars. Similarly, *Streptomyces maritimus* produced benzoate, a food preservative and plastics additive, at higher yields when engineered to grow on cellulose as compared to its growth on simple sugars (96). Furthermore, *Streptomyces lividans* modified with a *T. fusca* endoglucanase (for cellulose consumption) and two biosynthesis genes could produce 4-vinylphenol, a compound used in the plastics and electronics industries (95).

## 1.6 Conclusions

As our main source of antibiotics, Actinobacteria became famous for helping contribute to one of the greatest achievements of the twentieth century—vastly reducing the impact of infectious diseases. This century, we have an urgent and critical need for bio-based approaches that help meet our energy requirements while reducing carbon emissions. As described in this review,

efforts are underway to leverage the diverse ecology of Actinobacteria to identify enzymes, pathways, and biosynthetic products that can contribute to biofuel and chemical production. We believe that, in the twenty-first century, Actinobacteria have potential to further increase in prominence as they help us meet our global needs for renewable biofuels and other bioproducts while mitigating the dangers of climate change.

## 1.7 References

1. Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, et al. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. *ISME J.* 5(8):1323–31
2. Ahmed A, Earl J, Retchless A, Hillier SL, Rabe LK, et al. 2012. Comparative genomic analyses of 17 clinical isolates of *Gardnerella vaginalis* provide evidence of multiple genetically isolated clades consistent with subspeciation into genovars. *J. Bacteriol.* 194(15):3922–37
3. Atabani AE, Silitonga AS, Badruddin IA, Mahlia TMI, Masjuki HH, Mekhilef S. 2012. A comprehensive review on biodiesel as an alternative energy resource and its characteristics. *Renew. Sustain. Energy Rev.* 16:2070–93
4. Atsumi S, Hanai T, Liao JC. 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451:86–89
5. Barabas G, Vargha G, Szabo IM, Penyige A, Damjanovich S, et al. 2001. *n*-Alkane uptake and utilisation by *Streptomyces* strains. *Antonie Van Leeuwenhoek* 79:269–76
6. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, et al. 2016. Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* 80(1):1–43
7. Barr BK, Hsieh YL, Ganem B, Wilson DB. 1996. Identification of two functionally different classes of exocellulases. *Biochemistry* 35(2):586–92

8. Battistuzzi FU, Feijao A, Hedges SB. 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol. Biol.* 4:44
9. Beier S, Bertilsson S. 2011. Uncoupling of chitinase activity and uptake of hydrolysis products in freshwater bacterioplankton. *Limnol. Oceanogr.* 56(4):1179–88
10. Bell KS, Philp JC, Aw DWJ, Christofi N. 1998. A review: the genus *Rhodococcus*. *J. Appl. Microbiol.* 85:195–210
11. Beller HR, Lee TS, Katz L. 2015. Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids. *Nat. Prod. Rep.* 32:1508–26
12. Berenbaum MR, Eisner T. 2008. Bugs' bugs. *Science* 322:52–53
13. Berlemont R, Allison SD, Weihe C, Lu Y, Brodie EL, et al. 2014. Cellulolytic potential under environmental changes in microbial communities from grassland litter. *Front. Microbiol.* 5:639
14. Berlemont R, Martiny AC. 2013. Phylogenetic distribution of potential cellulases in bacteria. *Appl. Environ. Microbiol.* 79(5):1545–54
15. Blombach B, Riester T, Wieschalka S, Ziert C, Youn J-W, et al. 2011. *Corynebacterium glutamicum* tailored for efficient isobutanol production. *Appl. Environ. Microbiol.* 77(10):3300–10
16. Bontemps C, Toussaint M, Revol P-V, Hotel L, Jeanbille M, et al. 2013. Taxonomic and functional diversity of *Streptomyces* in a forest soil. *FEMS Microbiol. Lett.* 342(2):157–67
17. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, et al. 2014. Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. *Appl. Environ. Microbiol.* 80(15):4692–701

18. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Wendt-Pienkowski E, et al. 2016. Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression. *PLOS Biol.* 14(6):e1002475
19. Book AJ, Yennamalli RM, Takasuka TE, Currie CR, Phillips GN Jr, Fox BG. 2014. Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases. *Biotechnol. Biofuels* 7:109
20. Brecher G, Wigglesworth VB. 1944. The transmission of *Actinomyces rhodnii* Erikson in *Rhodnius prolixus* stal (Hemiptera) and its influence on the growth of the host. *Parasitology* 35(04):220–24
21. Brumbley SM, Petrasovits LA, Hermann SR, Young AJ, Croft BJ. 2006. Recent advances in the molecular biology of *Leifsonia xyli* subsp. *xyli*, causal organism of ratoon stunting disease. *Australas. Plant Pathol.* 35(6):681–89
22. Cafaro MJ, Currie CR. 2005. Phylogenetic analysis of mutualistic filamentous bacteria associated with fungus-growing ants. *Can. J. Microbiol.* 51(6):441–46
23. Cafaro MJ, Poulsen M, Little AEF, Price SL, Gerardo NM, et al. 2011. Specificity in the symbiotic association between fungus-growing ants and protective *Pseudonocardia* bacteria. *Proc. R. Soc. B* 278:1814–22
24. Caldera EJ, Currie CR. 2012. The population structure of antibiotic-producing bacterial symbionts of *Apterostigma dentigerum* ants: impacts of coevolution and multipartite symbiosis. *Am. Nat.* 180(5):604–17
25. Caspeta L, Nielsen J. 2013. Economic and environmental impacts of microbial biodiesel. *Nat. Biotechnol.* 31(9):789–93

26. Chandra G, Chater KF. 2014. Developmental biology of *Streptomyces* from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* 38:345–79
27. Choi YJ, Lee J, Jang Y-S, Lee SY. 2014. Metabolic engineering of microorganisms for the production of higher alcohols. *mBio* 5(5):e01524-14
28. Christopherson MR, Suen G, Bramhacharya S, Jewell KA, Aylward FO, et al. 2013. The genome sequences of *Cellulomonas fimi* and “*Cellvibrio gilvus*” reveal the cellulolytic strategies of two facultative anaerobes, transfer of “*Cellvibrio gilvus*” to the genus *Cellulomonas*, and proposal of *Cellulomonas gilvus* sp. nov. *PLOS ONE* 8(1):e53954
29. Chu S, Majumdar A. 2012. Opportunities and challenges for a sustainable energy future. *Nature* 488:294–303
30. Clawson ML, Bourret A, Benson DR. 2004. Assessing the phylogeny of *Frankia*-actinorhizal plant nitrogen-fixing root nodule symbioses with *Frankia* 16S rRNA and glutamine synthetase gene sequences. *Mol. Phylogenet. Evol.* 31:131–38
31. Colman DR, Toolson EC, Takacs-Vesbach CD. 2012. Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* 21:5124–37
32. Conn VM, Walker AR, Franco CM. 2008. Endophytic Actinobacteria induce defense pathways in *Arabidopsis thaliana*. *Mol. Plant-Microbe Interact.* 21(2):208–18
33. Coombs JT, Michelsen PP, Franco CMM. 2004. Evaluation of endophytic Actinobacteria as antagonists of *Gaeumannomyces graminis* var. *tritici* in wheat. *Biol. Control* 29(3):359–66
34. Crawford DL. 1978. Lignocellulose decomposition by selected *Streptomyces* strains. *Appl. Environ. Microbiol.* 35(6):1041–45
35. Cross T. 1981. Aquatic actinomycetes: a critical survey of the occurrence, growth and role of actinomycetes in aquatic habitats. *J. Appl. Bacteriol.* 50(3):397–423

36. Currie CR, Mueller UG, Malloch D. 1999. The agricultural pathology of ant fungus gardens. *PNAS* 96(14):7998–8002
37. Currie CR, Poulsen M, Mendenhall J, Boomsma JJ, Billen J. 2006. Coevolved crypts and exocrine glands support mutualistic bacteria in fungus-growing ants. *Science* 311(5757):81–83
38. Currie CR, Scott JA, Summerbell RC, Malloch D. 1999. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* 398:701–4
39. d’Espaux L, Mendez-Perez D, Li R, Keasling JD. 2015. Synthetic biology for microbial production of lipid-based biofuels. *Curr. Opin. Chem. Biol.* 29:58–65
40. Dale BE, Ong RG. 2012. Energy, wealth, and human development: Why and how biomass pretreatment research must improve. *Biotechnol. Prog.* 28(4):893–98
41. de Boer W, Folman LB, Summerbell RC, Boddy L. 2005. Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol. Rev.* 29:795–811
42. Demain AL. 2007. The business of biotechnology. *Ind. Biotechnol.* 3(3):269–83
43. Deng Y, Fisher AB, Fong SS. 2015. Systematic analysis of intracellular mechanisms of propanol production in the engineered *Thermobifida fusca* B6 strain. *Appl. Microbiol. Biotechnol.* 99(19):8089–8100
44. Deng Y, Fong SS. 2011. Metabolic engineering of *Thermobifida fusca* for direct aerobic bioconversion of untreated lignocellulosic biomass to 1-propanol. *Metab. Eng.* 13:570–77
45. Ding CH, Jiang ZQ, Li XT, Li LT, Kusakabe I. 2004. High activity xylanase production by *Streptomyces olivaceoviridis* E-86. *World J. Microbiol. Biotechnol.* 20:7–10
46. Eckert EM, Baumgartner M, Huber IM, Pernthaler J. 2013. Grazing resistant freshwater bacteria profit from chitin and cell-wall-derived organic carbon. *Environ. Microbiol.* 15(7):2019–30

47. Eichenlaub R, Gartemann K-H. 2011. The *Clavibacter michiganensis* subspecies: molecular investigation of gram-positive bacterial plant pathogens. *Annu. Rev. Phytopathol.* 49:445–64
48. Erikson D. 1941. Some studies on lake-mud strains of *Micromonospora*. *J. Bacteriol.* 41(3):277–300
49. Florez LV, Biedermann PHW, Engl T, Kaltenpoth M. 2015. Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat. Prod. Rep.* 32:904–36
50. Franco C, Michelsen P, Percy N, Conn V, Listiana E, et al. 2007. Actinobacterial endophytes for improved crop performance. *Australas. Plant Pathol.* 36:524–31
51. Fulton LM, Lynd LR, Korner A, Greene N, Tonachel LR. 2015. The need for biofuels as part of a low carbon energy future. *Biofuels Bioprod. Biorefining* 9:476–83
52. Gao B, Gupta RS. 2012. Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol. Mol. Biol. Rev.* 76(1):66–112
53. Gelfand I, Snapp SS, Robertson GP. 2010. Energy efficiency of conventional, organic, and alternative cropping systems at a site in the U.S. Midwest. *Environ. Sci. Technol.* 44(10):4006–11
54. Gittel A, Barta J, Kohoutova I, Mikutta R, Owens S, et al. 2014. Distinct microbial communities associated with buried soils in the Siberian tundra. *ISME J.* 8:841–53
55. Harris PV, Xu F, Kreel NE, Kang C, Fukuyama S. 2014. New enzyme insights drive advances in commercial ethanol production. *Curr. Opin. Chem. Biol.* 19:162–70
56. Herrmann RF, Shann JF. 1997. Microbial community changes during the composting of municipal solid waste. *Microb. Ecol.* 33:78–85
57. Hirsch AM, Valdes M. 2010. *Micromonospora*: an important microbe for biomedicine and potentially for biocontrol and biofuels. *Soil Biol. Biochem.* 42:536–42

58. Hopwood DA. 2007. *Streptomyces in Nature and Medicine: The Antibiotic Makers*. New York: Oxford Univ. Press
59. Hsing W, Canale-Parola E. 1992. Cellobiose chemotaxis by the cellulolytic bacterium *Cellulomonas gelida*. *J. Bacteriol.* 174(24):7996–8002
60. Int. Energy Agency. 2015. *2015 Key World Energy Statistics*. Paris: Int. Energy Agency
61. Ishaque M, Kluepfel D. 1981. Production of xylanolytic enzymes by *Streptomyces flavogriseus*. *Biotechnol. Lett.* 3(9):481–86
62. Janssen PH. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72(3):1719–28
63. Johansen KS. 2016. Discovery and industrial applications of lytic polysaccharide mono-oxygenases. *Biochem. Soc. Trans.* 44:143–49
64. Jones RT, Sanchez LG, Fierer N. 2013. A cross-taxon analysis of insect-associated bacterial diversity. *PLOS ONE* 8(4):e61218
65. Kaltenpoth M. 2009. Actinobacteria as mutualists: general healthcare for insects? *Trends Microbiol.* 17(12):529–35
66. Kaltenpoth M, Gottler W, Herzner G, Strohm E. 2005. Symbiotic bacteria protect wasp larvae from fungal infestation. *Curr. Biol.* 15:475–79
67. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772–80
68. Kawase T, Saito A, Sato T, Kanai R, Fujii T, et al. 2004. Distribution and phylogenetic analysis of family 19 chitinases in Actinobacteria. *Appl. Environ. Microbiol.* 70(2):1135–44



69. Kers JA, Cameron KD, Joshi MV, Bukhalid RA, Morello JE, et al. 2005. A large, mobile pathogenicity island confers plant pathogenicity on *Streptomyces* species. *Mol. Microbiol.* 55(4):1025–33
70. Kim M, Kim W-S, Tripathi BM, Adams J. 2014. Distinct bacterial communities dominate tropical and temperate zone leaf litter. *Microb. Ecol.* 67(4):837–48
71. Klaus M, Ostrowski MP, Austerjost J, Robbins T, Lowry B, et al. 2016. Protein–protein interactions, not substrate recognition, dominates the turnover of chimeric assembly line polyketide synthases. *J. Biol. Chem.* In press. doi: 10.1074/jbc.M116.730531
72. Kroiss J, Kaltenpoth M, Schneider B, Schwinger M-G, Hertweck C, et al. 2010. Symbiotic streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat. Chem. Biol.* 6(4):261–63
73. Kung Y, Runguphan W, Keasling JD. 2012. From fields to fuels: recent advances in the microbial production of biofuels. *ACS Synth. Biol.* 1:498–513
74. Kurosawa K, Boccazzi P, de Almeida NM, Sinskey AJ. 2010. High-cell-density batch fermentation of *Rhodococcus opacus* PD630 using a high glucose concentration for triacylglycerol production. *J. Biotechnol.* 147:212–18
75. Kurosawa K, Laser J, Sinskey AJ. 2015. Tolerance and adaptive evolution of triacylglycerol-producing *Rhodococcus opacus* to lignocellulose-derived inhibitors. *Biotechnol. Biofuels* 8:76
76. Kurosawa K, Wewetzer SJ, Sinskey AJ. 2013. Engineering xylose metabolism in triacylglycerol producing *Rhodococcus opacus* for lignocellulosic fuel production. *Biotechnol. Biofuels* 6:134

77. Lauber CL, Hamady M, Knight R, Fierer N. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 75(15):5111–20
78. Lawrence DB, Schoenike RE, Quispel A, Bond G. 1967. The role of *Dryas drummondii* in vegetation development following ice recession at Glacier Bay, Alaska, with special reference to its nitrogen fixation by root nodules. *J. Ecol.* 55(3):793–813
79. Liu J, Rice A, Mcglew K, Shaw V, Park H, et al. 2015. Metabolic engineering of oilseed crops to produce high levels of novel acetyl glyceride oils with reduced viscosity, freezing point and calorific value. *Plant Biotechnol. J.* 13:858–65
80. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42(1):D490–95
81. Loqman S, Barka EA, Clement C, Ouhdouch Y. 2009. Antagonistic actinomycetes from Moroccan soil to control the grapevine gray mold. *World J. Microbiol. Biotechnol.* 25:81–91
82. Loria R, Kers J, Joshi M. 2006. Evolution of plant pathogenicity in *Streptomyces*. *Annu. Rev. Phytopathol.* 44:469–87
83. Lu Y, Wang J, Deng Z, Wu H, Deng Q, et al. 2013. Isolation and characterization of fatty acid methyl ester (FAME)-producing *Streptomyces* sp. S161 from sheep (*Ovis aries*) faeces. *Lett. Appl. Microbiol.* 57:200–5
84. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, et al. 2012. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488:86–90
85. Margesin R, Mortelmaier C, Mair J. 2013. Low-temperature biodegradation of petroleum hydrocarbons (n-alkanes, phenol, anthracene, pyrene) by four actinobacterial strains. *Int. Biodeterior. Biodegrad.* 84:185–91

86. Marin ML, Lee JH, Murtha J, Ustunol Z, Pestka JJ. 1997. Differential cytokine production in clonal macrophage and T-cell lines cultured with *Bifidobacteria*. *J. Dairy Sci.* 80(11):2713–20
87. Marsh SE, Poulsen M, Pinto-Tomas A, Currie CR. 2014. Interaction between workers during a short time window is required for bacterial symbiont transmission in *Acromyrmex* leaf-cutting ants. *PLOS ONE* 9(7):e103269
88. Marushima K, Ohnishi Y, Horinouchi S. 2009. CebR as a master regulator for cellulose/celooligosaccharide catabolism affects morphological development in *Streptomyces griseus*. *J. Bacteriol.* 191(19):5930–40
89. Matulich KL, Weihe C, Allison SD, Amend AS, Berlemont R, et al. 2015. Temporal variation overshadows the response of leaf litter microbial communities to simulated global change. *ISME J.* 9(11):2477–89
90. McCarthy AJ. 1987. Lignocellulose-degrading actinomycetes. *FEMS Microbiol. Rev.* 46:145–63
91. McCarthy AJ, Williams ST. 1992. Actinomycetes as agents of biodegradation in the environment—a review. *Gene* 115:189–92
92. Metcalfe AC, Krsek M, Gooday GW, Prosser JI, Wellington EMH. 2002. Molecular analysis of a bacterial chitinolytic community in an upland pasture. *Appl. Environ. Microbiol.* 68(10):5042–50
93. Meyer F, Paarmann D, D’Souza M, Olson R, Glass EM, et al. 2008. The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386

94. Milani C, Turrone F, Duranti S, Lugli GA, Mancabelli L, et al. 2016. Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment. *Appl. Environ. Microbiol.* 82(4):980–91
95. Noda S, Kawai Y, Tanaka T, Kondo A. 2015. 4-Vinylphenol biosynthesis from cellulose as the sole carbon source using phenolic acid decarboxylase- and tyrosine ammonia lyase-expressing *Streptomyces lividans*. *Bioresour. Technol.* 180:59–65
96. Noda S, Kitazono E, Tanaka T, Ogino C, Kondo A. 2012. Benzoic acid fermentation from starch and cellulose via a plant-like  $\beta$ -oxidation pathway in *Streptomyces maritimus*. *Microb. Cell Fact.* 11:49
97. Oh D-C, Poulsen M, Currie CR, Clardy J. 2009. Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nat. Chem. Biol.* 5(6):391–93
98. Pasti MB, Pometto AL III, Nuti MP, Crawford DL. 1990. Lignin-solubilizing ability of actinomycetes isolated from termite (Termitidae) gut. *Appl. Environ. Microbiol.* 56(7):2213–18
99. Pauly M, Keegstra K. 2008. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 54:559–68
100. Paye JMD, Guseva A, Hammer SK, Gjersing E, Davis MF, et al. 2016. Biological lignocellulose solubilization: comparative evaluation of biocatalysts and enhancement via cotreatment. *Biotechnol. Biofuels* 9:8
101. Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, et al. 2015. Fungal cellulases. *Chem. Rev.* 115:1308–448
102. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. 2012. Microbial engineering for the production of advanced biofuels. *Nature* 488(7411):320–28

103. Phelan RM, Sekurova ON, Keasling JD, Zotchev SB. 2014. Engineering terpene biosynthesis in *Streptomyces* for production of the advanced biofuel precursor bisabolene. *ACS Synth. Biol.* 4:393–99
104. Phillips CM, Beeson WT IV, Cate JH, Marletta MA. 2011. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem. Biol.* 6:1399–406
105. Pokusaeva K, Fitzgerald GF, van Sinderen D. 2011. Carbohydrate metabolism in *Bifidobacteria*. *Genes Nutr.* 6:285–306
106. Poulsen M, Oh D-C, Clardy J, Currie CR. 2011. Chemical analyses of wasp-associated *Streptomyces* bacteria reveal a prolific potential for natural products discovery. *PLOS ONE* 6(2):e16763
107. Ransom C, Balan V, Biswas G, Dale B, Crockett E, Sticklen M. 2007. Heterologous *Acidothermus cellulolyticus* 1,4- $\beta$ -endoglucanase E1 produced within the corn biomass converts corn stover into glucose. *Appl. Biochem. Biotechnol.* 136–140:207–19
108. Razeghifard R. 2013. Algal biofuels. *Photosynth. Res.* 117:207–19
109. Rosas-Magallanes V, Deschavanne P, Quintana-Murci L, Brosch R, Gicquel B, Neyrolles O. 2006. Horizontal transfer of a virulence operon to the ancestor of *Mycobacterium tuberculosis*. *Mol. Biol. Evol.* 23(6):1129–35
110. Salem H, Bauer E, Strauss AS, Vogel H, Marz M, Kaltenpoth M. 2014. Vitamin supplementation by gut symbionts ensures metabolic homeostasis in an insect host. *Proc. R. Soc. B* 281:20141838

111. Saviola B, Bishai W. 2006. The genus *Mycobacterium*—medical. In *The Prokaryotes*, Vol. 3, ed. M Dworkin, S Falkow, E Rosenberg, K-H Schleifer, E Stackebrandt, pp. 919–33. New York: Springer. 3rd ed.
112. Schlosser A, Jantos J, Hackmann K, Schrempf H. 1999. Characterization of the binding protein dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Appl. Environ. Microbiol.* 65(6):2636–43
113. Schlosser A, Kampers T, Schrempf H. 1997. The *Streptomyces* ATP-binding component MsiK assists in cellobiose and maltose transport. *J. Bacteriol.* 179(6):2092–95
114. Schrempf H, Walter S. 1995. The cellulolytic system of *Streptomyces reticuli*. *Int. J. Biol. Macromol.* 17(6):353–55
115. Schrey SD, Erkenbrack E, Fruh E, Fengler S, Hommel K, et al. 2012. Production of fungal and bacterial growth modulating secondary metabolites is widespread among mycorrhiza-associated streptomycetes. *BMC Microbiol.* 12:164
116. Scott JJ, Budsberg KJ, Suen G, Wixon DL, Balser TC, Currie CR. 2010. Microbial community structure of leaf-cutter ant fungus gardens and refuse dumps. *PLOS ONE* 5(3):e9922
117. Scott JJ, Oh DC, Yuceer MC, Klepzig KD, Clardy J, Currie CR. 2008. Bacterial protection of beetle fungus mutualism. *Science* 322:63
118. Seeds JD, Bishop JG. 2009. Low *Frankia* inoculation potentials in primary successional sites at Mount St. Helens, Washington, USA. *Plant Soil* 323:225–33
119. Seipke RF, Kaltenpoth M, Hutchings MI. 2012. *Streptomyces* as symbionts: an emerging and widespread theme? *FEMS Microbiol. Rev.* 36:862–76

120. Semedo LTAS, Gomes RC, Linhares AA, Duarte GF, Nascimento RP, et al. 2004. *Streptomyces drozdowiczii* sp. nov., a novel cellulolytic streptomycete from soil in Brazil. *Int. J. Syst. Evol. Microbiol.* 54:1323–28
121. Servin AL. 2004. Antagonistic activities of lactobacilli and *Bifidobacteria* against microbial pathogens. *FEMS Microbiol. Rev.* 28:405–40
122. Smanski MJ, Schlatter DC, Kinkel LL. 2015. Leveraging ecological theory to guide natural product discovery. *J. Ind. Microbiol. Biotechnol.* 43:115–28
123. Smith KM, Cho K-M, Liao JC. 2010. Engineering *Corynebacterium glutamicum* for isobutanol production. *Appl. Microbiol. Biotechnol.* 87:1045–55
124. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–13
125. Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, et al. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463:559–62
126. Steffan SA, Chikaraishi Y, Currie CR, Horn H, Gaines-Day HR, et al. 2015. Microbes are trophic analogs of animals. *PNAS* 112(49):15119–24
127. Steger K, Jarvis A, Vasara T, Romantschuk M, Sundh I. 2007. Effects of differing temperature management on development of Actinobacteria populations during composting. *Res. Microbiol.* 158(7):617–24
128. Stes E, Francis I, Pertry I, Dolzblasz A, Depuydt S, Vereecke D. 2013. The leafy gall syndrome induced by *Rhodococcus fascians*. *FEMS Microbiol. Lett.* 342(2):187–94
129. Stes E, Vandeputte OM, El Jaziri M, Holsters M, Vereecke D. 2011. A successful bacterial coup d'état: how *Rhodococcus fascians* redirects plant development. *Annu. Rev. Phytopathol.* 49:69–86

130. Stutzenberger FJ. 1972. Cellulolytic activity of *Thermomonospora curvata*: nutritional requirements for cellulase production. *Appl. Microbiol.* 24(1):77–82
131. Takahashi T, Oka T, Iwana H, Kuwata T, Yamamoto Y. 1993. Immune response of mice to orally administered lactic acid bacteria. *Biosci. Biotechnol. Biochem.* 57(9):1557–60
132. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. 2013. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci. Rep.* 3:1–10
133. Tamura K, Battistuzzi FU, Billings-Ross P, Murillo O, Filipowski A, Kumar S. 2012. Estimating divergence times in large molecular phylogenies. *PNAS* 109(47):19333–38
134. Thompson JN. 1994. *The Coevolutionary Process*. Chicago: Univ. Chicago Press
135. Thompson PB. 2012. The agricultural ethics of biofuels: climate ethics and mitigation arguments. *Poiesis Prax.* 8:169–89
136. Tucker MP, Mohagheghi A, Grohmann K, Himmel ME. 1989. Ultra-thermostable cellulases from *Acidothermus cellulolyticus*: comparison of temperature optima with previously reported cellulases. *Nat. Biotechnol.* 7:817–20
137. Turroni F, van Sinderen D, Ventura M. 2011. Genomics and ecological overview of the genus *Bifidobacterium*. *Int. J. Food Microbiol.* 149:37–44
138. Tveit AT, Urich T, Svenning MM. 2014. Metatranscriptomic analysis of Arctic peat soil microbiota. *Appl. Environ. Microbiol.* 80(18):5761–72
139. US Energy Inf. Adm. 2016. *January 2016 Monthly Energy Review*. Washington, DC: US Energy Inf. Adm.
140. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, et al. 2010. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330:219–22



141. Vitousek PM, Walker LR. 1989. Biological invasion by *Myrica faya* in Hawai'i: plant demography, nitrogen fixation, ecosystem effects. *Ecol. Monogr.* 59(3):247–65
142. Vollmann J, Eynck C. 2015. Camelina as a sustainable oilseed crop: contributions of plant breeding and genetic engineering. *Biotechnol. J.* 10:525–35
143. von Graevenitz A, Bernard K. 2006. The genus *Corynebacterium*—medical. See Ref. 111, pp. 819–42
144. Wachinger G, Bronnenmeier K, Staudenbauer WL, Schrempf H. 1989. Identification of mycelium-associated cellulase from *Streptomyces reticuli*. *Appl. Environ. Microbiol.* 55(10):2653–57
145. Waksman SA. 1918. Studies in the metabolism of actinomycetes II. *J. Bacteriol.* 4:307–30
146. Waksman SA. 1931. Decomposition of the various chemical constituents etc. of complex plant materials by pure cultures of fungi and bacteria. *Arch. Mikrobiol.* 2(1):136–54
147. Walter S, Schrempf H. 1996. The synthesis of the *Streptomyces reticuli* cellulase (avicelase) is regulated by both activation and repression mechanisms. *Mol. Gen. Genet.* 251(2):186–95
148. Wang B, Wang J, Zhang W, Meldrum DR. 2012. Application of synthetic biology in cyanobacteria and algae. *Front. Microbiol.* 3:344
149. Wang C, Dong D, Wang H, Muller K, Qin Y, et al. 2016. Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition. *Biotechnol. Biofuels* 9:1–17
150. Watanabe Y, Martini JE, Ohmoto H. 2000. Geochemical evidence for terrestrial ecosystems 2.6 billion years ago. *Nature* 408:574–78
151. Watanabe Y, Shinzato N, Fukatsu T. 2003. Isolation of actinomycetes from termites' guts. *Biosci. Biotechnol. Biochem.* 67(8):1797–801

152. Weber NA. 1966. Fungus-growing ants. *Science* 153(3736):587–604
153. Webster NS, Taylor MW. 2012. Marine sponges and their microbial symbionts: love and other relationships. *Environ. Microbiol.* 14(2):335–46
154. Weise SE, van Wijk KJ, Sharkey TD. 2011. The role of transitory starch in C<sub>3</sub>, CAM, and C<sub>4</sub> metabolism and opportunities for engineering leaf starch accumulation. *J. Exp. Bot.* 62(9):3109–18
155. Wilson DB. 2004. Studies of *Thermobifida fusca* plant cell wall degrading enzymes. *Chem. Rec.* 4(2):72–82
156. Wilson DB. 2011. Microbial diversity of cellulose hydrolysis. *Curr. Opin. Microbiol.* 14:259–63
157. Wu M, Bu L, Vuong TV, Wilson DB, Crowley MF, et al. 2013. Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. *J. Biol. Chem.* 288(46):33107–17
158. Yamamoto S, Suda M, Niimi S, Inui M, Yukawa H. 2013. Strain optimization for efficient isobutanol production using *Corynebacterium glutamicum* under oxygen deprivation. *Biotechnol. Bioeng.* 110(11):2938–48
159. Yu H, Zeng G, Huang H, Xi X, Wang R, et al. 2007. Microbial community succession and lignocellulose degradation during agricultural waste composting. *Biodegradation* 18:793–802
160. Yuzawa S, Keasling JD, Katz L. 2016. Insights into polyketide biosynthesis gained from repurposing antibiotic-producing polyketide synthases to produce fuels and chemicals. *J. Antibiot.* In press. doi:10.1038/ja.2016.64

161. Ziemert N, Lechner A, Wietz M, Millán-Aguirre N, Chavarria KL, Jensen PR. 2014. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *PNAS* 111(12):e1130–39
162. Zifcakova L, Vetrovsky T, Howe A, Baldrian P. 2015. Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environ. Microbiol.* 18(1):288–301
163. Zimmermann W. 1990. Degradation of lignin by bacteria. *J. Biotechnol.* 13:119–30

## **Chapter 2: Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose**

**Previously published:** Adam J. Book\*, Gina R. Lewin\*, Bradon R. McDonald, Taichi E. Takasuka, Drew T. Doering, Aaron S. Adams, Joshua A.V. Blodgett, Jon Clardy, Kenneth F. Raffa, Brain G. Fox, and Cameron R. Currie; \*authors contributed equally (2014). Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. *Applied and Environmental Microbiology*, 80(15):4692-4701.

**Co-authors contributed in the following ways:** Conceived and designed the experiments: AJB, GRL, TET, BGF, CRC. Performed experiments and analyzed data for Figure 2.1: AJB, GRL, DTD; Figure 2.2: GRL, TET, AJB; Figure 2.3: AJB, BRM; Figure 2.4: GRL; AJB. Figure 2.5: AJB, GRL; Figure 2.6: AJB; Table 2.1: AJB; Table 2.2: AJB. Contributed reagents or materials: ASA, JAVB, JC, KFR. Wrote the paper: AJB, GRL, BGF, CRC.

### **2.1 Abstract**

Actinobacteria in the genus *Streptomyces* are critical players in microbial communities that decompose complex carbohydrates in the soil, and these bacteria have recently been implicated in the deconstruction of plant polysaccharides for some herbivorous insects. Despite the importance of *Streptomyces* to carbon cycling, the extent of their plant biomass-degrading ability remains largely unknown. In this study, we compared four strains of *Streptomyces* isolated from insect herbivores that attack pine trees: DpondAA-B6 (SDPB6) from the mountain pine beetle, SPB74 from the southern pine beetle, and SirexAA-E (SACTE) and SirexAA-G from the woodwasp, *Sirex noctilio*. Biochemical analysis of secreted enzymes demonstrated that only two of these strains, SACTE and SDPB6, were efficient at degrading plant biomass. Genomic

analyses indicated that SACTE and SDPB6 are closely related and that they share similar compositions of carbohydrate-active enzymes. Genome-wide proteomic and transcriptomic analyses revealed that the major exocellulases (GH6 and GH48), lytic polysaccharide monooxygenases (AA10), and mannanases (GH5) were conserved and secreted by both organisms, while the secreted endocellulases (GH5 and GH9 versus GH9 and GH12) were from diverged enzyme families. Together, these data identify two phylogenetically related insect-associated *Streptomyces* strains with high biomass-degrading activity and characterize key enzymatic similarities and differences used by these organisms to deconstruct plant biomass.

## 2.2 Introduction

Plant biomass is the most abundant terrestrial carbon source, and its deconstruction and subsequent catabolism are key components of global carbon cycling (1, 2). The majority of the energy in plant biomass is stored in plant cell walls, primarily in the recalcitrant polysaccharides cellulose and hemicellulose. Efficient breakdown of these insoluble polymers is difficult, and only a limited number of bacteria and fungi have this capability (3). Microbial decomposition in the soil is the primary pathway for carbon stored in terrestrial plant biomass to be returned to the atmosphere (1). Symbiotic cellulolytic microbes associated with a wide variety of herbivores also make significant contributions to global carbon flux (4). However, our molecular and biochemical understanding of how microbes deconstruct plant biomass is largely limited to a few model cellulolytic organisms (3, 5–9).

Actinobacteria in the genus *Streptomyces* are well represented in both culture-based and culture-independent studies of soil biodiversity (10). These bacteria are generally thought to make

important contributions to the carbon cycle through the decomposition of plant biomass and other biopolymers. However, individual cellulose-degrading enzymes have been isolated and biochemically tested for only a few strains (11–15). Genome-wide transcriptomic analyses have been completed for *Streptomyces coelicolor* and *S. griseus*, but not for these organisms grown on plant biomass as the carbon source (16–18). Very few *Streptomyces* strains have been shown to efficiently degrade crystalline cellulose (11), and fewer still have been analyzed at the molecular, biochemical, or system-wide level to identify key biomass-degrading enzymes.

*Streptomyces* organisms associated with a variety of insects that feed on living trees and dead wood have been demonstrated to deconstruct plant biomass and pure cellulose with high specific activity (19–22). Two distinct digestive models have emerged for insect-associated *Streptomyces*. In the internal digestive model, *Streptomyces* bacteria in the insect gut secrete enzymes to aid in digestion of recalcitrant plant biomass (21, 22). In the external digestive system model, *Streptomyces*, likely in combination with other bacteria and fungi, is inoculated directly into living pine trees during oviposition; bark beetles and woodwasps are insects that use this model (20, 23–25). The external microbial community inoculated into the pine tree by the insect secretes biomass-degrading enzymes that deconstruct plant cell walls and thus facilitate energy accumulation by the developing larvae inside the tree (26). *Streptomyces* sp. strain SirexAA-E (SACTE) is a cellulolytic strain that was isolated from the pine-boring woodwasp *Sirex noctilio* and fits into the external digestive model of insect-*Streptomyces* interactions (20). SACTE produces and secretes a suite of enzymes that efficiently deconstruct lignocellulosic plant biomass into small polysaccharides and free sugars (19). Genome-wide transcriptomic and proteomic analyses showed that coordinated expression and secretion of select enzymes were

tailored to the carbon sources present in the growth medium (19). In contrast to previous microarray-based transcriptomic analysis of SACTE, in this study, we used RNA-sequencing techniques and data analysis to compare gene expression between multiple strains of *Streptomyces*.

In this study, we characterized four strains of *Streptomyces* isolated from insects which, in combination with their host's symbiotic microbial community, feed on pine trees. These four strains come from insects that span two taxonomic orders and have a broad geographic distribution: the mountain pine beetle (*Dendroctonus ponderosae*), the southern pine beetle (*Dendroctonus frontalis*), and the *Sirex* woodwasp (*Sirex noctilio*) (Figure 2.1a; see also Figure A2.1 in the supplemental material). Isolation of three of these *Streptomyces* strains has been previously reported: SirexAA-E (SACTE) and SirexAA-G (SACTG) from the *Sirex* woodwasp (20) and SPB74 from the southern pine beetle (23). A fourth *Streptomyces* strain, DpondAA-B6 (SDPB6), was isolated from an adult mountain pine beetle collected from a naturally attacked *Pinus contorta* tree near Mackenzie, British Columbia, Canada (Figure 2.1a). Comparative genomics provided insight into the phylogenetic relationships of the *Streptomyces* strains from these different insect hosts and their genomic potential to degrade plant polymers. Microbiology, biochemistry, and next-generation sequencing were used to characterize the molecular responses of these organisms during growth on biomass and plant polysaccharides. Genes that were differentially expressed in response to growth on plant biomass versus glucose as a sole carbon source in the medium were identified by transcriptome sequencing (RNA-Seq), and proteins that were secreted during growth were identified using mass spectrometry. Together, these results provide new insight into how highly cellulolytic *Streptomyces* strains contribute to

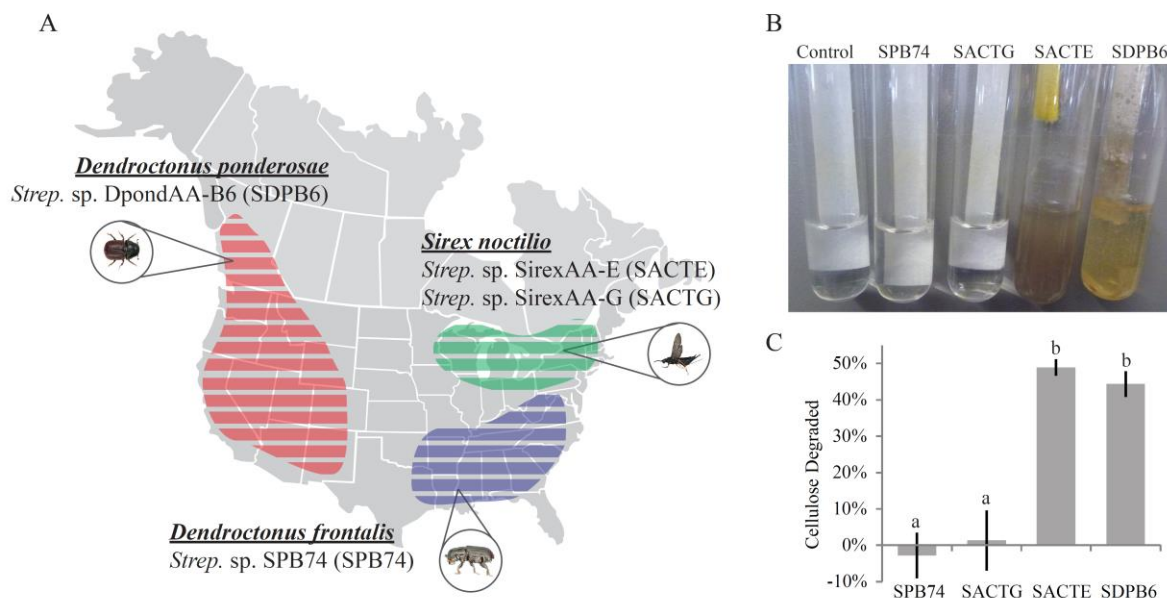


**Figure 2.1.** (A) General range of pine-boring insects *D. ponderosae*, *S. noctilio*, and *D. frontalis* in North America based on survey maps in the NAPIS database. Collection sites for individual insects and their associated *Streptomyces* strains are indicated. (B) *Streptomyces* organisms from these pine-boring insects were grown in minimal medium with cellulose filter paper as the sole carbon source. (C) Quantification of cellulose filter paper deconstruction in cultures grown for 10 days. Data represent the means  $\pm$  SDs from at least three experiments; lowercase letters represent statistically significant differences between strains (Tukey-Kramer method;  $P < 0.01$ ).

deconstruction of plant biomass in insect-herbivore systems.

## 2.3 Materials and Methods

### 2.3.1 Isolation of *Streptomyces*.

The isolation of *Streptomyces* strains SirexAA-E (SACTE; NCBI Taxonomy identification number [ID] 862751), SirexAA-G (SACTG; NCBI Taxonomy ID 683219) (20), and SPB74 (NCBI Taxonomy ID 465543) (23) was previously reported. *Streptomyces sp. DpondAA-B6* (SDPB6; NCBI Taxonomy ID 682311) was isolated from samples of *D. ponderosae* adults



collected from naturally attacked *P. contorta* trees near Mackenzie, British Columbia, Canada (27). Adult insects were surface sterilized in 95% ethanol for 1 min, rinsed twice in sterile phosphate-buffered solution (PBS), and homogenized in PBS, and dilutions were spread onto agar plates supplemented with chitin. Bacteria were isolated and maintained on yeast malt extract agar (YMEA) plates (Difco).

### **2.3.2 Culturing of organisms.**

Strains were grown in M63 minimal medium at pH 6.8 containing, per 1 liter, 10.72 g of  $\text{K}_2\text{HPO}_4$ , 5.24 g of  $\text{KH}_2\text{PO}_4$ , 2 g of  $(\text{NH}_4)_2\text{SO}_4$ , 0.5 ml of an iron solution (1 mg of  $\text{FeSO}_4$  dissolved in 1 ml of 0.01 M HCl), 1 ml of 1 M  $\text{MgSO}_4$ , 1 ml of thiamine (1 mg/ml), and 5 ml of trace element solution (SPV-4 with  $\text{Fe}^{\text{III}}$ -citrate substituted for  $\text{Fe}^{\text{II}}$ -chloride [28]). The sole carbon source (0.5%, wt/vol) in the medium was glucose, crystalline cellulose, Whatman no. 1 filter paper (GE Healthcare Life Sciences, Pittsburgh, PA), Sigmacell-20 (Sigma-Aldrich, St. Louis, MO), birch wood xylan (Sigma-Aldrich), or sterile AFEX-pretreated corn stover (GLBRC Corn Stover UW 2009 AFEX, AFEX ID 712-192). Cultures were incubated with shaking at 250 rpm at 30°C for the number of days indicated below.

### **2.3.3 Quantification of residual cellulose.**

Strains were grown in test tubes containing 8 ml of M63 medium with ~0.07 g of Whatman no. 1 filter paper (GE Healthcare Life Sciences, Pittsburgh, PA) as the sole carbon source. After 10 days of growth, residual cellulose concentrations were determined using an acid-detergent fiber method modified from the work of Weimer et al. (29). Briefly, 16 ml of acid detergent (30) was added to each culture. Tubes were crimp sealed with flanged rubber stoppers and autoclaved for

45 min at 121°C. One at a time, the tubes were vented, decapped, and immediately vacuum filtered through preweighed glass microfiber filters (Whatman GF/D, 2.7- $\mu$ m pore size [GE Healthcare Life Sciences, Pittsburgh, PA]). Then, filters were washed liberally with hot water. Filters were dried at 105°C overnight and reweighed immediately after removal from the oven. Net cellulose loss was calculated by subtracting the final residual cellulose mass from initial cellulose mass. This value was then divided by the initial cellulose mass to calculate percent cellulose degraded and corrected to account for cellulose and water loss in uninoculated samples.

#### **2.3.4 Genomic sequencing and phylogenetic analysis.**

Genome sequences were previously described for SACTE (Gold ID Gc01941) and SACTG (Gold ID Gi06834) (19, 20). A high-quality draft genome sequence for *Streptomyces* sp. DpondAA-B6 was determined at the Joint Genome Institute (JGI) using Illumina HiSeq2000 (Gold ID Gi0022546). All general aspects of library construction and sequencing performed at the Joint Genome Institute can be found at <http://www.jgi.doe.gov/>. The genome of SPB74 was generated at the Broad Institute (Gold ID Gi02993). Genome annotations were generated at JGI and the Broad Institute. Putative carbohydrate-active enzyme (CAZy) gene annotations were generated (19). Putative antiSMASH annotations were generated by uploading genomes to the antiSMASH online annotation server (31).

For multilocus phylogenetic analysis, proteins from all genomes included in the phylogeny were predicted using Prodigal and annotated using HMMer 3.0 models provided in TIGRfam 13.0, curated using the provided score cutoffs (32–34). A total of 646 TIGRfam families were found to be present in all 17 genomes (see Dataset A2.1). The protein sequences of each family from each

genome were aligned using the MAFFT 7.029 FFT-NS-i algorithm (35). If more than one protein from a single family was found in a genome, the protein with the highest bit score, i.e., best match to the respective HMMer model, was chosen as the representative ortholog. The protein sequences were then reverse translated to their respective DNA sequence, and all alignments were concatenated. The phylogeny was generated using RAxML 7.2.6 with the GTRGAMMA substitution model and 100 rapid bootstrap replicates (36).

### ***2.3.5 Preparation of secretomes.***

Supernatants were prepared from growing cultures by centrifuging the culture medium for 15 min at  $5,125 \times g$  to remove insoluble polysaccharides and cells. The supernatant fraction was then passed through a 0.22- $\mu$ m polyethersulfone (PES) membrane filter (GE Healthcare Life Sciences, Pittsburgh, PA) to remove remaining cells. For enzymatic assays, the supernatants were concentrated using a 10-kDa-cutoff PES membrane (Sartorius Stedim, NY). The concentration of secreted protein was determined by Bio-Rad protein assay.

### ***2.3.6 Enzyme activity measurements.***

Reducing-sugar assays were carried out by mixing concentrated (1 mg/ml) *Streptomyces* secretome preparations with polysaccharide-containing substrates. Control secretomes were prepared with Spezyme CP (Genencor), an industrial cellulolytic enzyme cocktail.

Polysaccharide substrates include Whatman no. 1 filter paper (GE Healthcare Life Sciences, Pittsburgh, PA), Sigmacell-20 (Sigma-Aldrich, St. Louis, MO), cellulose III (a generous gift from Bruce Dale at Michigan State University, East Lansing, MI [37]), phosphoric acid-swollen

cellulose (PASC) (**38**), birch wood xylan (Sigma-Aldrich), D-mannan (Megazyme, Ireland), corn stover (GLBRC 2009 Corn Stover), AFEX-pretreated corn stover (GLBRC Corn Stover UW 2009 AFEX, AFEX ID 712-192), and ionic-liquid pretreated switchgrass (a generous gift from Department of Energy Joint Bioenergy Institute, CA). Enzyme hydrolysis reactions were carried out using 0.1 mg/ml of secretome protein in 0.1 M sodium phosphate, pH 6, containing the different substrates (10 mg/ml) at 40°C for 20 h. Reducing-sugar content was determined by the dinitrosalicylic acid (DNS) assay (39). Glucose was used as the standard to quantify the amount of reducing ends produced.

### ***2.3.7 Extracellular proteomics.***

Extracellular proteins from culture supernatants were precipitated by trichloroacetic acid (TCA) precipitation of 100 ng of total secreted protein from 7-day-old culture supernatants. Protein samples were digested with trypsin (sequencing grade; Promega, Madison, WI) and desalted using C<sub>18</sub> pipette tips (Millipore, Billerica, MA). High-energy collision dissociation mass spectrometry employing capillary liquid chromatography-tandem mass spectrometry (LC-MS/MS) was performed on an electrospray ionization FT/ion-trap mass spectrometer (LTQ Orbitrap XL; Thermo Fisher Scientific, San Jose, CA) in the University of Wisconsin—Madison Biotechnology Center (Madison, WI) and database searched with Scaffold (Scaffold\_3\_00\_06; Proteome Software, Portland, OR). The spectra from SACTE were searched against the SACTE predicted proteome (translation of all protein coding sequences); a protein false-discovery rate of 0.0% (probabilistic method) was reported. The spectra from SDPB6 were searched against the SDPB6 predicted proteome; a protein false-discovery rate of 0.0% (probabilistic method) was reported. Homology between SACTE and SDPB6 proteins was generated by BLAST

comparison. Homology was reported when E values were less than  $1e^{-50}$  for more than 75% of the protein sequence.

### ***2.3.8 RNA extraction and preparation***

SDPB6 and SACTE were grown in M63 minimal medium plus either glucose or AFEX corn stover (0.5%, wt/vol) for 7 days with shaking at 30°C. The cell pellet was separated from the culture supernatant by centrifugation for 10 min at  $4,500 \times g$ . Total RNA extraction was adapted as previously described (19, 40). Briefly, 700  $\mu$ l of cells in 50 mM sodium acetate, pH 5.1, containing 10 mM EDTA was combined with 100  $\mu$ l of 10% (wt/vol) SDS, 700  $\mu$ l of saturated phenol (pH 4.3; Ameresco), 100  $\mu$ l of glass beads (400  $\mu$ m), and 100  $\mu$ l of zirconia-silica beads (1.0 mm). Cells were shaken in a Mini-Beadbeater-96 for 2 min to lyse the cells. Samples were heat shocked at 65°C for 10 min, followed by 2 more min in the beadbeater. Then, samples were centrifuged at  $12,000 \times g$  for 5 min, and the aqueous layer was transferred to a new tube. Samples were extracted once with phenol (pH 4.5), twice with phenol-chloroform (1:1, vol/vol), and once with chloroform. RNA was precipitated from solution by adding 0.2 volume of 8 M LiCl to the tube and incubating the tube overnight at  $-80^{\circ}\text{C}$ . RNA pellets were collected by centrifugation at  $12,000 \times g$  for 20 min and washed with 70% ethanol. Pellets were dried and resuspended in 40  $\mu$ l of RNase-free water. DNA contamination was removed from samples by incubation with 5  $\mu$ l of RQ1 RNase-free DNase (Promega, Madison, WI) for 60 min and then extracted once with phenol (pH 4.5) and twice with phenol-chloroform (1:1, vol/vol), followed by chloroform extraction. RNA was precipitated by addition of 0.1 volume of 3 M sodium acetate and 2 volumes of 100% ethanol and incubation overnight at  $-80^{\circ}\text{C}$ . Samples were then centrifuged at  $12,000 \times g$  for 20 min, washed twice with 70% ethanol, and air dried. Samples

were resuspended in RNase-free water and checked for quality and quantity with a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA). High-quality RNA preparations had a total RNA yield greater than 5 µg and an  $A_{260}/A_{280}$  ratio greater than 1.7. RNA stability was also checked by Agilent RNA PicoChip analysis. Finally, bacterial rRNA was removed from 5 µg of total RNA with the Ribo-Zero magnetic kit for bacteria (Epicentre, Madison, WI).

### ***2.3.9 RNA sequencing and analysis.***

cDNA library construction and Illumina HiSeq2000 sequencing were performed at the University of Wisconsin Biotechnology Center (Madison, WI). Single-ended 100-bp sequencing runs were performed, and data were demultiplexed using the CASAVA software package. Our analysis pipeline began with trimming the fastq sequences to remove low-quality bases, adapter sequences, and reads with fewer than 60 bases (41). Trimmed fastq files were then aligned to their respective genomes using the Burrows Wheeler Aligner BWA-mem (42). Data were then converted to BAM files using Samtools, and alignment statistics were calculated with Bamtools (43). Reads with overlapping genes were then counted to quantify expression level using the R package Rsamtools. Statistical analysis was performed with the R package DEseq to identify differentially expressed genes and generate heatmaps of expression data. Scatter plots were generated with the R plot function (44). Homology networks were created by BLAST comparison of all CAZy annotated genes in SACTE and SDPB6. Proteins with a BLAST E value of  $<1e^{-50}$  were connected with an edge, as were CAZy family annotations. A network was generated with Cytoscape (45).

### **2.3.10 Nucleotide sequence accession numbers.**

Genomic sequences are available at NCBI RefSeq as follows: SACTE, PRJNA72627; SACTG, PRJNA61899; SDPB6, PRJNA195846; and SPB74, PRJNA48415.

## **2.4 Results**

### **2.4.1 Cellulolytic activity screen.**

Three strains of *Streptomyces* isolated from pine tree-associated insects have been previously reported: SirexAA-E (SACTE) and SirexAA-G (SACTG) from the *Sirex* woodwasp (20) and SPB74 from the southern pine beetle (23). A fourth *Streptomyces* strain, DpondAA-B6 (SDPB6), was isolated from a surface-sterilized adult mountain pine beetle collected from a naturally attacked *Pinus contorta* tree near Mackenzie, British Columbia, Canada (Figure 2.1a) (27). Cellulolytic activities were significantly different among these four strains (Figure 2.1). Only two of the four strains, SACTE and SDPB6, grew well in liquid medium containing filter paper as the only carbon source (Figure 2.1b and 2.1c). Quantitative cellulose utilization assays were performed, and these two strains deconstructed  $48.9\% \pm 2.3\%$  and  $44.3\% \pm 3.6\%$  of crystalline cellulose in 10 days, respectively (Figure 2.1c). The other two strains, SPB74 and SACTG, showed little to no growth in the qualitative assay, and essentially no filter paper deconstruction was determined by quantitative analysis.

### **2.4.2 Biochemical analysis of secreted enzymes.**

Polysaccharide-hydrolyzing activities of the secreted proteins from these four strains supported significant differences between the two cellulolytic strains, SACTE and SDPB6, and the two

noncellulolytic strains, SPB74 and SACTG. While all strains grew visibly on ammonia fiber expansion-pretreated corn stover (AFEX-CS [37]) as the sole carbon source, there were significant differences in the amount of total protein secreted by each strain. SACTE gave the highest concentration of total secreted protein ( $167 \pm 5 \mu\text{g/ml}$ ), followed by SPB74 ( $75 \pm 11 \mu\text{g/ml}$ ), then SDPB6 ( $69 \pm 12 \mu\text{g/ml}$ ), and SACTG ( $43 \pm 9 \mu\text{g/ml}$ ). When normalized to total protein concentration, the secretomes from SACTE and SDPB6 had significantly higher specific activity for hydrolysis than SACTG and SPB74 in reactions with cellulosic substrates (filter paper, Sigmacell, PASC, and lichenan), hemicellulosic substrates (xylan and mannan), and plant biomass (AFEX-CS) ( $P < 0.01$  for all; Tukey-Kramer method) (Figure 2.2). The secretomes from SDPB6 had significantly higher activity than those of SACTE when reacted against crystalline cellulose (24%;  $P = 0.0012$ ) and cellulose III (83%;  $P = 0.0168$ ). In contrast, secretomes from SACTE were significantly more active than those of SDPB6 when reacted against birch wood xylan ( $P < 0.0001$ ), lichenan ( $P = 0.0409$ ), and AFEX-CS ( $P = 0.0160$ ). High-performance liquid chromatography (HPLC) analysis of the products produced by the reaction of both SACTE and SDPB6 secretomes with crystalline cellulose showed that cellobiose was the primary product.

#### ***2.4.3 Phylogenetic and genomic analyses.***

Given the significant differences in biomass deconstruction activities of the four strains, we compared their genomes to gain insight into genetic differences that contribute to the phenotypes. A finished genome sequence was previously described for SACTE, and a draft sequence was described for SACTG (19, 20). Draft genomes for SPB74 and SDPB6 were generated at the Broad Institute and Joint Genome Institute, respectively (Table 1). Multilocus



phylogenetic analysis of these four insect-associated *Streptomyces* strains, in addition to 13 strains with publicly available genome sequences, demonstrated that SDPB6 and SACTE are closely related species that form a monophyletic clade, as are SPB74 and SACTG. However, these two clades occur in separate lineages of *Streptomyces* (Figure 2.3).

Genome analyses for the presence of carbohydrate-active enzymes (CAZy) revealed that SACTE had the highest percentage of CAZy genes in its genome, followed by SACTG, then SDPB6, and finally SPB74 (Table 2.1; see also Dataset A2.2). Genomic differences between the cellulolytic and noncellulolytic strains

included specific families of glycoside hydrolases (GHs) (Table 2.2; see also Dataset A2.2). Notably, SPB74 and SACTG were missing genes that encode the important cellulose-degrading families GH5, GH9, GH74, and GH48 as well as the primary xylanase families GH10 and GH11. Furthermore, SPB74 and SACTG had significantly fewer carbohydrate binding modules (CBMs) in their genomes. The CBM2 family was expanded in the SACTE and SDPB6 genomes,

**Table 2.1.** Genome Statistics

|                                 | SPB74   | SACTG   | SACTE   | SDPB6   |
|---------------------------------|---------|---------|---------|---------|
| Genome size (bp)                | 6970553 | 7443083 | 7414440 | 7295516 |
| Scaffolds                       | 2       | 254     | 1       | 36      |
| G+C%                            | 68.13%  | 73.32%  | 71.75%  | 72.14%  |
| Predicted ORFs                  | 5808    | 6612    | 6647    | 6362    |
| Protein coding genes            | 5742    | 6547    | 6563    | 6272    |
| RNA genes                       | 66      | 65      | 84      | 90      |
| CAZy Proteins <sup>a</sup>      | 130     | 166     | 185     | 141     |
| Percent CAZy                    | 2.3%    | 2.5%    | 2.9%    | 2.3%    |
| antiSMASH Clusters <sup>b</sup> | 16      | 20      | 22      | 28      |

<sup>a</sup> Carbohydrate Active enZymes

<sup>b</sup> putative secondary metabolite gene clusters

**Table 2.2.** Genomic composition of putative biomass-degrading enzymes and accessory proteins

|       | SPB74 | SACTG | SACTE | SDPB6 | Predicted Function <sup>a</sup> |
|-------|-------|-------|-------|-------|---------------------------------|
| GH5   | 0     | 0     | 3     | 2     | endo-glucanase                  |
| GH8   | 0     | 0     | 0     | 0     |                                 |
| GH9   | 0     | 0     | 1     | 1     |                                 |
| GH12  | 1     | 1     | 1     | 1     |                                 |
| GH74  | 0     | 0     | 1     | 2     | cellobio-hydrolase              |
| GH6   | 2     | 3     | 1     | 2     |                                 |
| GH48  | 0     | 0     | 1     | 1     | beta-glucosidase                |
| GH1   | 3     | 3     | 6     | 6     |                                 |
| GH3   | 7     | 8     | 7     | 5     | LPMO                            |
| AA10  | 1     | 1     | 6     | 4     |                                 |
| GH10  | 0     | 0     | 1     | 1     | xylanase                        |
| GH11  | 0     | 0     | 1     | 2     |                                 |
| GH13  | 9     | 9     | 12    | 11    | chitinase                       |
| GH18  | 4     | 4     | 7     | 7     |                                 |
| GH19  | 1     | 1     | 3     | 2     | cellulose binding               |
| CBM1  | 0     | 0     | 0     | 0     |                                 |
| CBM2  | 2     | 2     | 10    | 10    |                                 |
| CBM3  | 1     | 1     | 0     | 0     |                                 |
| CBM4  | 0     | 0     | 1     | 1     | chitin binding                  |
| CBM5  | 1     | 1     | 3     | 6     |                                 |
| CBM12 | 0     | 0     | 2     | 7     |                                 |

<sup>a</sup> Predicted function based on homologs in the CAZy database

with 10 copies per genome, while the genomes of SPB74 and SACTG contain only two CBM2 domains each. Interestingly, the two cellulolytic strains, SACTE and SDPB6, also had the most secondary metabolite gene clusters, as predicted through antiSMASH genome-based analyses (31). Given the dissimilarities observed in these strain's phenotypic, biochemical, and genomic analyses, we chose to focus further molecular analyses on the two cellulolytic strains, SACTE and SDPB6, to gain insights into how these organisms successfully degrade plant biomass.

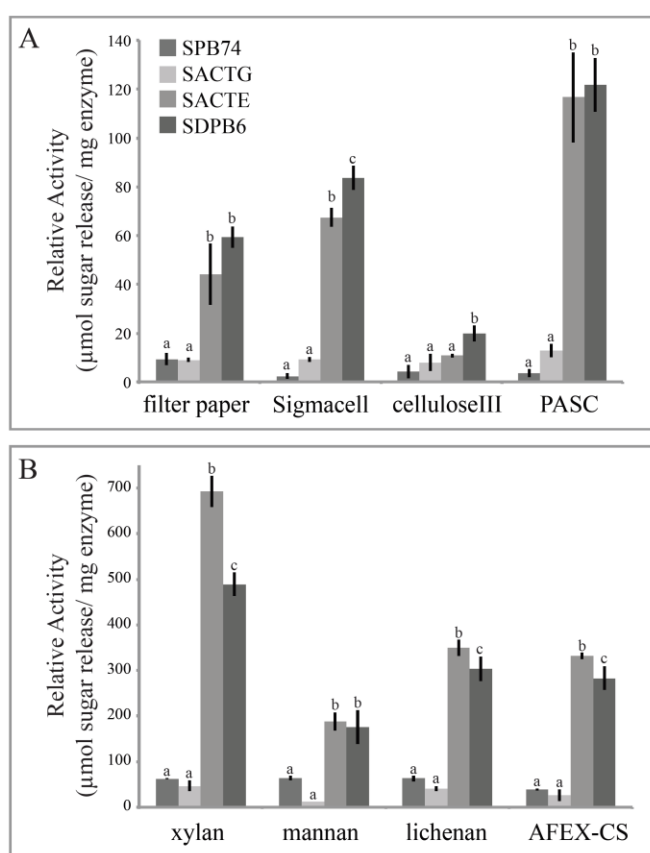


**Figure 2.2. Reactions of secretomes from 4 *Streptomyces* strains with polysaccharide substrates. (A)**

Reducing sugars released from the reaction of the secretomes of 4 *Streptomyces* strains with cellulose filter paper, crystalline cellulose (Sigmacell-20), cellulose III, and amorphous cellulose (PASC). (B) Reducing sugars released from reaction with hemicellulosic polymers xylan, mannan, lichenan, and AFEX-pretreated corn stover. Data represent the means  $\pm$  SDs from three experiments; lowercase letters represent statistically significant differences within each substrate (Tukey-Kramer method;  $P < 0.05$ ).

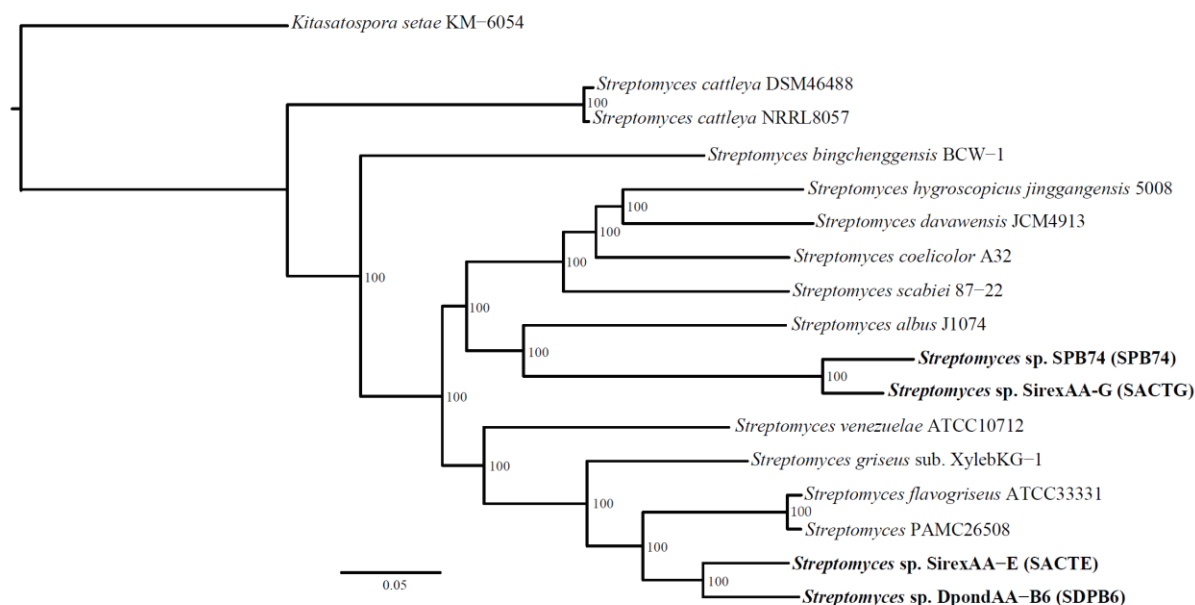


**Figure 2.3. Multilocus phylogenetic tree of the genus *Streptomyces*.** The maximum-likelihood tree was generated from a concatenated alignment of 646 orthologous genes conserved in all genomes of 16 strains of *Streptomyces* plus the closely related outgroup *Kitasatospora setae*. Bootstrap support for the maximum-likelihood analysis is given at each node.

#### 2.4.4 Secretome analysis on pure polysaccharides.

We used LC-MS/MS to identify proteins in the secretomes from SACTE and SDPB6 cultures grown in minimal medium containing cellulose, xylan, or glucose as the sole carbon source (see Data Sets S3 and S4 in the supplemental material). Relatively simple secretomes were found for both strains during growth on cellulose; 23 and 30 proteins constitute 95% of the total identified mass spectral counts from SACTE and SDPB6, respectively. In contrast, both SACTE and SDPB6 secreted a more diverse set of proteins in response to growth on either glucose (122 and 162 proteins, respectively) or xylan (92 and 125 proteins, respectively). Other than a constitutively secreted putative glucosylceramidase (SDPB6\_04637) in SDPB6, the dominant proteins observed in the secretomes from xylan- and cellulose-grown cultures were absent in the

secretomes of the glucose-grown cultures. In xylan-grown cultures, putative xylan-degrading enzymes were highly abundant: GH10 xylanases (SACTE\_0265 and SDPB6\_04909), putative xylose isomerase proteins (SACTE\_5230, SDPB6\_02230, and SDPB6\_05756), and a xylanase from the GH11 family in SACTE (SACTE\_0358) (see Data Sets S3 and S4).

In both the SACTE and SDPB6 cellulose secretomes, four homologous proteins represented 76% of the total spectral counts (Figure 2.4a) (the threshold for counting proteins was greater than 1% of total spectral counts; the homology threshold was a BLAST E value of  $<1e^{-50}$ ). All four proteins were annotated as CAZymes and include the nonredundant set of enzymes necessary to deconstruct crystalline cellulose to cellobiose and a beta-mannanase: reducing- and nonreducing-end 1,4-beta cellobiohydrolases (GH6 and GH48, respectively), a lytic polysaccharide monooxygenase (LPMO; AA10), and a beta-mannanase (GH5). When the threshold was relaxed to include proteins with more than 0.1% of the total spectral counts, the core secretome contained 14 homologous proteins that accounted for 87% of the total counts (Figure 2.4d).

One major difference between the SACTE and SDPB6 secretomes was the predominant type of enzyme produced for endocellulolytic hydrolysis of internal bonds in the cellulose chain (Figure 2.4b and 2.4c). SACTE secreted a GH5 enzyme (SACTE\_0482), while SDPB6 secreted two enzymes with predicted endoglucanase activity from the GH9 and GH12 families (SDPB6\_01939 and SDPB6\_04855, respectively). Furthermore, the percentage of the secretome representing endoglucanase activity was 10-fold greater in SDPB6 than in SACTE (23.1% and 2%, respectively). There were also significant differences in the detected abundance of putative cellulolytic LPMOs in each strain, 20% in SDPB6 and 7% in SACTE (SDPB6\_03538 and

SACTE\_3159, respectively). When added to the endo-1,4-beta-glucosidase activity, 43% of the total spectral counts in cellulose-grown SDPB6 were from enzymes that cleave cellulose internally, versus 9% in SACTE. Conversely, the cellulose-grown SACTE secretome was enriched in xylan-degrading enzymes (GH10 SACTE\_0265 and acetylxylin esterase SACTE\_0357). The SDPB6 secretome also contained proteins with putative 1,3-beta-glucosidase and peptidase functions and a protein with unknown function (SDPB6\_03112, SDPB6\_02147, and SDPB6\_02490, respectively). The absence of detectable spectra does not

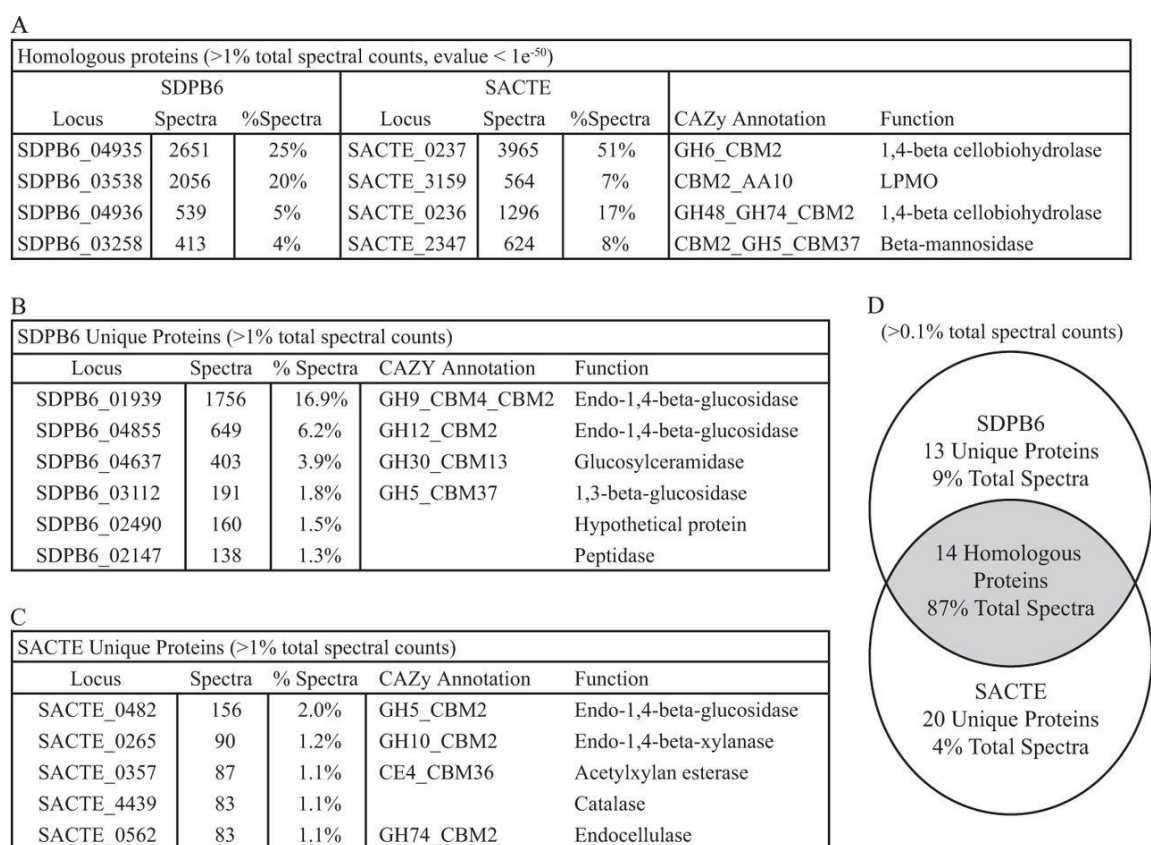
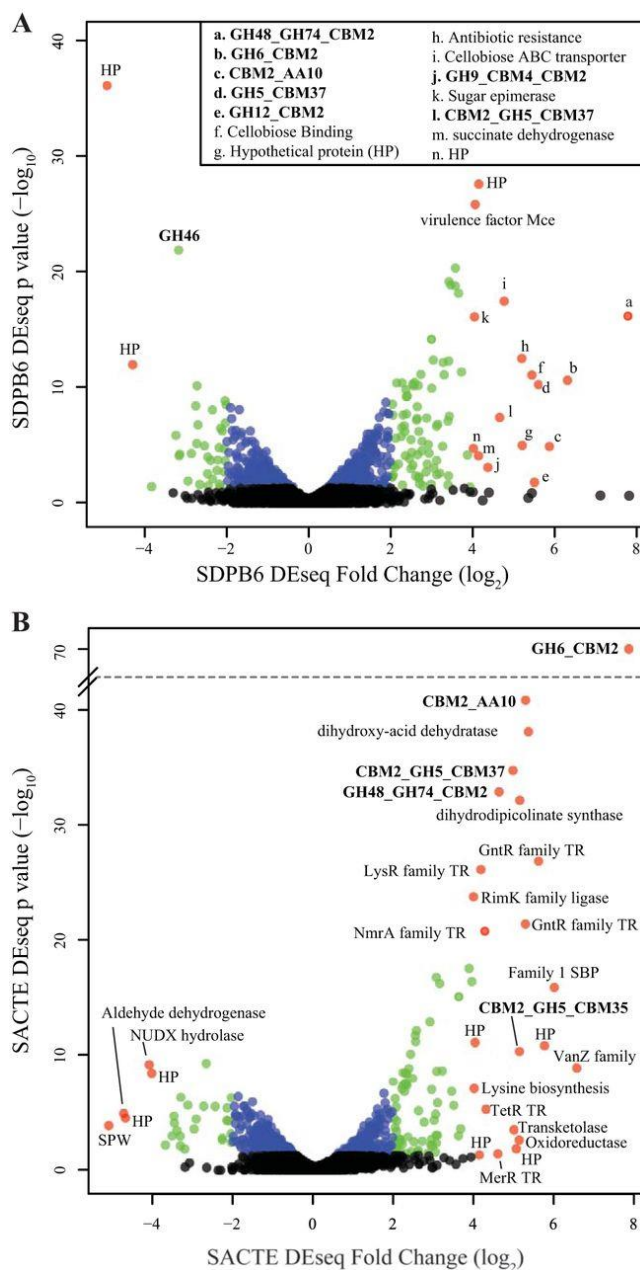


**Figure 2.4. Composition of secretomes from cellulose-grown cultures of SACTE and SDPB6 identified by LC-MS/MS.** (A) Homologous proteins identified in secretomes with more than 1% of the total spectral counts. (B and C) Proteins unique to secretomes from SDPB6 (B) and SACTE (C) with more than 1% of total spectral counts. (D) Homologous and unique proteins when threshold is lowered to proteins with more than 0.1% of the spectral counts.

always correlate with the absence of a protein in a sample. Thus, we investigated mRNA expression to gain further insight into these differences.

#### **2.4.5 RNA sequencing.**

To gain further insight into how SACTE and SDPB6 deconstruct and metabolize plant biomass, we analyzed the differential expression of mRNA in these strains when grown in medium containing either glucose or AFEX-CS as the sole carbon source. In total, approximately 800 Mbp and 1 Gbp were sequenced for SACTE and SDPB6, respectively. When these reads were mapped to their respective genomes, gene coverage was approximately 500 to 2,000 reads per open reading frame (ORF) (see Table A2.1). Variation between biological replicates was measured by pairwise comparison of normalized reads for all protein-coding genes (see Figure A2.2). Linear models (slope and  $r^2$  value) of the pairwise data demonstrate the actual fit of the data versus the theoretical perfect fit. SACTE glucose had the closest fit to the theoretical, followed by SDPB6 AFEX-CS, then SACTE AFEX-CS, and then SDPB6 glucose. Sample clustering of variance-stabilized data reflected our experimental design (i.e., glucose samples cluster together and AFEX-CS samples cluster together [see Figure A2.3]). Together, these data quality assessments indicate that our experiment produced biologically relevant data.



#### 2.4.6 Differential expression analysis.

Comparison of transcript abundance between growth in media containing either glucose or AFEX-CS as the sole carbon source identified 684 and 840 genes that were differentially expressed for SACTE and SDPB6, respectively ( $P < 0.05$ ) (Figure 2.5; see also Supplementary Datasets 2.3 and 2.6). SDPB6 differentially expressed 84 genes with >2-fold changes and 15 with >4-fold changes (Figure 2.5, green and red circles, respectively; all fold change descriptions are  $\log_2$  fold change). Similarly, in SACTE, 90 genes were differentially expressed with >2-fold changes and 23 with >4-fold changes. Interestingly, 22% of genes in the >4-fold differential expression (DE) category were CAZy annotated genes (Figure 2.5, bold annotations). To confirm this observation, we categorized all genes from SDPB6 and SACTE into broad KEGG and CAZy categories; the CAZy functional category had the greatest number of >4-fold differentially expressed genes in both strains (see Dataset A2.7). ABC transporters were also significantly upregulated in both strains, as were a number of metabolic pathways (carbon, starch, and benzoate) and a number of nonpolar and basic amino acid biosynthetic pathways (Val, Ile, Leu, Lys, Pro, and Arg). Lastly, SACTE upregulated five transcription factors 4-fold or greater, while SDPB6 did not have any transcription factors with a 4-fold change.

To further examine the genes in the CAZy family that may contribute to the deconstruction of plant biomass, we built a homology network of all CAZy proteins from SDPB6 and SACTE (Figure 2.6). Nodes in this network represent proteins (circles) or CAZy families (triangles), and edges represent annotations or homologous proteins (BLAST E value less than  $1e^{-50}$ ). The size of protein nodes corresponds to the fold change observed in the RNA-Seq data. Only a small number of the total CAZy genes were significantly differentially expressed (>2-fold; SDPB6,



12/131, and SACTE, 17/166 [Figure 2.6, inset; see also Supplementary Datasets 2.5 and 2.6]). However, a majority of genes that contain a CBM2 domain were significantly induced (SDPB6, 7/10, and SACTE, 8/11). Similar to the proteomic results, the RNA-Seq data identified a small combination of genes that encode enzymes to degrade crystalline cellulose to cellobiose (endoglucanase [GH5, GH9, and GH12], reducing-end exoglucanase [GH48], nonreducing-end exoglucanase [GH6], and polysaccharide monooxygenase [AA10]). The most highly induced genes in these two strains were both genes for exoglucanases: in SDPB6, the reducing-end GH48

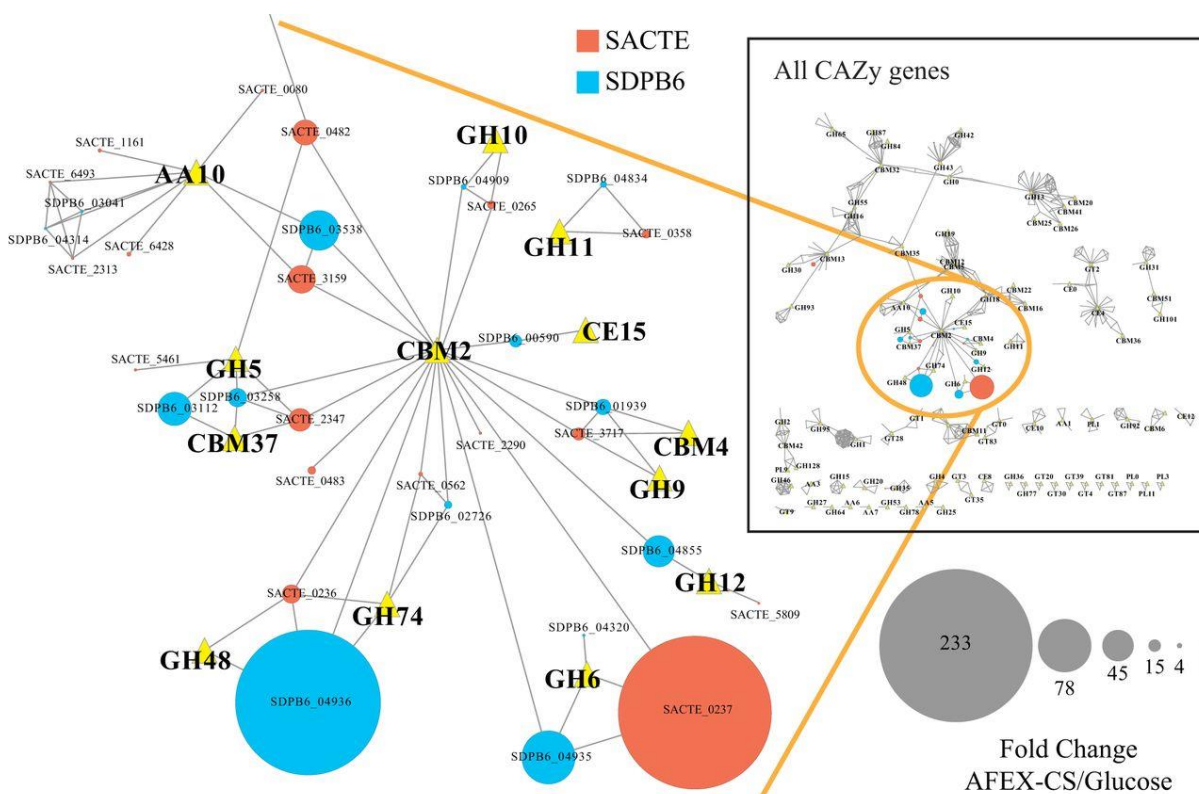


**Figure 2.6. Protein similarity network of CAZymes present in SDPB6 and SACTE.** Nodes are proteins (circles) or CAZy functional categories (yellow triangles); edges indicate that the gene belongs to the respective CAZy family or a BLAST similarity with an E value less than  $1e^{-50}$ . Node size represents the fold change in RNA abundance between AFEX-CS and glucose. The inset shows all CAZy proteins; the expanded region shows highly induced genes.

family enzyme SDPB6\_04936, and in SACTE, the nonreducing-end GH6 family enzyme SACTE\_0237. Both strains also induced the gene homologous to each of these exoglucanase genes, SACTE\_0236 and SDPB6\_04935, respectively. Interestingly, each pair of nonreducing- and reducing-end exocellulases was concurrent in their respective genomes, which suggests coregulated expression. These two *Streptomyces* strains also induced homologous AA10 LPMOs in response to growth on plant biomass (SDPB6\_03538 and SACTE\_3159).

While the two strains induced homologous GH9 endoglucanases (SDPB6\_01939 and SACTE\_3717), the most highly induced endoglucanases were from different GH families. SDPB6 expressed a GH12 enzyme (SDPB6\_04855), while SACTE expressed a GH5 enzyme (SACTE\_0482). The homolog of SDPB6\_04855 in SACTE (SACTE\_5809) does not contain a CBM2 domain and was not significantly expressed. Moreover, SDPB6 does not contain a homolog of the endocellulolytic SACTE GH5 enzyme (SACTE\_0482).

## 2.5 Discussion

*Streptomyces* strains are found in a wide variety of terrestrial and marine ecosystems. Classically, they have been viewed as free-living organisms, but recently they also have been found to be associated with eukaryotic hosts (20, 21,23, 26, 46). Despite the ubiquitous presence of *Streptomyces* in most ecosystems (10), their capability to degrade plant biomass remains largely uncharacterized. In this study, we compared four strains of *Streptomyces* isolated from pine-boring insect herbivores. Using a combination of biochemical and molecular tools, we have provided insight into the biomass-deconstructing capability of these *Streptomyces* and the similarities and differences in enzymes they use to break down plant biomass.

Phenotypic and biochemical data showed that two of the strains, SACTE and SDPB6, are highly efficient at degrading both plant biomass and pure polysaccharides, whereas SPB74 and SACTG have limited activity on all polysaccharide substrates tested (Figures 2.1 and 2.2). Our phylogenetic analysis showed that these two pairs of *Streptomyces* strains are in separate lineages of the genus. Analysis of the content of CAZy genes identified important genetic differences between the cellulolytic and noncellulolytic strains. While CAZy genes are abundant in the genomes of all four strains (2.3% to 2.9%), the SPB74 and SACTG genomes are missing genes encoding critical endoglucanase, cellobiohydrolase, and xylanase families and are also depleted in CBM2 domains. The absence of these enzymes and binding modules likely impacts the ability of these organisms to efficiently deconstruct the diverse sugar linkages found in plant cell walls. From an evolutionary perspective, it is interesting to consider whether the abundance of CAZy genes in SACTE and SDPB6 represents selection toward enhanced biomass-degrading activity or whether the absence of important CAZymes in SPB74 and SACTG genomes represents specialization for another purpose, such as secondary-metabolite production. SPB74 produces a novel secondary metabolite, mycangimycin, which inhibits growth of an antagonistic fungus of the beetle system (23, 47). These insect-associated strains may have coevolved with their hosts to play either a nutritional or protective role. Further microbiology and genomic analyses are needed to understand the evolution of these strains and their putative symbiotic roles in these insect-microbe systems.

Proteomic and transcriptomic analyses of SACTE and SDPB6 revealed homologous and relatively simple secreted enzyme cocktails used to deconstruct plant biomass. The major

exocellulases, LPMOs, and mannanases were conserved in both organisms. In contrast, the molecular data from SACTE and SDPB6 showed that these strains use endocellulases from different enzyme families (SACTE\_0482, GH5, and SDPB6\_04855, GH12). The genome of SDPB6 does not contain a homolog to SACTE\_0482. Conversely, while SACTE has a homologous GH12 enzyme in its genome, this gene does not contain a CBM2 (Figure 2.6). We hypothesize that together with their evolutionary relatedness, the endocellulolytic activities of the GH5 and GH12 enzymes have been selected by the insect-host and environment for slightly different functional activities. The SDPB6 secretome contains twice as many cellulolytic LPMOs as does that of SACTE based on mass spectral counts (Figure 2.4), suggesting differential reliance on oxidative cellulose deconstruction in the two strains. This variation in endocellulolytic and LPMO enzymes may contribute to the small but significant differences observed in the secretome activity assays for hydrolysis of crystalline cellulose (Figure 2.2).

All highly expressed enzymes in SACTE and SDPB6 shared conserved enzyme domain architectures; i.e., they all contained a CBM2 domain linked to a catalytic domain, even though the majority of CAZy genes in both genomes are not linked to a CBM2 domain. In contrast, the SPB74 and SACTG genomes contain only two CBM2 domains (GH6\_CBM2 and GH18\_CBM2). This conservation of enzyme domain architecture suggests that highly cellulolytic *Streptomyces* strains have converged on the use of secreted enzymes fused to an individual CBM2 domain. This pattern is similar to that for fungal secreted cellulases, which primarily consist of enzymes fused to CBM1 domains (48). It is possible that the similar, filamentous cellular morphology of fungi and *Streptomyces* may have selected for secretion of free enzymes with this similar domain architecture. Additionally, non-CAZy genes likely play a

role in plant biomass deconstruction. Genes classified as ABC transporter genes were significantly upregulated; the products of these genes may be involved in transport of solubilized sugars into the cell. Genes in the benzoate degradation pathway were also upregulated; these proteins may deconstruct aromatic rings in lignin (49).

This work identified two phylogenetically related strains of cellulolytic *Streptomyces* isolated from insect herbivores and established an overall similarity in the enzymatic apparatus they use for cellulose deconstruction. Further sampling from insect herbivore systems will help clarify how these *Streptomyces* and their relatively simple external aerobic enzyme complexes help insects efficiently unlock the vast amount of energy stored in the cell walls of plants. The simplicity of these cellulolytic enzyme complexes may also provide insights into how to modify the currently complex enzyme cocktails needed for the production of sustainable lignocellulosic biofuels.

## 2.6 References

1. Schlesinger WH, Andrews JA. 2000. Soil respiration and the global carbon cycle. *Biogeochemistry* 48:7–20. 10.1023/A:1006247623877
2. Klemm D, Heublein B, Fink HP, Bohn A. 2005. Cellulose: fascinating biopolymer and sustainable raw material. *Angew. Chem. Int. Ed. Engl.* 44:3358–3393. 10.1002/anie.200460587
3. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66:506–577. 10.1128/MMBR.66.3.506-577.2002

4. Singh BK, Bardgett RD, Smith P, Reay DS. 2010. Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nat. Rev. Microbiol.* 8:779–790.  
10.1038/nrmicro2439
5. Glass NL, Schmoll M, Cate JHD, Coradetti S. 2013. Plant cell wall deconstruction by ascomycete fungi. *Annu. Rev. Microbiol.* 67:477–498. 10.1146/annurev-micro-092611-150044
6. Desvaux M. 2005. *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol. Rev.* 29:741–764. 10.1016/j.femsre.2004.11.003
7. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MWW, Lunin VV, Himmel ME, Bomble YJ. 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* 342:1513–1516. 10.1126/science.1244273
8. Dam P, Kataeva I, Yang SJ, Zhou FF, Yin YB, Chou WC, Poole FL, Westpheling J, Hettich R, Giannone R, Lewis DL, Kelly R, Gilbert HJ, Henrissat B, Xu Y, Adams MWW. 2011. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Res.* 39:3240–3254. 10.1093/nar/gkq1281
9. Fontes CMGA, Gilbert HJ. 2010. Cellulosomes: highly efficient nanomachines designed to designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* 79:655–681. 10.1146/annurev-biochem-091208-085603
10. Janssen PH. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.* 72:1719–1728.  
10.1128/AEM.72.3.1719-1728.2006

11. Wachinger G, Bronnenmeier K, Staudenbauer WL, Schrempf H. 1989. Identification of mycelium-associated cellulase from *Streptomyces reticuli*. *Appl. Environ. Microbiol.* 55:2653–2657
12. Schlochtermeyer A, Walter S, Schroder J, Moorman M, Schrempf H. 1992. The gene encoding the cellulase (Avicelase) Cel1 from *Streptomyces reticuli* and analysis of protein domains. *Mol. Microbiol.* 6:3611–3621
13. Walter S, Schrempf H. 1996. Physiological studies of cellulase (Avicelase) synthesis in *Streptomyces reticuli*. *Appl. Environ. Microbiol.* 62:1065–1069
14. Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunaes AC, Stenstrom Y, MacKenzie A, Sorlie M, Horn SJ, Eijsink VGH. 2011. Cleavage of cellulose by a CBM33 protein. *Protein Sci.* 20:1479–1483. 10.1002/pro.689
15. Marushima K, Ohnishi Y, Horinouchi S. 2009. CebR as a master regulator for cellulose/celooligosaccharide catabolism affects morphological development in *Streptomyces griseus*. *J. Bacteriol.* 191:5930–5940. 10.1128/JB.00703-09
16. Yagüe P, Rodriguez-Garcia A, Lopez-Garcia MT, Martin JF, Rioseras B, Sanchez J, Manteca A. 2013. Transcriptomic analysis of *Streptomyces coelicolor* differentiation in solid sporulating cultures: first compartmentalized and second multinucleated mycelia have different and distinctive transcriptomes. *PLoS One.* 8(3):e60665. 10.1371/journal.pone.0060665
17. Rodríguez-García A, Barreiro C, Santos-Beneit F, Sola-Landa A, Martin JF. 2007. Genome-wide transcriptomic and proteomic analysis of the primary response to phosphate limitation in *Streptomyces coelicolor* M145 and in a  $\Delta phoP$  mutant. *Proteomics* 7:2410–2429. 10.1002/pmic.200600883

18. Hara H, Ohnishi Y, Horinouchi S. 2009. DNA microarray analysis of global gene regulation by A-factor in *Streptomyces griseus*. *Microbiology* 155:2197–2210.  
10.1099/mic.0.027862-0
19. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. 2013. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci. Rep.* 3:1030.  
10.1038/srep01030
20. Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. *ISME J.* 5:1323–1331. 10.1038/ismej.2011.14
21. Pasti MB, Belli ML. 1985. Cellulolytic activity of actinomycetes isolated from termites (Termitidae) gut. *FEMS Microbiol. Lett.* 26:107–112
22. Watanabe Y, Shinzato N, Fukatsu T. 2003. Isolation of actinomycetes from termites' guts. *Biosci. Biotechnol. Biochem.* 67:1797–1801. 10.1271/bbb.67.1797
23. Scott JJ, Oh DC, Yuceer MC, Klepzig KD, Clardy J, Currie CR. 2008. Bacterial protection of beetle-fungus mutualism. *Science* 322:63. 10.1126/science.1160423
24. Slippers B, Coutinho TA, Wingfield BD, Wingfield MJ. 2003. A review of the genus *Amylostereum* and its association with woodwasps. *S. Afr J. Sci.* 99:70–74
25. Vasanthakumar A, Handelsman J, Schloss PD, Bauer LS, Raffa KF. 2008. Gut microbiota of an invasive subcortical beetle, *Agrilus planipennis* Fairmaire, across various life stages. *Environ. Entomol.* 37:1344–1353.
26. Seipke RF, Kaltenpoth M, Hutchings MI. 2012. *Streptomyces* as symbionts: an emerging and widespread theme? *FEMS Microbiol. Rev.* 36:862–876. 10.1111/j.1574-6976.2011.00313.x



27. Adams AS, Aylward FO, Adams SM, Erbilgin N, Aukema BH, Currie CR, Suen G, 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.* 79:3468–3475. 10.1128/AEM.00068-13
28. Balows A. 1992. The prokaryotes: a handbook on the biology of bacteria. Ecophysiology, isolation, identification, applications. Springer-Verlag, New York, NY
29. Weimer PJ, Lopezguisa JM, French AD. 1990. Effect of cellulose fine-structure on kinetics of its digestion by mixed ruminal microorganisms in vitro. *Appl. Environ. Microbiol.* 56:2421–2429
30. Goering HK, Van Soest PJ. 1970. Forage fiber analyses (apparatus, reagents, procedures, and some applications). US Agricultural Research Service, Washington, DC
31. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39:W339–W346. 10.1093/nar/gkr466
32. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. 10.1186/1471-2105-11-119
33. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. 10.1371/journal.pcbi.1002195
34. Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. *Nucleic Acids Res.* 31:371–373. 10.1093/nar/gkg128
35. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:

- improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.  
10.1093/molbev/mst010
36. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 10.1093/bioinformatics/btu033
37. Teymouri F, Laureano-Perez L, Alizadeh H, Dale BE. 2004. Ammonia fiber explosion treatment of corn stover. *Appl. Biochem. Biotechnol.* 113:951–963.  
10.1385/ABAB:115:1-3:0951
38. Schülein M. 1997. Enzymatic properties of cellulases from *Humicola insolens*. *J. Biotechnol.* 57:71–81. 10.1016/S0168-1656(97)00090-4
39. Miller GL. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* 31:426–428. 10.1021/ac60147a030
40. Chen JQ, Weimer PJ. 2001. Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. *Microbiology* 147:21–30
41. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, Usadel B. 2012. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* 40:W622–W627. 10.1093/nar/gks540
42. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. 10.1093/bioinformatics/btp324
43. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.  
10.1093/bioinformatics/btp352
44. Anders S, Huber W. 2010. Differential expression analysis for sequence count data.

Genome Biol. 11(10):R106. 10.1186/gb-2010-11-10-r106

45. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13:2498–2504. 10.1101/gr.1239303
46. Poulsen M, Oh DC, Clardy J, Currie CR. 2011. Chemical analyses of wasp-associated *Streptomyces* bacteria reveal a prolific potential for natural products discovery. *PLoS One* 6(2):e16763. 10.1371/journal.pone.0016763
47. Oh DC, Scott JJ, Currie CR, Clardy J. 2009. Mycangimycin, a polyene peroxide from a mutualist *Streptomyces* sp. *Org. Lett.* 11:633–636. 10.1021/ol802709x
48. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37:D233–D238. 10.1093/nar/gkn663
49. Bianchetti CM, Harmann CH, Takasuka TE, Hura GL, Dyer K, Fox BG. 2013. Fusion of dioxygenase and lignin-binding domains in a novel secreted enzyme from cellulolytic *Streptomyces* sp. SirexAA-E. *J. Biol. Chem.* 288:18574–18587. 10.1074/jbc.M113.475848

### **Chapter 3: Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression**

**Previously published:** Adam J. Book\*, Gina R. Lewin\*, Bradon R. McDonald, Taichi E.

Takasuka, Evelyn Wendt- Pienkowski, Drew T. Doering, Steven Suh, Kenneth F. Raffa, Brain G. Fox, and Cameron R. Currie; \*authors contributed equally (2016). Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression. *PLoS Biology*. 14(6): e1002475.

**Co-authors contributed in the following ways:** Conceived and designed the experiments: AJB, GRL, BRM, TET, BGF, CRC. Performed experiments and analyzed data for Figure 3.1: GRL, AJB, BRM, TET, DTD; Figure 3.2: AJB, BRM, GRL; Figure 3.3: AJB, GRL, SS; Figure 3.4: EWP, GRL, AJB; Figure 3.5: AJB, BRM. Contributed reagents or materials: KFR, BGF, CRC. Wrote the paper: AJB, GRL, BGF, CRC.

#### **3.1 Abstract**

The evolution of cellulose degradation was a defining event in the history of life. Without efficient decomposition and recycling, dead plant biomass would quickly accumulate and become inaccessible to terrestrial food webs and the global carbon cycle. On land, the primary drivers of plant biomass deconstruction are fungi and bacteria in the soil or associated with herbivorous eukaryotes. While the ecological importance of plant-decomposing microbes is well established, little is known about the distribution or evolution of cellulolytic activity in any bacterial genus. Here we show that in *Streptomyces*, a genus of Actinobacteria abundant in soil and symbiotic niches, the ability to rapidly degrade cellulose is largely restricted to two clades of host-associated strains and is not a conserved characteristic of the *Streptomyces* genus or host-

associated strains. Our comparative genomics identify that while plant biomass degrading genes (CAZy) are widespread in *Streptomyces*, key enzyme families are enriched in highly cellulolytic strains. Transcriptomic analyses demonstrate that cellulolytic strains express a suite of multi-domain CAZy enzymes that are coregulated by the CebR transcriptional regulator. Using targeted gene deletions, we verify the importance of a highly expressed cellulase (GH6 family cellobiohydrolase) and the CebR transcriptional repressor to the cellulolytic phenotype. Evolutionary analyses identify complex genomic modifications that drive plant biomass deconstruction in *Streptomyces*, including acquisition and selective retention of CAZy genes and transcriptional regulators. Our results suggest that host-associated niches have selected some symbiotic *Streptomyces* for increased cellulose degrading activity and that symbiotic bacteria are a rich biochemical and enzymatic resource for biotechnology.

### **3.2 Introduction**

Cellulose is the principal component of plant cell walls and the most abundant organic compound in terrestrial ecosystems [1]. While cellulose deconstruction is critical to ecosystem functioning and the global carbon cycle, only select lineages of fungi and bacteria have evolved the ability to efficiently degrade this highly recalcitrant substrate [2–7]. These cellulolytic microbes are abundant in the soil and are also commonly associated with herbivorous eukaryotes. While a number of these biomass-degrading microbes have been isolated and characterized with genomic, transcriptomic, and biochemical analyses [8–12], relatively little is known about the evolution of this activity in any bacterial genus. Comparative genomic studies demonstrate that important plant biomass-degrading genes are highly variable across multiple taxonomic levels in both fungi and bacteria, suggesting that ecological niche likely drives

evolution of cellulolytic ability [3,4,13,14]. Understanding the distribution and evolution of cellulolytic activity in ecologically important bacteria connects microbial diversity to ecosystem function and the global carbon cycle and also provides a powerful approach to identify enzymes for the production of biofuels and other biocommodities.

Actinobacteria from the genus *Streptomyces* are thought to be significant players in the deconstruction of plant biomass [15–17]. These organisms are one of the most abundant genera in soils and recently have been identified living in symbiotic relationships with a wide variety of eukaryotic hosts [18–20]. *Streptomyces* grow as a filamentous hyphal mass, often in low nutrient environments [16,21,22]. They secrete extracellular enzymes (cellulases and proteases) and diverse secondary metabolites with antimicrobial activities [15,23]. Genomic analyses demonstrate that nearly all *Streptomyces* genomes contain Carbohydrate-Active Enzymes (CAZy) and accessory proteins [3]. Individual cellulolytic strains have been isolated from soil and host-associated niches [24,25], and recent work has characterized the transcriptomic and proteomic response of insect-associated strains when grown on plant biomass [11,26].

While *Streptomyces* have been extensively investigated for their ability to produce antibiotic compounds [21,27,28], relatively little is known about the distribution and evolution of plant biomass deconstruction across this, or any, genus of bacteria. Our comparative genomic, transcriptomic, and phylogenetic analyses demonstrate that a small subset of strains evolved the ability to deconstruct cellulose through modified gene content, gene expression, and enzyme architecture. We show that catalytic enzymes and transcription factors are critical for rapid cellulose degradation using targeted gene deletions. These results demonstrate that complex

evolutionary processes, including acquisition and selective retention of multiple genes and transcriptional regulators, drive the evolution of plant biomass deconstruction in bacteria.

### 3.3 Results and Discussion

#### 3.3.1 Genus-wide cellulolytic ability

We systematically compared the phylogenetic diversity of over 1,100 strains of *Streptomyces* and measured the cellulose degrading activity of 223 diverse strains isolated from free-living (soil and marine) and eukaryotic host-associated niches (Figure 3.1, Table A3.1). We identified strains that exhibited high rates of cellulose degrading activity, deconstructing filter-paper (FP) with comparable activity to the cellulolytic strain *Streptomyces* sp. SirexAA-E [11,20]. Rapid deconstruction of FP was restricted to just 13% of strains tested (29/223, Figure 3.1), and out of these 29 highly cellulolytic strains, 86% grouped in two phylogenetically distinct clades (I and III). Interestingly, 93% of all strains in these two clades were isolated from taxonomically diverse eukaryotic hosts (e.g., woodwasps, mountain pine beetles, ants, honeybees, and white-rot fungi, Figure 3.1a). These results suggest that rapid deconstruction of cellulose evolved in these two clades of host-associated strains and is not a conserved characteristic across the *Streptomyces* genus or across all host-associated strains. Rapid cellulose degradation is not limited to clades I and III, as we identify two distantly-related host-associated strains, *Streptomyces* spp. Amel2xE9 and LamerLS-31b, with high cellulolytic activity (Figure 3.1a). Taken together, these results suggest that the symbiotic environment selects for increased rates of cellulose deconstruction in some lineages of host-associated *Streptomyces*. High cellulolytic activity is not restricted to symbiotic strains, as it has been reported in *Streptomyces reticuli* [29] and several strains isolated from forest soil [24,30]. Future sampling will likely identify

additional strains or perhaps even clades with high activity. This cellulose degradation assay identifies strains that are capable of degrading cellulose in pure culture, through the expression and secretion of a full suite of cellulose degrading enzymes. It is likely that some of the low activity strains are capable of deconstructing cellulose in nature; however, in our assay, they are

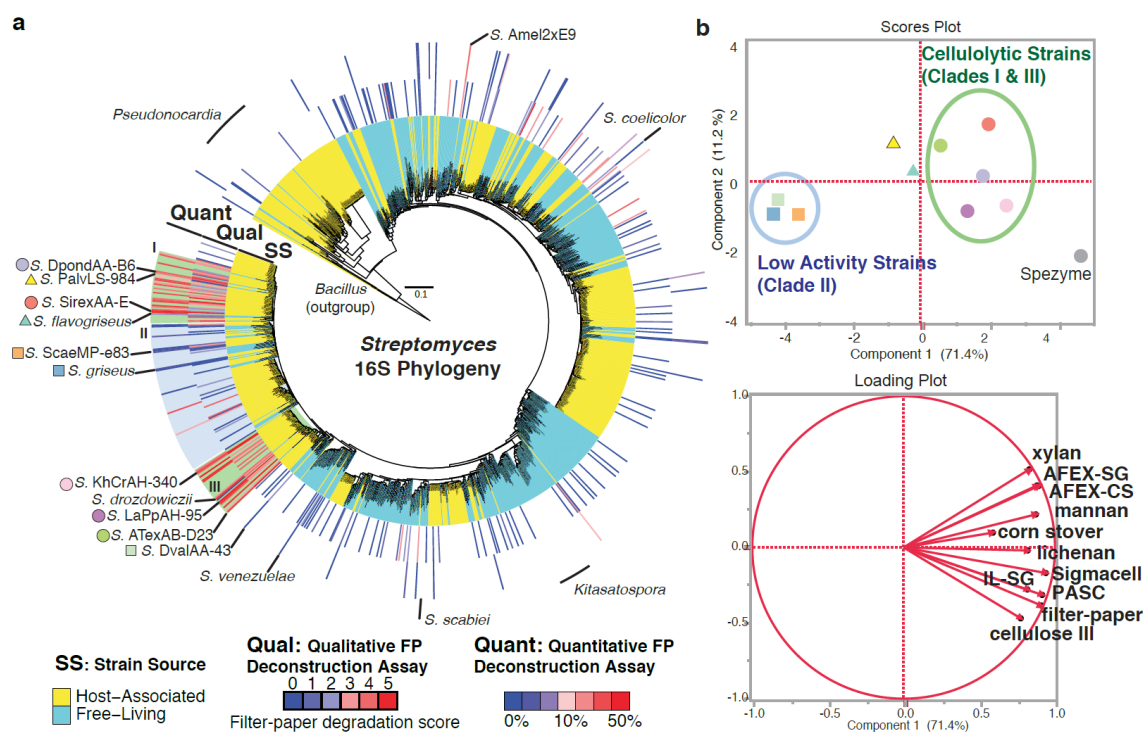


**Figure 3.1. Distribution of cellulolytic ability in the genus *Streptomyces*.** (a) 16S rRNA gene phylogenetic tree of 1,141 *Streptomyces* strains from free-living (cyan) and host-associated (yellow) environments. The tree is annotated with qualitative cellulose (filter-paper) degradation scores (0: no growth in 3 wk, 5: filter-paper deconstruction in 1 wk) and quantitative cellulose degrading activities (% filter-paper degraded in 10 d). Shading indicates highly cellulolytic clades I and III (green) and related low activity clade II (blue). (b) Principle component analysis of cellulose, hemicellulose, and plant biomass degrading activity of *Streptomyces* secretomes. Strains are identified by colored shapes on the tree in panel A. Scores plot shows similarity of polysaccharide degrading activity. Loading plot indicates which substrates influence components 1 and 2 of the scores plot.



missing critical interactions with environmental community members (e.g., biofilm formation [31], succession of decomposition from fungi [32], mutualistic or parasitic partners) or environmental conditions (e.g., nitrogen limitation, pH, temperature [33,34]).

### **3.3.2 Polysaccharide-degrading activity of *Streptomyces* secretomes.**

To further study the evolution of cellulose degradation in *Streptomyces*, we focused our studies on a subset of the genus consisting of two clades (I and III) that rapidly deconstruct crystalline cellulose, and a low-activity sister clade II (Figure 3.1). First, we quantified and visualized total secreted proteins produced by representative strains from all three clades when grown with glucose or Ammonia Fiber Expansion (AFEX) pre-treated corn stover as the sole carbon source. The total secreted protein concentration was higher for strains from clades I and III than strains in clade II when grown on AFEX corn stover (Figure A3.1). Furthermore, the protein profile produced by strains in clades I and III showed a suite of distinct and abundant polypeptides, while strains in clade II produced a low abundance smear of protein across the lane (Figure A3.1a). To investigate which components of plant biomass these strains deconstruct, we isolated their secreted proteins and assayed activity on a variety of polysaccharide substrates including cellulose, hemicellulose, and plant biomass (Figure A3.2). Principal component analysis of our biochemical assays showed that strains in clades I and III grouped separately from those in clade II (Figure 3.1b). Some cellulolytic strains from clades I and III had cellulose and biomass degrading activities comparable to a commercial enzyme mixture derived from the secreted proteins of *Trichoderma reesei* Rut-C30 (Spezyme CP, Figure A3.2). These results show that some host-associated cellulolytic *Streptomyces* secrete enzymes that efficiently deconstruct crystalline cellulose and plant biomass and thus may contribute to nutrient acquisition of the

herbivorous host. Furthermore, these strains may be a promising and largely untapped source of enzymes for the production of biocommodities from plant biomass [35].

### **3.3.3 Comparative genomic analysis of plant biomass degrading genes.**

To examine the hypothesis that genomic content is a critical factor determining cellulolytic ability, we sequenced 28 genomes of host-associated *Streptomyces*. These new genomes, combined with previously sequenced genomes, created a dataset containing 67 host-associated and 57 free-living strains, which spans the genus and represents both rapid cellulose degrading strains and low activity strains (Table A3.2, Figure 3.2 and Figure A3.3). Multilocus phylogenetic analysis shows that clade I is composed of 2 subclades of highly similar strains collected from different hosts and geographies (Figure 3.2, Table A3.1), clade III is composed of three unique species, and clade II contains >15 putative species. Thus, the cellulolytic clades I and III contain relatively few unique species based on genome content but represent a wide variety of host species (e.g., mountain pine beetle, Sirex woodwasp, honey bee, termite, leaf-cutter ants, and arboreal ants), while clade II represents greater bacterial species diversity. Additionally, this phylogenetic analysis identified clades of host-associated strains with low cellulolytic activity (Figure A3.3). We hypothesize that these strains may contribute a different function to the host, such as defense from pathogens (e.g., *Streptomyces* sp. SPB74 [36]).

Comparative genomic analysis of CAZy content demonstrated that strains across the genus encode the genes necessary to deconstruct plant biomass (Figure 3.2 and Figure A3.3). Genes in glycoside hydrolase (GH) families with putative endoglucanase, cellobiohydrolase,  $\beta$ -glucosidase, mannanase, xylanase, chitinase, and lytic polysaccharide monooxygenases (LPMO,

AA10) activities were widespread across the genus (Figure 3.2 and Figure A3.3). However, highly cellulolytic clades I and III were enriched in total CAZy genes (percent CAZy), in important enzyme families to deconstruct cellulose (GH5, GH6, GH9, GH12, GH48, and AA10) and xylan (GH10, GH11, GH30), and in cellulose-targeting carbohydrate-binding module

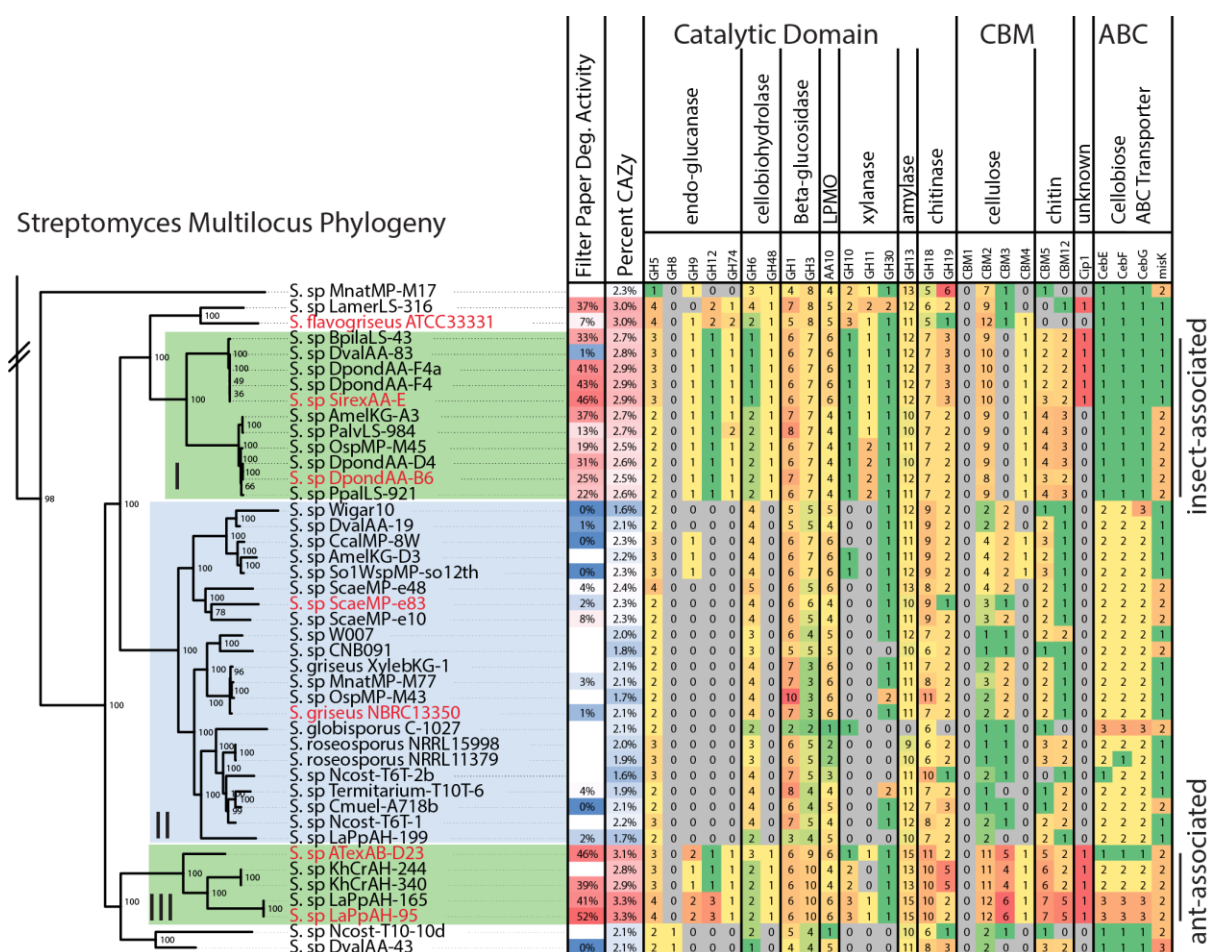


**Figure 3.2. Comparative analysis of CAZy genes across the *Streptomyces* genus.** Multilocus phylogenetic tree of clades I–III (full tree: Figure A3.3). Taxonomy tree is a RAxML tree calculated from an alignment of 97 genes conserved across all species. Tree is rooted with the outgroup *Kitasatospora setae*, and bootstrap support for each node is indicated (100 bootstraps). Tree is annotated with filter-paper degradation activity and select data from the CAZy and ABC transporter analyses. The number of genes present in each functional category is indicated. Rapid cellulose degrading clades I and III are indicated in green, low activity clade II in blue. Red strains were selected for RNA-seq analysis.

(CBM) families 2 and 3. Clustering analysis of CAZy profiles demonstrated that cellulolytic clades I and III have similar CAZy gene content, while clade II groups with non-phylogenetically related strains (Figure A3.4). Clade II strains all lack a reducing end cellobiohydrolase (GH48) and an endoglucanase (GH9, GH12, GH74), and most strains do not have xylanase genes (GH10, GH11) (Figure 3.2). The absence of these non-redundant activities likely reduces their ability to deconstruct cellulose and plant biomass. These results suggest an ancient origin for some CAZy genes within this genus followed by selective expansion of specific families in the cellulolytic host-associated clades. Furthermore, they suggest that free-living *Streptomyces* in the soil have not evolved the capacity to rapidly utilize all the components of plant biomass. Interestingly, the CAZy gene content of phytopathogenic strains *Streptomyces scabiei* and *Streptomyces acidiscabies* clusters closely with the cellulolytic strains in clades I and III, suggesting that cellulolytic activity contributes to phytopathogenesis.

Transport of plant biomass breakdown products, free sugars and small oligosaccharides, into the cell is critical for utilization of these recalcitrant carbon sources [37,38]. We analyzed all strains in our genomic dataset for ABC sugar transporter genes (Figure 3.2 and Figures A3.3 and A3.5 and Dataset 3.1). The cellobiose binding and permease proteins *cebE*, *cebF*, and *cebG* as well as the *MsiK* ATPase were conserved across all strains in clades I–III and nearly all strains in the genus. Similarly, xylose, N-acetylglucosamine, and multiple-sugar transporters were conserved across most of the genus (Figure A3.5). Together, the genomic analyses show variability of the plant biomass deconstructing enzymes across the genus and high conservation of the transporters. Thus, the ability to rapidly deconstruct crystalline cellulose and hemicellulose is not widespread in the *Streptomyces*, whereas utilization of their breakdown products is common to

the genus and likely a major carbon source for free-living strains. These “free-living” strains live in complex multi-species communities, where they could cooperate with other microbes to deconstruct plant biomass, or they might specialize in competing for the breakdown products released by the cellulolytic enzymes produced by other microbes.

### **3.3.4 Comparative gene expression analysis.**

RNA-seq analysis of seven *Streptomyces* from clades I, II, and III (Figure 3.2, red labels) identified a dramatic disparity in differential gene expression between cellulolytic and non-cellulolytic strains when grown with plant biomass or glucose as the carbon source (Dataset A3.2). When grown on AFEX corn stover, the cellulolytic strains SirexAA-E, DpondAA-B6, ATexAB-D23, and LaPpAH-95 induced 23, 16, 49, and 39 genes greater than 4-fold, respectively (Figure 3.3a), while the clade II strains *Streptomyces griseus* and ScaeMP-e83 induced only two and seven genes greater than 4-fold, respectively, compared to growth on glucose (Figure 3.3b). CAZy genes represented a large percentage of the highly induced genes in cellulolytic strains (22%–48%), while *S. griseus* and ScaeMP-e83 did not induce any CAZy genes greater than 4-fold. When we consider the genes significantly upregulated 2-fold or greater, these patterns are similar. DpondAA-B6, SirexAA-E, ATexAB-D23, and LaPpAH-95 induced nine, 12, 15, and 32 CAZy genes, respectively, including at least six cellulases and two xylanases (Table A3.3, Dataset A3.2). *S. griseus* and ScaeMP-e83 upregulated four and 12 CAZy genes, respectively, greater than 2-fold, including only one cellulase gene, coding for a GH6, and no xylanase genes. Additionally, all strains induced ABC transporter genes, including the celooligosaccharide sugar binding protein, *cebE*, indicating that all the strains had the capability to uptake cellulose breakdown products (Table A3.3, Dataset A3.2). Finally, all

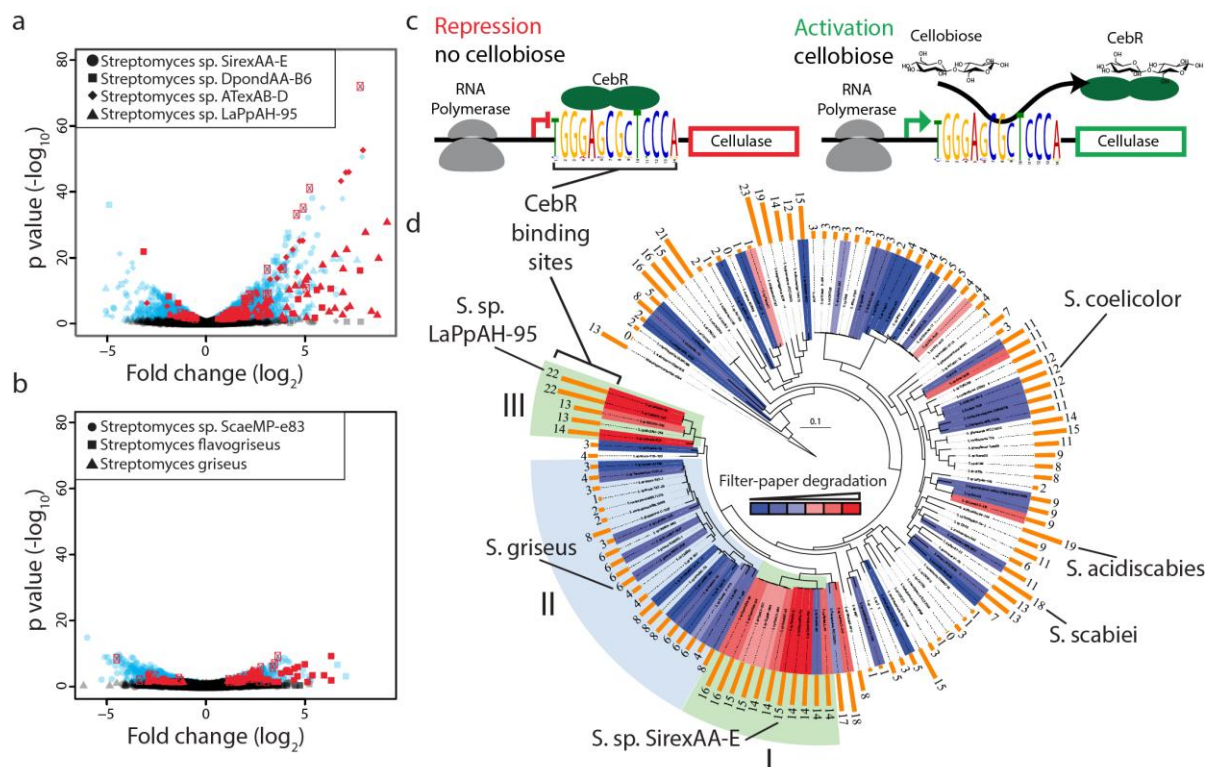


**Figure 3.3. Differential expression and coregulation of biomass degrading genes.** Differential expression (DE) of genes from clades I and III **(a)** and clade II **(b)**. Strains were grown with glucose or AFEX pretreated corn stover as the sole carbon source. Each point represents a gene; the shape indicates the strain. The x-axis shows the fold change between carbon sources, and the y-axis shows the statistical support of the fold change. Black points indicate non-significant DE ( $p$ -value  $> 0.05$ ). Red points identify CAZy annotated genes with significant DE. **(c)** Model for negative transcriptional control of CAZy genes by the CebR transcriptional regulator. The CebR binding sequence is the consensus of the top 25 coexpressed genes from each of the four cellulolytic strains. **(d)** *Streptomyces* multilocus phylogenetic tree annotated with quantitative cellulose (filter-paper) degrading activity (red and blue heatmap) and the number of CebR transcriptional regulator binding sites (TGGGAGCGCTCCCA) in the genome (orange bars).

strains, except DpondAA-B6, induced expression of multiple transcriptional regulators greater than 2-fold when grown on AFEX corn stover (Table A3.3, Dataset A3.2).

### 3.3.5 Coordinated expression of CAZy genes controlled by the CebR transcription regulator.

The striking contrast in gene expression between rapid cellulose-degrading strains and low activity strains led us to hypothesize that coordinated expression of CAZy genes contributed to the difference in cellulolytic ability. We analyzed the promoter regions of the top 25 induced genes in each RNA-seq dataset and identified a statistically significant 14 bp palindromic sequence (TGGGAGCGCTCCCA,  $e$ -value  $9.2e^{-162}$ ) present in the cellulolytic strains that matches the binding element for the CebR transcriptional regulator (Figure A3.6a). In Actinobacteria, CebR proteins are LacI/GalR-like transcription factors that control gene expression in response to the presence of cellobiose or other small cello-oligosaccharides (Figure 3.3c) [39–42]. Genus-wide, there was a positive correlation between the number of CebR binding sites in a genome and the cellulolytic ability of the strain (Figure 3.3d and Figure 3.6b). We analyzed the genomes of the RNA-seq strains for genes predicted to be directly controlled by CebR and mapped on experimentally observed differential expression values (Table A3.4). Interestingly, strains from clades I and III had more genes predicted to be CebR controlled, and a higher percentage of these genes were significantly differentially expressed greater than 2-fold: Clade I strains SirexAA-E had 10/18 CebR controlled genes expressed >2-fold, DpondAA-B6 11/19; Clade III strains ATexAB-D23 14/19, LaPpAH-95 21/30; and Clade II strains *S. griseus* 2/8, ScaeMP-e83 3/6. Thus, identification of strains enriched with CebR binding sites likely predicts increased cellulolytic ability. Remarkably, the phytopathogenic strains *S. scabiei* and *S. acidiscabies* have 18 and 19 CebR sites in their genomes, respectively, including a site in the promoter region of the thaxtomin toxin biosynthetic gene that controls toxin production [43]. We confirmed that CebR controls coordinated expression of CAZy genes by generating a CebR deletion strain in SirexAA-E (SACTE\_2285,  $\Delta$ cebr, Figure 3.4). This strain constitutively

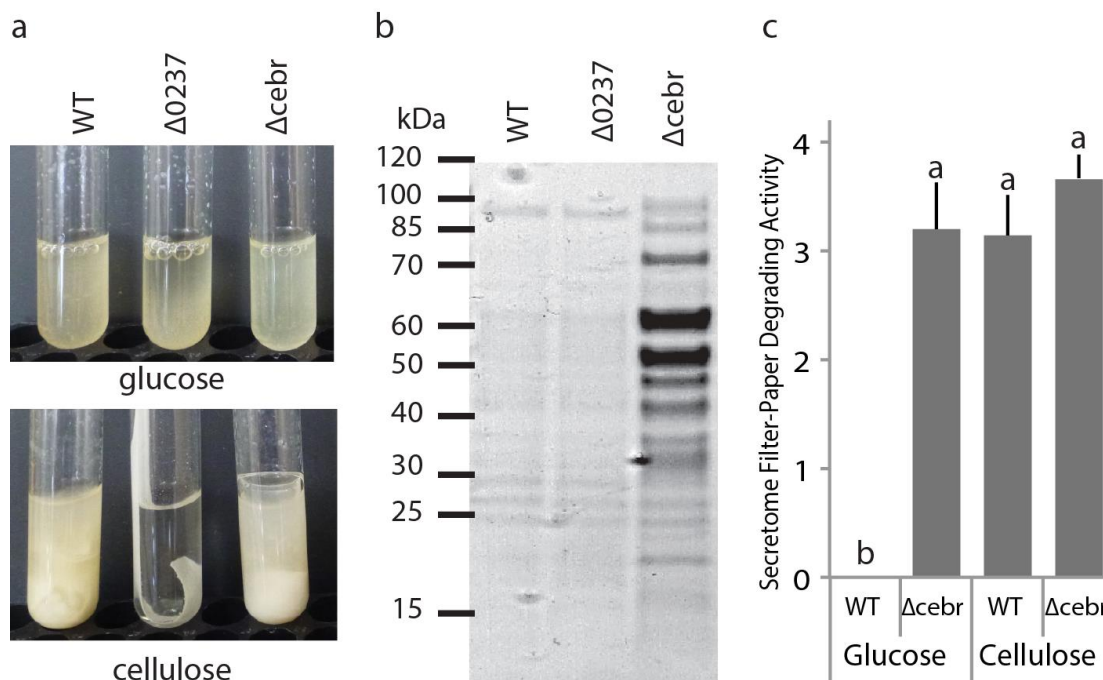


**Figure 3.4. Genetic analysis of *Streptomyces* sp. SirexAA-E.** (a) Phenotypic analysis of *Streptomyces* sp. SirexAA-E wild-type (wt), Δ0237 (GH6), and Δcebr (cellulase transcriptional regulator) strains grown on glucose or cellulose as the sole carbon source. (b) Secreted protein profile of wt, Δ0237, and Δcebr strains grown in glucose media. (c) Filter-paper degrading activity (mg glucose released per mg total protein) of secreted proteins isolated from wt and Δcebr strains grown with glucose or cellulose as the carbon source. Statistically significant differences are indicated.

expressed and secreted a suite of cellulolytic enzymes in cellulose-free media (Figure 3.4b–3.4c and Figure A3.7). Constitutive expression of a transgenic CebR in the Δcebr strain resulted in wild-type levels of secreted cellulases in glucose media and reduced protein when grown in cellulose media (Figure A3.8b).

Comparison of a phylogenetic tree of the CebR repressor with the *Streptomyces* phylogeny shows incongruence of this transcription factor with the evolutionary history of the strains (Figure A3.9a). Cellulolytic strains with a large number of CebR binding sites in their genomes



contained more similar CebR gene sequences than the related species with low activity.

Molecular evolutionary analysis showed that CebR from the low activity strains contained only residues with neutral ( $1.2 > \omega > 1$ ) or purifying selection ( $\omega < 1.2$ , Figure A3.9b), while residues in the cellulolytic clades showed positive selection to diversify ( $\omega > 1.2$ ). Sites with positive selection were located on the surface of the protein and may influence ligand or DNA binding directly or through allosteric interactions. Together, the gene expression and molecular evolution data suggest an important connection between the evolution of the CebR transcriptional regulator, the number of CebR binding sequences in the genome, and the cellulolytic ability of the strain.

### ***3.3.6 Analysis of Important Enzyme Families***

In three of the four cellulolytic strains (SirexAA-E, ATexAB-D23, and LaPpAH-95), a GH6 family non-reducing end cellobiohydrolase was the most highly induced gene when grown on AFEX corn stover and was significantly induced greater than 2-fold in all strains (Dataset A3.2). To test the importance of the GH6 enzyme to the cellulolytic phenotype, we generated an in-frame deletion of the GH6 gene (SACTE\_0237) in the highly cellulolytic strain SirexAA-E (Figure 3.4a and Figure A3.7). The mutant strain ( $\Delta 0237$ ) grew normally in glucose media; however, no growth, secreted proteins, or significant cellulolytic activity was observed in cellulose media (Figure A3.7). To confirm that this phenotype was due solely to the loss of the GH6 protein, we generated a  $\Delta 0237 \Delta \text{cebR}$  double mutant (Figure A3.7) to constitutively express the cellulases. We observed that the secretome was nearly identical to the  $\Delta \text{cebR}$  single mutant except for the missing GH6 protein band (Figure A3.7e). The filter-paper degrading activity of the double mutant was reduced to near zero, similar to the single  $\Delta 0237$  mutant

(Figure A3.7d). The mutant phenotype was rescued by integration of a constitutively expressed transgene (*ermE\*::SACTE\_0237*) into the  $\Delta 0237$  strain (Figure A3.8b). These results support a growing body of literature that suggest the importance of individual CAZy genes for cellulolytic ability [44]. Furthermore, they demonstrate that rapid cellulolytic activity is the product of a complete suite of enzymes (endoglucanase, reducing and non-reducing end cellobiohydrolase, LPMO), and absence of any one of these enzymes results in limited activity. Thus, the absence of the endoglucanase genes (GH9, GH12, GH74), a reducing end cellobiohydrolase gene (GH48), and xylanase genes (GH10 and GH11) in most clade II strains likely has a major impact on their cellulose and plant biomass degrading ability in isolation.

The importance of lytic polysaccharide monooxygenases (LPMOs, AA10) to the deconstruction of recalcitrant polysaccharides is increasingly recognized [45–48]. LPMOs are abundant in *Streptomyces* genomes and include both chitinolytic and cellulolytic AA10 genes [49]. Our genus-wide genomic analysis shows that AA10 genes are present in all strains, and there is little difference in the number of genes between highly cellulolytic clades I and III and low activity clade II (Figure 3.2). To gain further insight into their variation and utilization, we examined the phylogenetic distribution of AA10 proteins from clades I–III mapped with RNA-seq gene expression data and transcriptional regulator binding sites (CebR, Figure A3.10). This tree shows that strains in clade I and III have three homologs in the cellulolytic clade and three homologs in the chitinolytic clade, while clade II strains (e.g., *S. griseus*) have two cellulolytic AA10s and four chitinolytic AA10s. The most highly expressed homolog (clade A1, Figure A3.10) was conserved in strains from all clades. All protein sequences in clade A1 contained a CBM2 domain, and analysis of the upstream DNA identified a CebR transcription factor binding site.

However, differential expression of the *S. griseus* gene was less than half of those from strains in clades I and III (SirexAA-E, DpondAA-B6, LaPpAH-95, and ATexAB-D23) (Figure A3.10 and S2 Dataset). The second most highly expressed homolog (clade A2, Figure A3.10) only contained sequences from the rapid cellulose-degrading clade I and III strains. These sequences were single domain proteins, and the upstream genomic DNA encodes a CebR binding site. Most genes in the chitinolytic clade had low and insignificant differential expression values when grown on AFEX corn stover, suggesting a different regulatory system. Together, these results suggest small but potentially significant differences in AA10 homologs between rapid crystalline cellulose degrading strains and strains with low activity.

Interestingly, three cellulolytic strains (SirexAA-E, ATexAB-D23, and LaPpAH-95) strongly induced expression of a homolog of Cip1 from *T. reesei* (42.2% identity). Cip1 is highly expressed in *T. reesei* during growth on cellulose, but the biochemical function is unknown [50]. Cip1 homologs are present in only 3 bacterial phyla (Acintobacteria, Proteobacteria, and Chloroflexi) and present in just 21% of *Streptomyces* genomes in this dataset (Figure A3.2 and Figure A3.11). Phylogenetic analysis of Cip1 protein sequences identifies monophyletic clades for the Ascomycota, Actinobacteria, and Proteobacteria phyla (Figure A3.11). *Streptomyces* Cip1 sequences are polyphyletic and incongruent with the multilocus phylogeny (Figure A3.3 and Figure A3.11). These data suggest possible lateral transfer of Cip1 within the *Streptomyces* genus. Additionally, the four cellulolytic strains upregulated a combined 20 genes (greater than 4-fold) annotated as unknown function, suggesting that these *Streptomyces* may contain additional proteins with novel biochemistry.

### 3.3.7 Evolution of CAZy genes in cellulolytic strains.

Our comparative genomic and transcriptomic analyses also identified differences in domain architecture of cellulolytic enzymes, which is critical for substrate specificity and specific activity [9,51]. Clade I is enriched in CBM2 domains, while clade III is enriched in both CBM2 and CBM3 domains (Figure 3.2). Additionally, a protein homology network showed that nearly all highly expressed CAZy enzymes contained both a catalytic and a CBM domain (Figure 3.5a). We hypothesized that expansion of multi-domain proteins in cellulolytic strains contributes to their biomass degrading ability. To test this hypothesis, we first modeled the evolutionary history of each CAZy gene family in *Streptomyces* including gene origin (lateral gene transfer, gene duplication, or ancestral) and loss events (Figure A3.12). Although the CBM2 genes are the most abundant family in extant genomes, only 55% of total possible genes were retained across the evolutionary history of the genus. When gain and loss of CBM2 genes were plotted on individual species, we observed a dramatic difference between the rapid crystalline cellulose degrading strains and strains with low activity (Figure 3.5b and Figure A3.13). Clades I and III were enriched in the total number of CBM2 domains and had higher retention rates than clade II (79% versus 27% retained on average, respectively). The largest inferred source of CBM2 genes in cellulolytic clades was horizontal gene transfer, followed by gene duplication, and then vertical inheritance (ancestral genes, Figure 3.5b). Interestingly, CBM3 genes were expanded and retained almost exclusively in clade III (Figure A3.13). In contrast, the GH6 family had a lower retention rate in the highly cellulolytic clades than the low activity clade, 40% and 53%, respectively (Figure A3.13). These results indicate that while the CBM2 and CBM3 families of genes are under selection to expand and diversify in the cellulolytic strains, there is no selective advantage for additional GH6 family genes in these strains. Together, these data suggest that

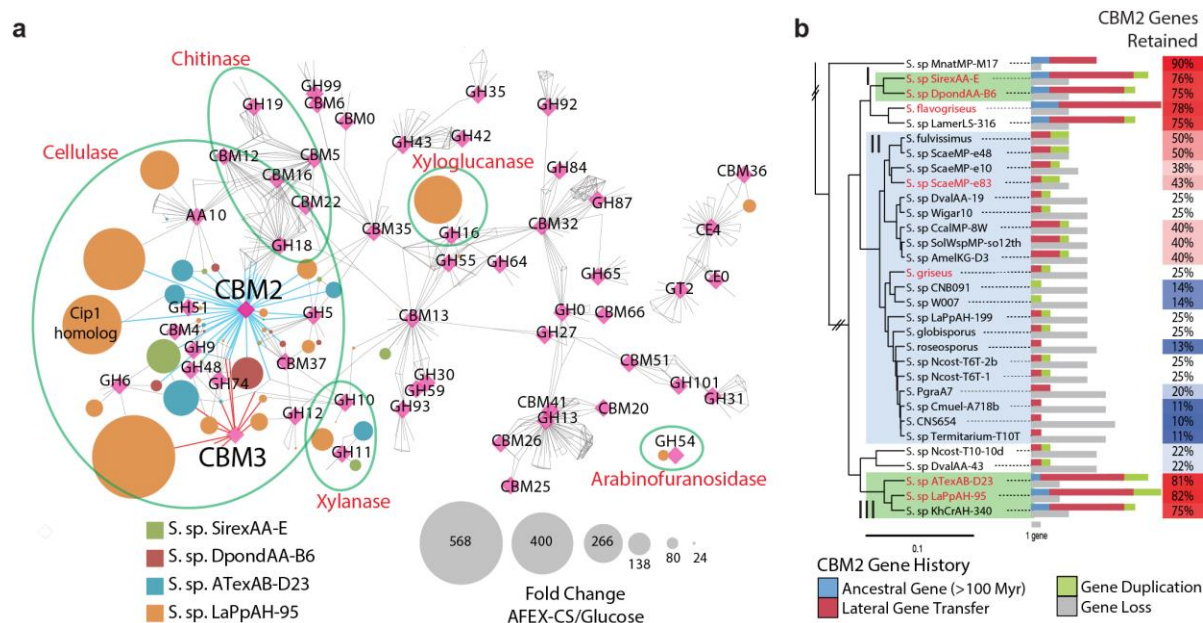


**Figure 3.5. Multi-domain enzyme expression and evolution. (a)** Protein similarity network of CAZymes present in four cellulolytic *Streptomyces* strains. Nodes are proteins (circles) or CAZy functional categories (magenta diamonds); edges indicate that the gene belongs to the respective CAZy family or BLAST similarity with an  $e$ -value  $< 1 \times 10^{-50}$ . Node size represents the fold-change in RNA abundance between glucose and AFEX corn stover grown cells. CBM2 and CBM3 linked proteins are indicated with blue and red edges, respectively. **(b)** Evolutionary expansion of CBM2 domains in cellulolytic *Streptomyces*. CBM2 domain gain and loss events for strains in cellulolytic clades I & III (green shading) and clade II (blue shading) mapped onto the multilocus phylogenetic tree. CBM2 retention rates (genes retained / total possible genes) are identified by the heatmap.

targeted expansion of the CBM2 and CBM3 families through horizontal transfer and selective gene retention likely contributed to the evolution of cellulolytic ability in host-associated *Streptomyces*. Furthermore, they emphasize that clade II strains have not been selected for increased ability to bind and deconstruct crystalline cellulose, and thus these strains may obtain sufficient carbon from plant biomass breakdown products.

Our genomic, proteomic, and transcriptomic analyses of *Streptomyces* identify key genome modifications that increase cellulolytic ability, and we determine the evolutionary processes that contributed to these changes. While genomic content predicts that nearly all *Streptomyces* are capable of crystalline cellulose deconstruction, our results demonstrate that high activity is limited to select lineages of host-associated strains. Selective pressure from the symbiotic environment likely selects these strains for increased CAZy gene content, retention of horizontally transferred genes, and expansion of regulatory elements. In contrast, free-living soil strains, having evolved within diverse communities, likely have experienced limited selection for rapid cellulose degrading activity in isolation. Evolution-focused multiomic studies of cellulolytic microbes are critical for understanding the functional contributions of environmental microbes to the carbon cycle, ecosystem processes, and climate change, as well as providing high-interest organisms and enzymes for the biofuels industry.

### **3.4 Materials and Methods**

#### ***3.4.1 Isolation of Streptomyces.***

Free-living strains were obtained from the USDA ARS Culture Collection at the National Center for Agricultural Utilization Research in Peoria, Illinois. Host-associated strains were isolated by the Currie-lab from diverse insects, plants, and fungi (S1 Table). In all cases, hosts were surface sterilized in 95% ethanol for 1 min, rinsed twice in sterile phosphate-buffered saline (PBS), homogenized in PBS, and dilution plated onto Chitin minimal medium [25]. Colony forming units were isolated and maintained on ISP2 (Difco).

### ***3.4.2 Culturing of organisms.***

Strains were grown in M63 minimal medium at pH 6.8 [11, 52]. Qualitative filter-paper deconstructive assays were performed with triplicate test tubes containing 5 mL M63 with either 0.5% w/v glucose or a 1 x 10 cm strip of Whatman #1 filter-paper (Millipore, Billerica, MA) as the sole carbon source. Tubes were observed over 3 wk and scored: 0 = no growth in 3 wk; 1 = growth in 2–3 wk, no degradation; 2 = growth in 1 wk, no degradation; 3 = degradation in 2–3 weeks; 4 = low degradation in 1 wk and heavy degradation in 2 wk; 5 = heavy degradation in 1 wk.

Large-scale growth was performed in flasks containing 50 mL M63 with either 0.5% w/v glucose or sterile AFEX-pretreated corn stover (GLBRC Corn Stover UW 2009 AFEX, AFEX ID 712–192) and was collected after 1 wk and used for enzyme assays and transcriptomics. All cultures were incubated at 30°C with shaking at 250 rpm.

### ***3.4.3 Quantification of residual cellulose.***

Cellulose degradation was quantified using previously described methods [26]. At least three replicates were performed for each strain.

### ***3.4.4 Analysis of secreted enzymes.***

Supernatants were prepared from growing cultures by first centrifugation, then filtration, and enzyme activity was measured using dinitrosalicylic acid (DNS) assays [11,26,53]. Reducing sugar assays were performed against Whatman #1 filter-paper (GE Healthcare Life Sciences, Pittsburgh, PA), Sigmacell-20 (Sigma-Aldrich, St. Louis, MO), cellulose-III (a generous gift

from Dr. Dale at Michigan State University, East Lansing [54]), phosphoric acid swollen cellulose (PASC) [55], birch-wood xylan (Sigma-Aldrich, St. Louis, MO), D-mannan (Megazyme, Ireland), corn stover (GLBRC 2009 Corn Stover), AFEX-pretreated corn stover (GLBRC Corn Stover UW 2009 AFEX, AFEX ID 712–192), and ionic-liquid pretreated switchgrass (a generous gift from Department of Energy Joint Bioenergy Institute, CA). The industrial enzyme cocktail, Spezyme CP (Genencor, Palo Alto, CA), was used as a comparison. Statistical analyses were performed in JMP Pro (SAS, Cary, NC).

For SDS-PAGE analysis, strains were grown with either glucose or AFEX-corn stover as the sole carbon source for 7 d and harvested by centrifugation and filtration. Total protein in each secretome was measured with a Bio-Rad protein assay (Bio-Rad, Hercules, CA), and BSA was used to generate a standard curve. Secretomes were normalized to equal protein concentrations and separated on a 4%–20% Bio-Rad Criterion gel. Gel was stained with Bio-Rad Bio-Safe coomassie stain.

#### ***3.4.5 Genomic sequencing and phylogenetic analysis.***

16S rRNA gene sequences for *Streptomyces* and representative Actinobacteria were downloaded from NCBI refseq. These were combined with sequences from strains isolated by the Currie-lab [25] and sequences extracted from *Streptomyces* genome assemblies. The full set of sequences was aligned using MAFFT [56] and trimmed based on visual inspection. The alignment was used to generate a phylogeny using RAxML-7.2.6. with the GTRGAMMA substitution model [57].



Genome sequences for 28 host-associated *Streptomyces* were generated (S2 Table) at the Joint Genome Institute (JGI [11,20]). Library construction and sequencing methods used at the JGI can be found at <http://www.jgi.doe.gov/>. Genome annotations were generated at the JGI. Putative CAZy gene annotations were generated using a custom script [11]. These 28 genomes were combined with 99 other high-quality *Streptomyces* genomes to create a dataset of 127 genomes. Multilocus phylogenetic analysis was performed as described in Book et al. 2014 [26] using 97 conserved genes families that were found to be present in all 127 genomes. CAZy gene profile similarity tree was generated by calculating a similarity matrix with the Spearman algorithm and then generating a neighbor-joining tree with phylip. ABC sugar transporters were identified by GhostKOALA analysis of all 127 *Streptomyces* genomes [58]. Genes annotated as ABC sugar transporters as defined by standard KEGG ontology were parsed and collected in S1 Dataset.

CebR phylogenetic trees were constructed as described in Book et al. 2014 [49]. Briefly, CebR homolog protein sequences were identified by BLAST analysis of NCBI and sequences were aligned with MUSCLE [59]. The phylogenetic tree was generated using the MrBayes [60] code on CIPRES Science Gateway with a calculated standard deviation of  $\leq 0.05$ . Non-default parameters were set to mcmc, ngen = 10,000,000, temp = 0.200, burninfrac = 0.25, stoprule = No, sump burnin = 4,000, and sumt burnin = 4,000. Phylogenetic analysis of Cip1 and AA10 were generated with similar methods. The Cip1 dataset was collected with a NCBI refseq database BLAST search with an *e*-value cutoff of  $1e^{-10}$ . The AA10 data set was generated by collecting all AA10 annotated genes from the above CAZy analysis in Clade I-III strains.

### 3.4.6 RNA analysis.

RNA was purified from *Streptomyces* strains using a modified phenol-chloroform extraction protocol, DNase treatment, and ribosomal RNA removal [11,26,61]. cDNA library construction and single-end 100 bp Illumina HiSeq2000 sequencing was performed at the University of Wisconsin Biotechnology Center (Madison, WI). Data were demultiplexed using the CASAVA software package, and analysis was performed using a previously published pipeline [26]. BLAST comparison of all CAZy annotated genes in *Streptomyces* strains SirexAA-E, DpondAA-B6, LaPpAH-95, and ATexAB-D23 were used to make homology networks in Cytoscape [62]. Proteins (circles) with a BLAST  $e$ -value  $< 1e^{-50}$  were connected with an edge, as were CAZy family annotations (triangles).

### 3.4.7 CebR analysis.

Promoter sequence-based motifs were identified using MEME [63]. We analyzed DNA sequences 500 bp upstream of the start codon for the top 25 most differently expressed genes in each RNA-seq dataset. The occurrence of motifs in the sequence was assumed to be distributed zero to three per sequence. The consensus motif was generated by collecting a dataset that combines promoter sequences (500 bp upstream of start) from *Streptomyces* spp. SirexAA-E, DpondAA-B6, LaPpAH-95, *flavogriseus*, and ATexAB-D23. This dataset of 125 sequences was then analyzed with MEME, and the consensus motif and position weight matrix were identified. CebR binding site analysis was performed using the FIMO analysis in the MEME suite using the CebR consensus motif generated by MEME and searched against all *Streptomyces* genomes in our dataset. Results were filtered to an  $e$ -value cutoff of  $1e^{-7}$  [64].

Evolutionary rate estimation was done as described in Book et al. 2014 [49]. DNA coding sequences for CebR genes were collected from NCBI and codon alignments were generated with MUSCLE. Codon alignments were masked with Zorro to generate quality scores for codon positions [65], and then a phylogenetic tree was generated with RAxML using the masking scores [57]. Site-specific codon substitution models were generated using the CODEML program in PAML [66], with model = 0, NSsites = 3, ncatG = 3, fix\_kappa = 0, fix\_omega = 0, cleandata = 1, and fix\_blength = 2. Protein homology models were generated by aligning proteins sequences to the 1BDH structure with ITASSER [67].

#### ***3.4.8 Lateral Gene Transfer Analysis.***

The AnGST algorithm [68] was used to infer horizontal gene transfer, gene duplication, and gene loss of CAZy domains across the *Streptomyces* phylogeny. The protein sequences for dbCAN annotated domain families were aligned using MAFFT [56]. These alignments were then reverse translated to codon alignments, RAxML-7.2.6 [57] was used to generate 10 bootstrapped alignments, and FastTree 2.1.7 [69] was used to generate phylogenies from these bootstrap alignments. The 10 bootstrap phylogenies were then used as the input for AnGST with default parameters. A molecular clock-based species tree of *Streptomyces* genomes, plus an additional 20 Actinobacterial genomes used as outgroups, was generated using Reltime [70] with Cyanobacterial fossils and the *Salmonella/Escherichia coli* divergence time used as calibration points.

### 3.4.9 *Streptomyces* sp. *SirexAA-E* Genetics.

*E. coli* XL1 Blue MR (Stratagene, La Jolla, CA) and *E. coli* DH10b (Life Technologies, Carlsbad, CA) were used for routine subcloning, plasmid preparations, and cosmid library preparations. *E. coli* S17-1 [16] was the donor strain for intergeneric conjugation. The pGEM series of vectors (Promega, Madison, WI) and Litmus 28 and 38 (New England Biolabs, Beverly, MA) were from commercial sources. pANT841 [71], pOJ260 [72], and pSET152 [72] were described previously. The *ermE*\* promoter was obtained from pWHM1250 [73]. SuperCos1 (Stratagene) was used for the cosmid library. DNA isolations and manipulations in *E. coli* [71] and *Streptomyces* [16] were carried out according to standard procedures.

A cosmid library of *Streptomyces* sp. *SirexAA-E* chromosomal DNA was constructed by partially digesting DNA with *Sau3AI*, dephosphorylating, and ligating into the *BamHI* site of SuperCos1. The ligation mixture was packaged using Gigapak III XL packaging mix (Agilent Technologies, Santa Clara, CA) and transduced into *E. coli* XLI Blue MR cells. The library was screened by colony hybridizations using digoxigenin-labeled PCR products of the *CebR* gene and the *GH6* gene to find cosmids containing these genes within the center portion of the inserts. Cosmids were end sequenced to determine their exact content. A ~7.5 kb *KpnI* fragment of DNA containing the *GH6* (SACTE\_0237) gene was subcloned from the cosmid library into pGEM7Zf. A ~5 kb *PstI-HindIII* fragment was transferred into Litmus38, and the resulting plasmid was cut with *SacI* and religated, removing 942 bp, thereby creating an in-frame deletion within the SACTE\_0237 gene. The entire insert was then transferred into pOJ260, yielding pOJ260Δ0237. This construct was transformed into *E. coli* S17-1 in preparation for intergeneric conjugation.

To generate the CebR deletion, a ~9 kb XhoI fragment was isolated from the cosmid library and cloned into a standard cloning vector. From this, a 2.1 kb *SphI-PshAI* fragment and a 2 kb *MluI* (Klenow treated)-*BamHI* fragment were isolated. Both fragments were ligated into pANT841, thereby creating a ~4 kb insert with an 810 bp in-frame deletion in the CebR gene. The entire region was removed as an *EcoRI-HindIII* fragment and cloned into pOJ260, yielding pOJ260 $\Delta$ cebr. This construct was transformed into *E. coli* S17-1 in preparation for intergeneric conjugation.

*E. coli*–*Streptomyces* intergeneric conjugation was carried out as described in Galm et al. [74] using standard literature conditions [16] with minor modifications. Exconjugants were picked to fresh supplemented ISP4 plates containing naladixic acid and apramycin. Colonies that grew were screened via PCR to ensure they had both the WT and the mutant copies of the gene in the chromosome. These “single crossover” mutants were then grown for several rounds on nonselective media and tested again by PCR for the loss of the WT copy of the gene. Those mutants showing only the deleted copy of the gene were then confirmed via Southern hybridization. Southern analysis and colony lifts, digoxigenin labeling of DNA probes, blotting to Hybond-N+, hybridization, and detection were performed according to the protocols provided by the manufacturers (GE Healthcare, Piscataway, NJ and Roche Diagnostics Corp., Indianapolis, IN).

Mutations were rescued by inserting a WT copy of each gene of interest behind the constitutive *ermE*\* promoter into the site-specific integrating vector pSET152 and introducing this plasmid into the respective mutant strains via intergeneric conjugation using *E. coli* S17-1 as the donor

strain. Exconjugants were picked to fresh supplemented ISP4 plates containing naladixic acid and apramycin. Colonies that grew were screened via PCR to ensure they had the WT copy of the gene present behind the *ermE\** promoter. Empty pSET152 was used as a control.

Protein profiles were obtained from strains of interest by growing 5 ml starter cultures in R2YE for 2–3 d. A 100 µl aliquot of these cultures was then used to inoculate 5 ml of M63 media supplemented with either glucose or cellulosic filter-paper as the sole carbon source. Cultures were grown in a 30°C shaking incubator overnight and aliquots were centrifuged, run on a 4%–20% Criterion TGX Gel, and visualized using Coomassie Stain (Bio-Rad, Hercules, CA). Reducing sugar assays were performed as described above with filter-paper as the carbon source and a 30°C incubation for 20 h.

#### **3.4.10 Accession numbers.**

The genome sequence data are available at NCBI (see S2 Table for accession numbers), and the RNA-seq data are available from the Sequence Read Archive (PRJNA290090).

### **3.5 References**

1. Schlesinger WH, Andrews JA. Soil respiration and the global carbon cycle. *Biogeochemistry*. 2000;48(1):7–20. doi: 10.1023/a:1006247623877
2. Zhao ZT, Liu HQ, Wang CF, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2013;14:274. doi: 10.1186/1471-2164-14-274

3. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases in bacteria. *Appl Environ Microbiol.* 2013;79(5):1545–54. doi: 10.1128/aem.03305-12
4. Glass NL, Schmoll M, Cate JH, Coradetti S. Plant cell wall deconstruction by ascomycete fungi. *Annu Rev Microbiol.* 2013;67:477–98. doi: 10.1146/annurev-micro-092611-150044. pmid:23808333
5. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002;66(3):506–77. pmid:12209002 doi: 10.1128/mmbr.66.3.506-577.2002
6. Matulich KL, Martiny JBH. Microbial composition alters the response of litter decomposition to environmental change. *Ecology.* 2015;96(1):154–63. pmid:26236900 doi: 10.1890/14-0357.1
7. McGuire KL, Treseder KK. Microbial communities and their relevance for ecosystem models: Decomposition as a case study. *Soil Biol Biochem.* 2010;42(4):529–35. doi: 10.1016/j.soilbio.2009.11.016
8. Bouws H, Wattenberg A, Zorn H. Fungal secretomes—nature's toolbox for white biotechnology. *Appl Microbiol Biotechnol.* 2008;80(3):381–8. doi: 10.1007/s00253-008-1572-5. pmid:18636256
9. Gilbert HJ. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Mol Microbiol.* 2007;63(6):1568–76. pmid:17367380 doi: 10.1111/j.1365-2958.2007.05640.x
10. Schuster A, Schmoll M. Biology and biotechnology of *Trichoderma*. *Appl Microbiol Biotechnol.* 2010;87(3):787–99. doi: 10.1007/s00253-010-2632-1. pmid:20461510
11. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci Rep.* 2013;3:1030. doi: 10.1038/srep01030

12. Wilson DB. Microbial diversity of cellulose hydrolysis. *Curr Opin Microbiol.* 2011;14(3):259–63. doi: 10.1016/j.mib.2011.04.004
13. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science.* 2012;336(6089):1715–9. doi: 10.1126/science.1221748. pmid:22745431
14. Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, et al. Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nat Genet.* 2015;47(4):410–5. doi: 10.1038/ng.3223. pmid:25706625
15. Chater KF, Biro S, Lee KJ, Palmer T, Schrempf H. The complex extracellular biology of *Streptomyces*. *FEMS Microbiol Rev.* 2010;34(2):171–98. doi: 10.1111/j.1574-6976.2009.00206.x. pmid:20088961
16. Kieser T, Buttner MJ, Chater KF, Hopwood DA. *Practical Streptomyces genetics*. Colney: John Innes Foundation; 2000.
17. Schrempf H. *Biology of Streptomycetes*. 3rd. ed. Dworkin M, editor. New York: Springer-Verlag; 2007.
18. Aylward FO, Suen G, Biedermann PHW, Adams AS, Scott JJ, Malfatti SA, et al. Convergent bacterial microbiotas in the fungal agricultural systems of insects. *Mbio.* 2014;5(6):e02077. doi: 10.1128/mbio.02077-14
19. Janssen PH. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl Environ Microbiol.* 2006;72(3):1719–28. doi: 10.1128/aem.72.3.1719-1728.2006



20. Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, et al. Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. ISME J. 2011;5(8):1323–31. doi: 10.1038/ismej.2011.14. pmid:21368904
21. Schatz A, Bugle E, Waksman SA. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. Exp Biol Med. 1944;55(1):66–9. doi: 10.3181/00379727-55-14461
22. Hopwood DA. *Streptomyces* in nature and medicine: the antibiotic makers. New York: Oxford University Press; 2007.
23. Doroghazi JR, Metcalf WW. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC genomics. 2013;14:611. doi: 10.1186/1471-2164-14-611. pmid:24020438
24. Bontemps C, Toussaint M, Revol PV, Hotel L, Jeanbille M, Uroz S, et al. Taxonomic and functional diversity of *Streptomyces* in a forest soil. FEMS Microbiol Lett. 2013;342(2):157–67. doi: 10.1111/1574-6968.12126. pmid:23489323
25. Hanshew AS, McDonald BR, Diaz Diaz C, Djieto-Lordon C, Blatrix R, Currie CR. Characterization of Actinobacteria associated with three ant-plant mutualisms. Microb Ecol. 2015;69(1):192–203. doi: 10.1007/s00248-014-0469-3. pmid:25096989
26. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, et al. Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. Appl Environ Microbiol. 2014;80(15):4692–701. doi: 10.1128/aem.01133-14
27. Bibb MJ. Regulation of secondary metabolism in streptomycetes. Curr Opin Microbiol. 2005;8(2):208–15. doi: 10.1016/j.mib.2005.02.016

28. Hopwood DA, Chater KF, Bibb MJ. Genetics of antibiotic production in *Streptomyces coelicolor* A3(2), a model streptomycete. *Biotechnology*. 1995;28:65–102. pmid:8688641 doi: 10.1016/b978-0-7506-9095-9.50009-5
29. Wachinger G, Bronnenmeier K, Staudenbauer WL, Schrempf H. Identification of mycelium-associated cellulase from *Streptomyces reticuli*. *Appl Environ Microbiol*. 1989;55(10):2653–7.
30. Saratale GD, Saratale RG, Oh SE. Production and characterization of multiple cellulolytic enzymes by isolated *Streptomyces* sp. *MDS. Biomass Bioenergy*. 2012;47:302–15. doi: 10.1016/j.biombioe.2012.09.030
31. Whiteley M, Bangera MG, Bumgarner RE, Parsek MR, Teitzel GM, Lory S, et al. Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature*. 2001;413(6858):860–4. pmid:11677611 doi: 10.1038/35101627
32. de Boer W, Folman LB, Summerbell RC, Boddy L. Living in a fungal world: impact of fungi on soil bacterial niche development. *FEMS Microbiol Rev*. 2005;29(4):795–811. pmid:16102603 doi: 10.1016/j.femsre.2004.11.005
33. Pullan ST, Chandra G, Bibb MJ, Merrick M. Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes. *BMC genomics*. 2011;12:175. doi: 10.1186/1471-2164-12-175. pmid:21463507
34. Liao CH, Yao L, Xu Y, Liu WB, Zhou Y, Ye BC. Nitrogen regulator GlnR controls uptake and utilization of non-phosphotransferase-system carbon sources in actinomycetes. *Proc Natl Acad Sci U S A*. 2015;112(51):15630–5. doi: 10.1073/pnas.1508465112. pmid:26644570
35. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science*. 2007;315(5813):804–7. pmid:17289988 doi: 10.1126/science.1137016

36. Scott JJ, Oh DC, Yuceer MC, Klepzig KD, Clardy J, Currie CR. Bacterial protection of beetle-fungus mutualism. *Science*. 2008;322(5898):63. doi: 10.1126/science.1160423. pmid:18832638
37. Schlosser A, Schrempf H. A lipid-anchored binding protein is a component of an ATP-dependent cellobiose/cellotriose-transport system from the cellulose degrader *Streptomyces reticuli*. *Eur J Biochem*. 1996;242(2):332–8. pmid:8973652 doi: 10.1111/j.1432-1033.1996.0332r.x
38. Schlosser A, Jantos J, Hackmann K, Schrempf H. Characterization of the binding protein-dependent cellobiose and cellotriose transport system of the cellulose degrader *Streptomyces reticuli*. *Appl Environ Microbiol*. 1999;65(6):2636–43. pmid:10347054 doi: 10.1111/j.1432-1033.1996.0332r.x
39. Marushima K, Ohnishi Y, Horinouchi S. CebR as a master regulator for cellulose/celooligosaccharide catabolism affects morphological development in *Streptomyces griseus*. *J Bacteriol*. 2009;191(19):5930–40. doi: 10.1128/JB.00703-09. pmid:19648249
40. Walter S, Schrempf H. The synthesis of the *Streptomyces reticuli* cellulase (Avicelase) is regulated by both activation and repression mechanisms. *Mol Gen Genet*. 1996;251(2):186–95. pmid:8668129 doi: 10.1007/bf02172917
41. Anderson I, Abt B, Lykidis A, Klenk HP, Kyrpides N, Ivanova N. Genomics of aerobic cellulose utilization systems in Actinobacteria. *PLoS ONE*. 2012;7(6):e39331. doi: 10.1371/journal.pone.0039331
42. Spiridonov NA, Wilson DB. Characterization and cloning of CelR, a transcriptional regulator of cellulase genes from *Thermomonospora fusca*. *J Biol Chem*. 1999;274(19):13127–32. pmid:10224066 doi: 10.1074/jbc.274.19.13127

43. Francis IM, Jourdan S, Fanara S, Loria R, Rigali S. The cellobiose sensor CebR Is the gatekeeper of *Streptomyces scabies* pathogenicity. *Mbio*. 2015;6(2):e02018–14. doi: 10.1128/mBio.02018-14. pmid:25714708
44. Tolonen AC, Chilaka AC, Church GM. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Mol Microbiol*. 2009;74(6):1300–13. doi: 10.1111/j.1365-2958.2009.06890.x. pmid:19775243
45. Beeson WT, Vu VV, Span EA, Phillips CM, Marletta MA. Cellulose degradation by polysaccharide monooxygenases. *Annu Rev Biochem*. 2015;84:923–46. doi: 10.1146/annurev-biochem-060614-034439. pmid:25784051
46. Forsberg Z, Mackenzie AK, Sorlie M, Rohr AK, Helland R, Arvai AS, et al. Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc Natl Acad Sci U S A*. 2014;111(23):8446–51. doi: 10.1073/pnas.1402771111
47. Forsberg Z, Vaaje-Kolstad G, Westereng B, Bunaes AC, Stenstrom Y, MacKenzie A, et al. Cleavage of cellulose by a CBM33 protein. *Protein Sci*. 2011;20(9):1479–83. doi: 10.1002/pro.689. pmid:21748815
48. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu ZL, Zhai H, Sorlie M, et al. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*. 2010;330(6001):219–22. doi: 10.1126/science.1192231. pmid:20929773
49. Book AJ, Yennamalli RM, Takasuka TE, Currie CR, Phillips GN Jr, Fox BG. Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases. *Biotechnol Biofuels*. 2014;7:109. doi: 10.1186/1754-6834-7-109

50. Foreman PK, Brown D, Dankmeyer L, Dean R, Diener S, Dunn-Coleman NS, et al. Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. J Biol Chem. 2003;278(34):31988–97. pmid:12788920 doi: 10.1074/jbc.m304750200
51. Henrissat B. Cellulases and their interaction with cellulose. Cellulose. 1994;1(3):169–96. doi: 10.1007/bf00813506
52. Balows A. The Prokaryotes: A handbook on the biology of bacteria. 2nd ed. New York: Springer-Verlag; 1992.
53. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem. 1959;31(3):426–8. doi: 10.1021/ac60147a030
54. Teymouri F, Laureano-Perez L, Alizadeh H, Dale BE. Ammonia fiber explosion treatment of corn stover. Appl Biochem Biotechnol. 2004;113:951–63. doi: 10.1385/abab:115:1-3:0951
55. Schulein M. Enzymatic properties of cellulases from *Humicola insolens*. J Biotechnol. 1997;57(1–3):71–81. doi: 10.1016/s0168-1656(97)00090-4
56. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013;30(4):772–80. doi: 10.1093/molbev/mst010. pmid:23329690
57. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30(9):1312–3. doi: 10.1093/bioinformatics/btu033
58. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016; 428(4):726–31. doi: 10.1016/j.jmb.2015.11.006

59. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7. pmid:15034147 doi: 10.1093/nar/gkh340
60. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–42. doi: 10.1093/sysbio/sys029. pmid:22357727
61. Chen JQ, Weimer PJ. Competition among three predominant ruminal cellulolytic bacteria in the absence or presence of non-cellulolytic bacteria. *Microbiology.* 2001;147:21–30. doi: 10.1099/00221287-147-1-21
62. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504. pmid:14597658 doi: 10.1101/gr.1239303
63. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34(Web Server issue):W369–73. pmid:16845028 doi: 10.1093/nar/gkl198
64. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics.* 2011;27(7):1017–8. doi: 10.1093/bioinformatics/btr064. pmid:21330290
65. Wu M, Chatterji S, Eisen JA. Accounting for alignment uncertainty in phylogenomics. *PLoS ONE.* 2012;7(1):e30288. doi: 10.1371/journal.pone.0030288. pmid:22272325
66. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 1997;13(5):555–6. pmid:9367129 doi: 10.1093/bioinformatics/13.5.555
67. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725–38. doi: 10.1038/nprot.2010.5. pmid:20360767

68. David LA, Alm EJ. Rapid evolutionary innovation during an Archaeal genetic expansion. *Nature*. 2011;469(7328):93–6. doi: 10.1038/nature09649. pmid:21170026
69. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5(3):e9490. doi: 10.1371/journal.pone.0009490
70. Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipinski A, Kumar S. Estimating divergence times in large molecular phylogenies. *Proc Natl Acad Sci U S A*. 2012;109(47):19333–8. doi: 10.1073/pnas.1213199109. pmid:23129628
71. Sambrook J, Russell DW. The condensed protocols from molecular cloning: a laboratory manual. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2006.
72. Bierman M, Logan R, Obrien K, Seno ET, Rao RN, Schonher BE. Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene*. 1992;116(1):43–9. pmid:1628843 doi: 10.1016/0378-1119(92)90627-2
73. Madduri K, Kennedy J, Rivola G, Inveni-Solari A, Filippini S, Zanuso G, et al. Production of the antitumor drug epirubicin (4'-epidoxorubicin) and its precursor by a genetically engineered strain of *Streptomyces peucetius*. *Nat Biotechnol*. 1998;16(1):69–74. pmid:9447597 doi: 10.1038/nbt0198-69
74. Galm U, Wendt-Pienkowski E, Wang LY, George NP, Oh TJ, Yi F, et al. The biosynthetic gene cluster of zorbamycin, a member of the bleomycin family of antitumor antibiotics, from *Streptomyces flavoviridis* ATCC 21892. *Mol Biosyst*. 2009;5(1):77–90. doi: 10.1039/b814075h
75. Adams AS, Adams SM, Currie CR, Gillette NE, Raffa KF. Geographic variation in bacterial communities associated with the red turpentine beetle (*Coleoptera:Curculionidae*). *Environ Entomol*. 2010;39(2):406–14. doi: 10.1603/EN09221. pmid:20388269

76. Adams AS, Aylward FO, Adams SM, Erbilgin N, Aukema BH, Currie CR, et al. 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*. 2013;79(11):3468–75. doi: 10.1128/aem.00068-13
77. Carr G, Poulsen M, Klassen JL, Hou YP, Wyche TP, Bugni TS, et al. Microtermolides A and B from termite-associated *Streptomyces* sp. and structural revision of vinylamycin. *Org Lett*. 2012;14(11):2822–5. doi: 10.1021/ol301043p. pmid:22591554
78. Grubbs KJ. Microbial communities associated with ambrosia beetles and the European honey bee, *Apis mellifera*. Ph.D. Thesis, University of Wisconsin—Madison. 2013. Available: <https://search.library.wisc.edu/catalog/9910160552302121>
79. Hanshew AS. Microbial ecology of tropical arboreal ants. Ph.D. Thesis, University of Wisconsin—Madison. 2013. Available: <https://search.library.wisc.edu/catalog/9910190787102121>
80. Poulsen M, Oh DC, Clardy J, Currie CR. Chemical analyses of wasp-associated *Streptomyces* bacteria reveal a prolific potential for natural products discovery. *PLoS ONE*. 2011;6(2):e16763. doi: 10.1371/journal.pone.0016763
81. Schwab L. Investigation of endosymbiotic Actinobacteria occurring in higher fungi and medicinal prairie plants. M.Sc. Thesis, University of Wisconsin—Madison. 2011.
82. Visser AA, Nobre T, Currie CR, Aanen DK, Poulsen M. Exploring the potential for Actinobacteria as defensive symbionts in fungus-growing termites. *Microb Ecol*. 2012;63(4):975–85. doi: 10.1007/s00248-011-9987-4. pmid:22173371



## **Chapter 4: Cellulose-Enriched Microbial Communities from Leaf-Cutter Ant (*Atta colombica*) Refuse Dumps Vary in Taxonomic Composition and Degradation Ability**

**Previously published:** Gina R. Lewin, Amanda L. Johnson, Rolando D. Moreira-Soto, Kailene Perry, Adam J. Book, Heidi A. Horn, Adrián A. Pinto-Tomás, and Cameron R. Currie (2016).

Cellulose-Enriched Microbial Communities from Leaf-Cutter Ant (*Atta colombica*) Refuse Dumps Vary in Taxonomic Composition and Degradation Ability. *PLoS ONE*, 11(3): e0151840.

**Co-authors contributed in the following ways:** Conceived and designed the experiments:

GRL, ALJ, RDMS, AJB, AAPT, CRC. Performed experiments and analyzed data for Figure 4.2: RDMS; Figure 4.3: ALJ, GRL, HAH; Figure 4.4: GRL; Figure 4.5: GRL, KP; Figure 4.6: GRL; Table 4.1: GRL, ALJ; Table 4.2: GRL. Wrote the paper: GRL, RDMS, CRC.

### **4.1 Abstract**

Deconstruction of the cellulose in plant cell walls is critical for carbon flow through ecosystems and for the production of sustainable cellulosic biofuels. Our understanding of cellulose deconstruction is largely limited to the study of microbes in isolation, but in nature, this process is driven by microbes within complex communities. In Neotropical forests, microbes in leaf-cutter ant refuse dumps are important for carbon turnover. These dumps consist of decaying plant material and a diverse bacterial community, as shown here by electron microscopy. To study the portion of the community capable of cellulose degradation, we performed enrichments on cellulose using material from five *Atta colombica* refuse dumps. The ability of enriched communities to degrade cellulose varied significantly across refuse dumps. 16S rRNA gene amplicon sequencing of enriched samples identified that the community structure correlated with refuse dump and with degradation ability. Overall, samples were dominated by Bacteroidetes,

Gammaproteobacteria, and Betaproteobacteria. Half of abundant operational taxonomic units (OTUs) across samples were classified within genera containing known cellulose degraders, including *Acidovorax*, the most abundant OTU detected across samples, which was positively correlated with cellulolytic ability. A representative *Acidovorax* strain was isolated, but did not grow on cellulose alone. Phenotypic and compositional analyses of enrichment cultures, such as those presented here, help link community composition with cellulolytic ability and provide insight into the complexity of community-based cellulose degradation.

## 4.2 Introduction

The complex polysaccharides stored in the plant cell wall are the most abundant source of organic carbon in terrestrial ecosystems [1]. Select lineages of bacteria and fungi have evolved the ability to enzymatically deconstruct the primary component of plant cell walls, cellulose, a crystal of  $\beta$ -1,4-linked glucose molecules [1,2]. These microbes are critical for driving the terrestrial carbon cycle. Furthermore, they are a valuable resource to identify cellulase enzymes for the sustainable, economical production of cellulosic biofuels [3].

The enzymes used to break down cellulose are well characterized for a small set of microbial isolates [4–9]. However, in natural systems, organisms degrade plant biomass within communities [10–14]. Interactions between species influence cellulose degradation [4,15,16], but the complexity of natural systems hinders a full understanding of how plant biomass break down is altered by microbial interactions and by the underlying diversity of communities. Enrichments of environmental samples with either cellulose or plant biomass as the sole carbon source are an effective method to select for the portion of microbial communities capable of plant biomass

degradation [10,17–19]. Critically, this method preserves the community interactions necessary for cellulose break down, allowing for the analysis of cellulolytic organisms within a community and for insight into the ecology of cellulolytic communities as a whole.

In Central and South American tropical forests and savannahs, leaf-cutter ants and their symbiotic microbes are dominant herbivores and therefore important for carbon cycling. An individual mature leaf-cutter ant colony harvests hundreds of kilograms of leaf material per year (Figure 4.1a) [20]. In tropical savannahs, these ants are estimated to harvest as much as 13-17% of total leaves produced [21]. Harvested leaves are partially decomposed in subterranean chambers by a mutualistic fungus that the ants cultivate as their food source [22]. However, the fungal cultivar only degrades ~50% of the total leaf material including only 30% of the cellulose in the leaves [23]. Ants move the remaining, cellulose-enriched leaf material to refuse dumps



**Figure 4.1. Leaf-cutter ants concentrate decaying plant matter in refuse dumps. (a)** Leaf-cutter ants are dominant herbivores in Central and South American rain forests. An *Atta* worker carries a leaf fragment back to her nest in Costa Rica. **(b)** The leaf structure is still visible in the dump material tended by this *Atta* worker in the Currie lab at University of Wisconsin-Madison. **(c)** A vertically cross-sectioned *A. colombica* refuse dump in Costa Rica. **Photo credits:** Don Parsons (a,b), Gina Lewin (c)

(Figures 4.1b and 4.1c). These dumps function as compost piles where complex microbial communities drive the degradation of the cellulose-rich recalcitrant plant material [24,25].

Mature colonies of the leaf-cutter ant species *Atta colombica* maintain a large, aboveground refuse dump downslope from the main nest [20,26]. A single colony of up to 2 million worker ants can dispose of more than 100 refuse particles per minute [27], totaling to over one hundred kilograms (wet weight) of refuse material per year [20]. As material accumulates, the refuse dump can grow to be over one meter in height and two meters in diameter, acquiring vertical stratification with the freshest material in the top strata and the oldest, most recalcitrant material in the bottom strata [26]. Similar to human-produced compost piles, there are high levels of metabolic activity within refuse dumps as microbes degrade the cellulose and other recalcitrant material deposited by the ants [26,28]. The concentration of cellulose quickly decreases from 110  $\mu\text{g/ml}$  in the bottom of the fungus garden to 43  $\mu\text{g/ml}$  in the top strata of the dump to 30  $\mu\text{g/ml}$  in the lower strata of the dump [23,25]. Refuse dumps are also enriched for nitrogen, phosphorus, and other nutrients compared with the nutrient-poor tropical soil [26,28]. However, microorganisms deplete these nutrient levels to those of surrounding soil within one year of a colony dying or moving locations [28].

Culture-independent work has demonstrated that there is a highly diverse but unique community of microbes within refuse dumps dominated by Proteobacteria, Actinobacteria, and Bacteroidetes [24,25]. The microbial community structure of refuse dumps is generally similar among colonies. However, community succession occurs between the upper, middle, and lower strata of the dump piles, reflecting differences in abiotic properties such as oxygen levels (more anaerobic

in lower layers) and biomass composition (more recalcitrant in lower layers) [25]. Additionally, the refuse dump community differs from the *Enterobacteriaceae*-dominated community in fungus gardens [24,29] and from the Acidobacteria- and Proteobacteria-dominated community found in the nutrient-depleted tropical forest soil [30,31]. Overall, the refuse dump microbial community is well defined by these analyses, but its complexity has prevented a clear understanding of the organisms that may contribute to cellulose degradation.

Here, we analyzed microbial communities in leaf-cutter ant refuse dumps to study plant biomass degradation within a community context. We used electron microscopy to observe the degradation of plant cells and microbial communities in the leaf-cutter ant refuse dump. Then, to analyze simple communities capable of cellulose degradation, we performed enrichments on cellulosic filter paper and measured the ability of microbial communities to degrade cellulose across three layers of five *A. colombica* refuse dumps. We identified the microbial community composition in a range of samples using 16S rRNA gene amplicon sequencing. The alpha diversity of each sample and the beta diversity between samples were calculated to identify patterns in the community structure that correlated with degradation ability, ant colony, or refuse dump layer. Additionally, we isolated the most abundant community member detected across samples, an *Acidovorax* sp., and analyzed its ability to degrade cellulose. Through these methods, we identified community members that are important for cellulose degradation, and we improved our understanding of the interplay between microbial community composition and the ability to degrade cellulose.

### 4.3 Materials and Methods

#### 4.3.1 Sample Collection

All samples used for this study are covered by the Resolution Number 009, from the Comisión Institucional de Biodiversidad of the University of Costa Rica, and no protected species were sampled in this study. Samples for electron microscopy were aseptically collected from the top of *A. colombica* refuse dumps in May 2010 and in April 2011 with permission on private land in La Palma, Osa, Costa Rica. Dump fragments were immersed in Karnovsky fixative (2.5% glutaraldehyde and 2% paraformaldehyde in 0.1 M phosphate buffer, pH 7.4) and kept at 4°C for transportation to the Center of Research in Microscopic Structures at the University of Costa Rica.

For enrichments, the top, middle, and bottom layers of five *A. colombica* colonies were aseptically collected in July 2012 on protected land at Carara National Park, Costa Rica. These collections were permitted by Resolution ACOPAC-INV-006-10 from the Área de Conservación Pacífico Central, Sistema de Áreas de Conservación of the Ministerio del Ambiente, Energía y Telecomunicaciones. *A. colombica* dumps were cross-sectioned vertically, with the material on the exterior of the dump collected as “top”, material in the middle third designated as “middle”, and material in the bottom third labelled as “bottom” (Figure A4.1a). Dumps were all within 500 m of each other. All material was stored at 4°C.

### **4.3.2 Electron Microscopy**

To analyze the ultrastructure of *A. colombica* refuse dumps and look for the presence of microorganisms, we used scanning and transmission electron microscopy (SEM and TEM, respectively) techniques (n=2 for SEM and n=2 for TEM). Samples in Karnovsky fixative were left at 4°C overnight. Samples were post-fixed in 1% osmium tetroxide for at least 1 hour and dehydrated with ethanol (SEM; 30%, 50%, 70%, 80%, 90%, 95%, 100%) or acetone (TEM; 30%, 50%, 70%, 90%, 100%). SEM samples were dried by sublimation in a freeze dryer (VFD-20, VD Inc.) after immersion in tert-butanol. Dry samples were mounted on aluminum stubs, coated in gold with an Ion coater (IB-3, Giko), and examined on Hitachi S-570, S-2360N, and S-3700N electron microscopes. TEM samples were infiltrated with epoxy resin (Spurr) after dehydration and sectioned with a PT-PC PowerTome ultramicrotome (RMC products). Sections (70 nm) were stained with uranyl acetate (4% in 50% ethanol) and 2% Sato's Triple lead and examined on Hitachi H-7000 and H-7100 transmission electron microscopes.

### **4.3.3 Overview of Enrichment Design**

From each layer of each refuse dump, six pieces of approximately 3 mg (~2 mm diameter) of refuse dump material were inoculated into individual test tubes containing 5 mL of media and a 1x10 cm strip of Whatman #1 filter paper pressed against the side of the tube as the sole carbon source (Figure A4.1b). M63 minimal medium was used, containing in 1 L: 61.5 mM potassium phosphate dibasic (Acros, Geel, Belgium), 38.5 mM potassium phosphate monobasic (Acros, Geel, Belgium), 15.1 mM ammonium sulfate (Gibco, Grand Island, NY), 0.5 mL of an iron solution (1 mg/ml iron sulfate in 0.01 M HCl), 1 mL of 1M magnesium sulfate solution, 1 mL of 1 mg/ml thiamine solution (Acros, Geel, Belgium), and 5 mL of SPV-4 trace elements solution

[32]. These filter paper test tubes are a useful tool for determining the ability of a microbial community to grow on cellulose; aerobic cellulolytic communities grow directly on the filter paper and eventually break it into two pieces at the air/liquid interface. Additionally, the minimal media conditions were chosen to limit nutrients available to the microbes, requiring the degradation of cellulose for carbon and encouraging cross-feeding between community members. Samples were grown at 30°C (to replicate the temperature of the refuse dumps when samples were collected), shaking at 250 rpm. After 14 days, samples were vortexed, and 200  $\mu$ L was transferred into two sets of tubes: (1) three qualitative tubes each containing a strip of filter paper to determine how many days were necessary to break the filter paper and (2) three quantitative tubes with pre-weighed, submerged filter paper to compare the percentage of cellulose degraded across samples (Figure A4.1c). As detailed below, this combination of qualitative and quantitative cultures allowed us to analyze the community composition and the cellulolytic ability of the enrichment communities, while ensuring that the high concentrations of nutrients from the original dump material did not influence our data.

Qualitative tubes contained a strip of 1x10 cm filter paper in M63 minimal media. Cultures were grown at 30°C, 250 rpm and were checked daily for visible signs of growth on the filter paper or break down of the filter paper. Three days after the filter paper broke in half, samples were vortexed, and 1.5 mL samples were collected for future DNA extraction and amplicon sequencing (Figure A4.1d). This three day time point was chosen to provide a representative sample of the cellulose-degrading community, while allowing for enough biomass for DNA extraction and sequencing. For communities where the filter paper did not break, samples for DNA extraction were collected after 14 days. DNA samples were centrifuged in a benchtop



centrifuge at 16,100 x g for 10 min, the supernatant was removed, and the cell pellets were frozen at -20°C.

Quantitative tubes contained two 1x4 cm strips of pre-weighed filter paper and 8 mL of M63 media. Controls contained filter paper with no inoculum. Cultures were grown shaking at 250 rpm at 30°C for 10 days. A single time point was used for quantitative tube sampling to allow for comparison across all samples; ten days provided of a wide range of degradation values. The percentage of cellulose degraded was measured using a previously published acid detergent-based method [33,34].

Samples were named based on the refuse dump (colony 1-5), followed by the layer, followed by the inoculation replicate (A-F). For example, inoculation A from the top layer of dump 2 is “2 Top A”. No data were collected for samples 4 Middle A and 4 Bottom C because of technical difficulties.

#### **4.3.4 DNA Extraction**

DNA was extracted using the PowerSoil® DNA Isolation Kit (MoBio, Carlsbad, CA) with the following modifications. PowerBeads were added to the microcentrifuge tube with the thawed cell pellet, mixed, and then transferred back into PowerBead tube. After addition of solution C1, tubes were incubated at 70°C for 10 min. Then, instead of vortexing for lysis, tubes went through three rounds of bead beating for 2 min (Mini-Beadbeater-96, Biospec Products, Bartlesville, OK) then freezing at -80°C for 2.5 min.

#### **4.3.5 16S rRNA Gene Amplicon Sequencing**

Extracted DNA was PCR amplified in triplicate. Each 25  $\mu$ L reaction contained 12.75  $\mu$ L sterile UltraPure Distilled Water (Invitrogen, Grand Island, NY), 0.5  $\mu$ L DMSO at a final concentration of 0.28 M, 0.25  $\mu$ L Herculase II DNA polymerase (Agilent Technologies, Santa Clara, CA), 5  $\mu$ L buffer, 1 mM dNPTs, 50 ng DNA, and 0.625  $\mu$ L each of 10  $\mu$ M bacterial specific primers 926F (5'-AAACTYAAAKGAATTGACGG-3') and 1392R (5'-ACGGGCGGTGTGTRC-3') to a final concentration of 0.25  $\mu$ M with standard 454-Titanium adapters and multiplex identifiers for Lib-L (Roche 454 Sequencing, Madison, WI). PCR reaction conditions included 2 min at 95°C; 30 cycles of 95°C for 20 s, 66.3°C for 20 s, 72°C for 30 s; and a final 3 min elongation at 72°C. PCR products were verified on a 0.8% agarose gel. Pooled triplicates were run on a 2% low-melt agarose gel and cleaned up using a Zymoclean™ Gel DNA Recovery Kit (Zymo Research, Irvine, CA) followed by three rounds of Agencourt AMPure XP beads (Beckman Coulter, Brea, CA) using manufacturer's protocols. Samples were quantified using a Qubit® Fluorometer (Life Technologies, Grand Island, NY) and pooled to 10<sup>6</sup> DNA molecules/ $\mu$ L. DNA was sequenced on a GS Junior with FLX Titanium chemistry using previously published long-read modifications [35].

#### **4.3.6 Amplicon Sequence Processing**

Raw data were analyzed using the following steps in mothur v.1.33.3 [36]. Flowgrams were removed that did not contain an exact match to the barcode or primer using the command trim.flows. Then, the command shhh.flows was used to denoise sequences using the flowgrams. Sequences were trimmed to a minimum length of 200 bp using default settings. We aligned unique sequences to the Silva 16S rRNA gene sequence database, version 102, using the default

kmer-based search methods [37]. Reads that did not align over the region of interest were removed, and we ran filter.seqs command with “trump=.” to remove excess alignment columns. We removed chimeras identified using UCHIME [38]. Taxonomy was assigned using a mothur-formatted version of the Ribosomal Database Project (RDP) taxonomy training set 9 [39] with a cut-off of 60% identity. OTUs were clustered at 97% identity using the dist.seqs and cluster commands. OTUs were named based on their total abundance across samples. For example, OTU1 was the most abundant OTU overall. Amplicon sequencing data were deposited under Sequence Read Archive accession number SRP059774.

#### ***4.3.7 Statistical Analyses***

Comparisons of the qualitative degradation data across dumps and layers were performed in JMP (SAS Institute, Cary, NC) using a survival analysis and the Wilcoxon Group Homogeneity Test. Comparisons of the quantitative degradation data across dumps and layers were performed in JMP using analysis of variance (ANOVA) followed by the Tukey-Kramer honest significant difference (HSD) test.

All sequencing analyses were performed using data rarified to 1701 reads and OTUs classified at 97% similarity. We tested for correlations between the number of reads of an OTU in our sequencing data and the percentage of cellulose degraded by each sample (quantitative data) using ANOVA in JMP. Alpha diversity metrics were calculated in mothur using the summary.single command. We tested for differences in alpha diversity between colonies and dump layers using the Tukey-Kramer HSD test in JMP. Mothur was used to cluster samples with the Morisita-Horn Index using the tree.shared command. A parsimony test to analyze the

significance of the cladogram topology with colony, layer, or degradation ability (a categorical ranking based off the level of degradation by the sample) also was performed in *mothur*. We next analyzed the distance between samples using both the Morisita-Horn Index (*dist.shared* command) and a weighted Unifrac analysis (*unifrac.weighted* command) in *mothur*. The *pcoa* command in *mothur* was used to display these distance matrices with a principle coordinate analysis (PCoA). We tested if samples in the PCoAs clustered by colony, layer, or degradation level using an analysis of molecular variance (AMOVA) in *mothur*. Finally, the *corr.axes* command was used in *mothur* to determine the Pearson correlation of the coordinates of the PCoA with the percentage of cellulose degradation and with each OTU. For this correlation analyses, negative values of cellulose degradation relative to the control (within error of zero) were represented as 0%.

#### ***4.3.8 Isolation of dominant community members***

Highly cellulolytic communities 1 Top A, 1 Middle E, 1 Bottom E, and 3 Bottom C were dilution plated on the following media types to try to isolate dominant members: M63 cellulose agar: M63 medium (see above) with 15.0 g agar (Amresco, Solon, OH) in 1 L, a 4x4 cm piece of filter paper laid on top of the plate as the sole carbon source; AO Agar: 0.5 g Bacto™ yeast extract (BD, Franklin Lakes, NJ), 0.2 g Bacto™ beef extract (BD, Franklin Lakes, NJ), 0.5 g tryptone (Fisher Scientific, Pittsburg, PA), 2.4 mM sodium acetate (Fisher Scientific, Pittsburg, PA), and 9.0 g agar in 1 L [40]; R2A Agar: 0.5 g Bacto™ yeast extract, 0.5 g proteose peptone #3 (Remel, San Diego, CA), 0.5 g casamino acids (Fisher Scientific, Pittsburg, PA), 0.5 g D-glucose (Research Products International Corp., Mount Prospect, IL), 0.5 g Difco™ soluble starch (BD, Franklin Lakes, NJ), 0.3 g sodium pyruvate (Alfa Aesar, Haverhill, MA), 0.3 g

potassium phosphate dibasic, 0.05 g magnesium sulfate, and 15.0 g agar in 1 L [41]; AGS media: 1.0 g arginine monohydrochloride (Sigma-Aldrich, St. Louis, MO), 12.5 g glycerol (Fisher Scientific, Pittsburg, PA), 1.0 g potassium phosphate dibasic, 1.0 g sodium chloride (Fisher Scientific, Pittsburg, PA), 0.5 g magnesium sulfate heptahydrate (Sigma-Aldrich, St. Louis, MO), 0.02 g iron(III) sulfate hexahydrate (Mallinckrodt Chemical Works, St. Louis, MO), 0.001 g copper(II) sulfate pentahydrate (Fisher Scientific, Pittsburg, PA), 0.001 g zinc sulfate heptahydrate (Fisher Scientific, Pittsburg, PA), 0.001 g manganese(II) sulfate monohydrate (Fisher Scientific, Pittsburg, PA), and 15.0 g agar in 1L [42]; Yeast Malt Extract Agar: 4.0 g Bacto™ yeast extract; 10.0 g Bacto™ malt extract (BD, Franklin Lakes, NJ); 4.0 g D-glucose; and 15.0 g agar in 1 L.

Isolated strains were identified through sequencing their 16S rRNA gene using the following procedure with general bacterial primers. A colony of cells was added to 20 µL of microLYSIS®-PLUS (Gel Company, San Francisco, CA), then lysed using the thermal cycler profile: 65°C for 15 min, 96°C for 2 min, 65°C for 4 min, 96°C for 1 min, 65°C for 1 min, and 96°C for 30 s. Two microliters of this lysis solution were mixed with 8.5 µL of water, 1 µL each of 10 µM general bacterial 16S rRNA gene primers 27F (5'-AGAGTTTGATCMTGGCTC-3') and 1492R (5'-TACGGYTACCTTGTTACGACTT-3') at a final concentration of 0.4 µM [43], and 12.5 µL of EconoTaq Plus Green 2X Master Mix (Lucigen, Middleton, WI). This mixture was run in a thermal cycler using the parameters: 95°C for 5 min; 30 cycles of 95°C for 1 min, 55°C for 1 min, and 72°C for 2 min; and a final extension of 72°C for 10 min. After verifying amplification using DNA gel electrophoresis with a 0.8% agarose gel, fluorescent dyes were incorporated using reactions consisting of 1 µL of BigDye polymerase (University of Wisconsin-

Madison Biotech Center), 1.5 µL of BigDye Buffer, 0.5 µL of primer (either 27F or 1492R) to a final concentration of 0.5 µM, 6.5 µL of water, and 0.5 µL of amplified DNA. Amplification conditions were 95°C for 3 min; 35 cycles of 96°C for 10 s and 58°C for 3 min; and 72°C for 7 min. Samples were sequenced using Sanger sequencing at the University of Wisconsin-Madison Biotech Center (Madison, WI) and analyzed using SeqMan Pro in the DNASTAR Lasergene 11 suite (Madison, WI). The 16S rRNA gene sequence was matched to the RDP database using the options: Type, Isolate,  $\geq 1200$  bp, good quality [44].

#### **4.3.9 *Acidovorax* Taxonomy and Growth Assay and Identification**

To compare the taxonomy of our *Acidovorax* isolate to the organisms comprising OTU1 in the communities, all 10,803 sequences that clustered into OTU1 in our amplicon data were extracted using the bin.seqs command in mothur and clustered at 100% identity using CD-HIT [45–47]. Clusters containing three or less reads were removed. Representative sequences from each of the seven remaining clusters were aligned in MEGA6 (Muscle, default parameters) [48] with the 16S rRNA gene sequences for all *Acidovorax* type strains and the outgroup *Variovorax paradoxus* from RDP [39]. A maximum likelihood tree was built from the trimmed alignment using RAxML-HPC2 on XSEDE through Cipres with rapid bootstrapping [49,50]. The 16S rRNA gene sequence for this strain, named *Acidovorax* sp. AcolKP-3D, was deposited in GenBank, accession number KT150251.

To test for growth on cellulose, the *Acidovorax* isolate was grown in triplicate in test tubes containing 5 mL of liquid M63 minimal media and supplemented with either 27.8 mM D-glucose, 27.8 mM cellobiose (Sigma-Aldrich, St. Louis, MO), a 1x10 cm strip of filter paper,

0.025 g phosphoric acid swollen cellulose (PASC [51]), or 0.025g crystalline cellulose (Sigmacell, Sigma-Aldrich, St. Louis, MO) as the carbon source. Additionally, the isolate was grown in triplicate in LB broth (Lennox L, Research Products International Corp., Mount Prospect, IL). All cultures were grown at 30°C, shaking at 300 rpm and observed daily for fourteen days. To confirm visual growth observations, 200 µL of each culture was plated onto LB agar plates (LB broth with 15 g agar) after fourteen days. Plates were incubated at 30°C for two days.

## **4.4 Results**

### ***4.4.1 Electron Microscopy of Leaf-Cutter Ant Refuse Dumps***

SEM imaging of leaf material from the top layer of a refuse dump demonstrated that the stoma of the plant and the waxy cuticle remained intact (Figures 4.2a and 4.2b, representative images). However, analyses by TEM indicated that the plant cells walls were degraded or partially degraded and the internal structure of the cells was abnormal (red boxes, Figure 4.2c, representative image). Numerous rod-shaped bacteria and cocci were visible on the plant material and surrounding the plant cells. Interestingly, multiple small clusters of cells were visible, and different morphologies of bacteria appeared to be in close proximity of each other (Figure 4.2).

### ***4.4.2 Differences in Cellulolytic Ability across Refuse Dumps***

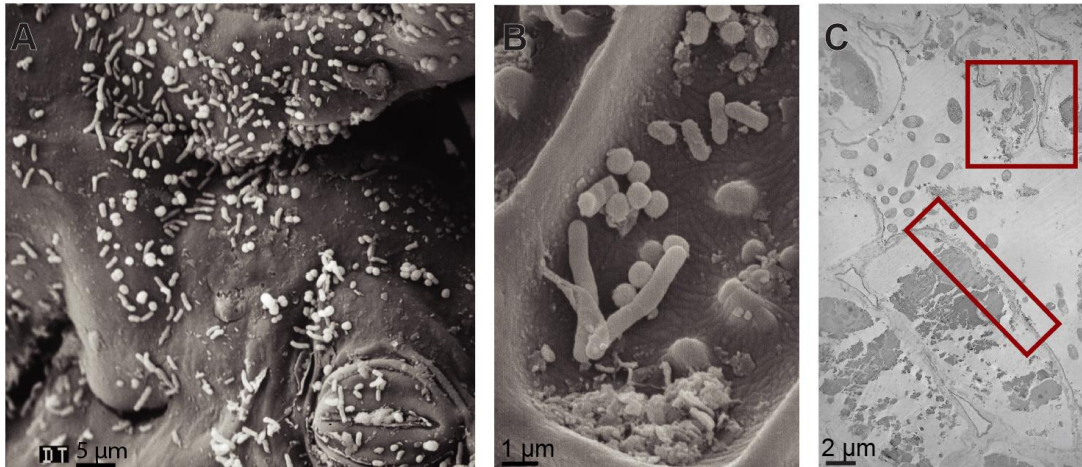


**Figure 4.2. Electron microscopy of leaf-cutter ant refuse dumps. (a and b)** Scanning electron microscopy shows the ultrastructure of refuse dump leaf material and different bacterial morphologies. **(c)** Transmission electron microscopy shows leaf cells and surrounding bacteria. Red boxes indicate degraded plant cell wall and abnormal, clumped internal cell structure. **Photo credits:** Rolando Moreira Soto

Enrichments of refuse dump material on cellulosic filter paper selected for the portion of the refuse dump microbial communities able to grow on and degrade cellulose as the sole carbon source (Figures A4.1 and A4.2 Figs; Table A4.1). The four fastest microbial communities broke the filter paper in two days (“qualitative assay”; Figure 4.3a and 4.3b). In contrast, 26 of the 88 communities tested did not break the filter paper in the 14-day experiment, although 20 of these 26 communities did show visible signs of growth on the filter paper. There were significant differences in the time to degradation when samples were grouped by the ant colony they originated from (Wilcoxon Group Homogeneity Test,  $p < 0.0001$ ; Figure 4.3b). All microbial communities from colony 5 degraded the filter paper within 8 days, with an average of  $4.6 \pm 0.3$  days (SE). In contrast, only 4/16 (22%) of the microbial communities from colony 4 degraded the filter paper over the 14-day experiment. Communities from colonies 1, 2, and 3 fell between these two extremes, with 67%, 83%, and 72% degrading the cellulose within 14 days, respectively. In contrast, there were no significant differences in degradation when samples were



grouped based on whether they were collected from the top, middle, or bottom layer of the dump (Wilcoxon Group Homogeneity Test,  $p = 0.8765$ , Figure 4.3b).

We quantified the percentage of cellulose degraded in 10 days of growth (“quantitative assay”), and the average cellulose degradation was  $35.1 \pm 3.1\%$  (Standard Error [SE], Figures 4.3c and

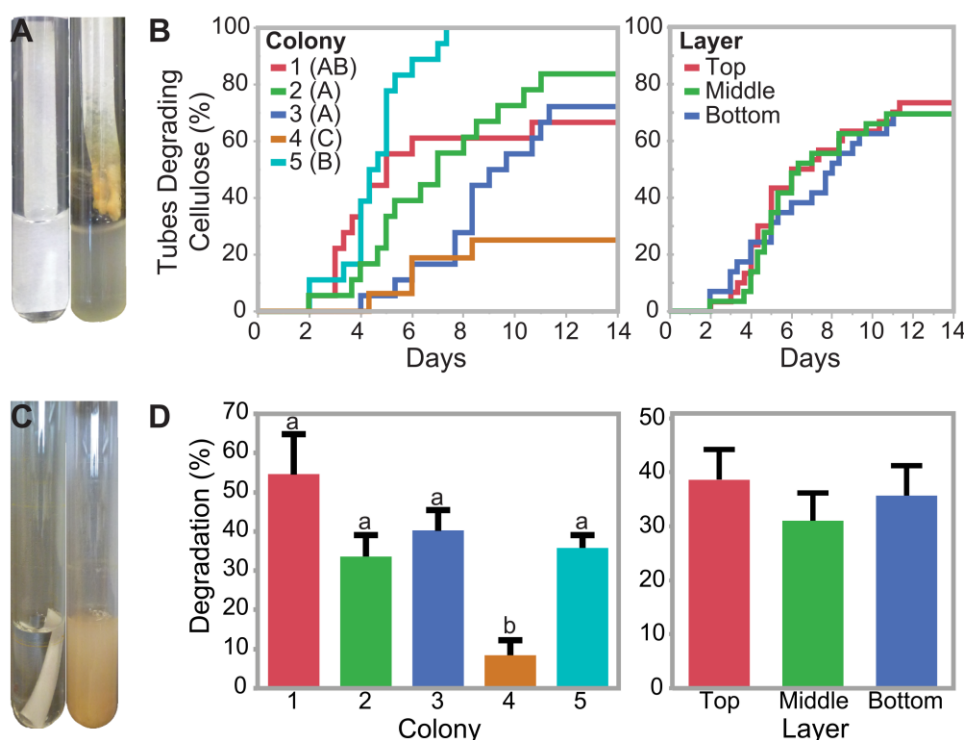


**Figure 4.3. Comparison of degradation ability across colonies and layers. (a and b) Qualitative Assay Data.**

Test tubes containing carbon-free minimal media and a strip of cellulosic filter paper were used to enrich for cellulolytic communities. Failure plots, indicating when the filter paper broke apart in each culture, were fit with Kaplan Meier curves and analyzed using the Wilcoxon method to determine significant differences among colonies (indicated by letters A-C) and layers (no significant differences). **(c and d) Quantitative Assay Data.** Pre-weighed, submerged cellulosic filter paper allowed quantification of cellulose degradation after 10 days. Samples are grouped by colony or dump layer. Error bars represent one standard error from the mean.

Significant differences were determined using Tukey's HSD test and are indicated above the data. **Photo**

**credits:** Gina Lewin (a, c)

4.3d; Table A4.1). In the most cellulolytic sample, nearly all cellulose was degraded after 10 days (1 Middle D;  $98.6 \pm 1.8\%$ ), and 5 of the 88 samples degraded over 95% of the cellulose. In contrast, 16 of the 88 samples degraded less than five percent of detectable cellulose.

Quantitative analyses supported the qualitative assays in demonstrating that the microbial communities from colony 4 were significantly less cellulolytic than communities from the other four refuse dumps ( $8.4 \pm 3.8\%$  degradation [SE]; Tukey-Kramer HSD test,  $p < 0.01$  for all).

However, there were no significant differences in the average percentage of cellulose degraded between colony 1, 2, 3, and 5 ( $54.6 \pm 10.2\%$ ,  $33.6 \pm 5.5\%$ ,  $40.3 \pm 5.2\%$ , and  $35.8 \pm 3.3\%$  degradation, respectively; SE). We also did not measure a significant difference between enrichments from the different strata of the dump (ANOVA,  $F = 0.49$ ,  $df = 3$  and  $85$ ,  $p = 0.6120$ , Figure 4.3d).

#### ***4.4.3 Amplicon Sequencing Analysis***

To understand how microbial community structure differed among enrichments from different dumps and different degradation abilities, we sequenced the V6-8 region of the bacterial 16S rRNA gene from a subset of enriched samples representing each layer of each dump and a range of degradation abilities (Table 4.1). We were not able to amplify any fungal sequences from our enrichments using universal fungal primers for internal transcribed spacer sequences (A. Johnson, unpublished data). After sequence processing, there were 104,044 total sequences and 1409 unique sequences across all samples. Sequences per sample ranged from 1701 to 9616 reads (Table 4.1). To standardize, we subsampled each sample to 1701 reads for all analyses.

**Table 4.1.** Degradation, sequencing depth, and alpha diversity metrics for sequenced samples at a 97% OTU definition.

| Sample <sup>a</sup> | Cellulose Degradation |                    |                                  | Number of Reads | Good's Coverage <sup>c</sup> | Observed OTUs <sup>c</sup> | Chao1 Estimator <sup>c</sup> | Inverse Simpson's Metric <sup>c</sup> | Berger Parker Index <sup>c</sup> |
|---------------------|-----------------------|--------------------|----------------------------------|-----------------|------------------------------|----------------------------|------------------------------|---------------------------------------|----------------------------------|
|                     | Average               | Standard Deviation | Degradation Ranking <sup>b</sup> |                 |                              |                            |                              |                                       |                                  |
| 1 Top A             | 88.3%                 | 1.5%               | High                             | 3478            | 99.1%                        | 55                         | 68                           | 2.8                                   | 0.57                             |
| 1 Top C             | 95.0%                 | 2.2%               | High                             | 4586            | 99.0%                        | 49                         | 68                           | 3.7                                   | 0.38                             |
| 1 Middle C          | -3.9%                 | 1.1%               | Low                              | 4045            | 98.6%                        | 70                         | 102                          | 7.0                                   | 0.30                             |
| 1 Middle E          | 98.6%                 | 1.8%               | High                             | 4183            | 99.1%                        | 39                         | 64                           | 5.1                                   | 0.37                             |
| 1 Bottom D          | 4.1%                  | 2.0%               | Low                              | 3586            | 98.9%                        | 54                         | 77                           | 6.8                                   | 0.31                             |
| 1 Bottom E          | 92.9%                 | 3.6%               | High                             | 5380            | 99.1%                        | 40                         | 61                           | 5.0                                   | 0.37                             |
| 2 Top A             | 22.8%                 | 8.4%               | Medium                           | 4543            | 99.7%                        | 22                         | 29                           | 3.5                                   | 0.47                             |
| 2 Top B             | 15.8%                 | 1.8%               | Medium                           | 4389            | 99.6%                        | 27                         | 38                           | 5.7                                   | 0.26                             |
| 2 Middle B          | 67.1%                 | 5.2%               | High                             | 3768            | 99.3%                        | 50                         | 65                           | 7.4                                   | 0.30                             |
| 2 Middle D          | 64.2%                 | 4.4%               | High                             | 2964            | 99.4%                        | 35                         | 43                           | 3.2                                   | 0.53                             |
| 2 Bottom A          | 29.6%                 | 2.6%               | Medium                           | 4436            | 99.0%                        | 50                         | 71                           | 5.5                                   | 0.34                             |
| 2 Bottom D          | 1.6%                  | 2.1%               | Low                              | 2354            | 98.6%                        | 63                         | 84                           | 6.3                                   | 0.24                             |
| 3 Top A             | 24.1%                 | 2.6%               | Low                              | 1702            | 98.8%                        | 78                         | 94                           | 14.3                                  | 0.18                             |
| 3 Top F             | 17.8%                 | 1.7%               | Low                              | 4191            | 98.9%                        | 60                         | 85                           | 12.3                                  | 0.16                             |
| 3 Middle A          | 41.0%                 | 8.0%               | Medium                           | 3485            | 98.7%                        | 73                         | 93                           | 5.4                                   | 0.40                             |
| 3 Middle E          | 21.6%                 | 5.0%               | Low                              | 4750            | 98.8%                        | 60                         | 88                           | 10.0                                  | 0.24                             |
| 3 Bottom B          | 56.7%                 | 0.0%               | Medium                           | 5499            | 98.3%                        | 81                         | 117                          | 12.9                                  | 0.21                             |
| 3 Bottom C          | 90.6%                 | 5.2%               | High                             | 9616            | 98.7%                        | 63                         | 94                           | 5.8                                   | 0.34                             |
| 3 Bottom D          | 56.6%                 | 3.6%               | Medium                           | 3704            | 98.7%                        | 66                         | 92                           | 8.7                                   | 0.26                             |
| 3 Bottom F          | 20.5%                 | 6.3%               | Medium                           | 1701            | 98.9%                        | 60                         | 89                           | 7.2                                   | 0.31                             |
| 4 Top C             | 49.8%                 | 1.4%               | Low                              | 4401            | 99.1%                        | 54                         | 72                           | 8.4                                   | 0.25                             |
| 4 Middle D          | -5.9%                 | 0.6%               | Low                              | 4321            | 98.7%                        | 64                         | 89                           | 2.9                                   | 0.57                             |
| 4 Bottom E          | 3.1%                  | 2.4%               | Low                              | 3253            | 99.1%                        | 56                         | 79                           | 6.3                                   | 0.37                             |
| 5 Top A             | 24.6%                 | 1.6%               | Medium                           | 3241            | 98.4%                        | 91                         | 125                          | 14.6                                  | 0.17                             |
| 5 Middle E          | 36.0%                 | 2.9%               | Medium                           | 3709            | 98.4%                        | 96                         | 129                          | 12.3                                  | 0.18                             |
| 5 Bottom C          | 52.7%                 | 3.3%               | High                             | 2759            | 99.1%                        | 57                         | 73                           | 4.4                                   | 0.45                             |

<sup>a</sup>Sample name formatting refers to colony sampled, layer of dump, and inoculation replicate.<sup>b</sup>Used for analyses that require a categorical variable.<sup>c</sup>Metrics calculated using data subsampled to 1701 reads.

Overall, samples were dominated by Bacteroidetes, Gammaproteobacteria, and Betaproteobacteria ( $30.7 \pm 3.5\%$ ,  $16.3 \pm 2.8\%$ , and  $15.7 \pm 3.5\%$  of total reads, respectively [SE]; Figure 4.4; Dataset A4.1). The most abundant operational taxonomic units (OTUs) across all samples were classified as *Acidovorax*, *Leadbetterella*, *Flavobacteriaceae*, *Dokdonella*, and

TM7 (Table 4.2). Twenty of the 30 most abundant OTUs were identified to the genus level, and species from ten of these twenty genera have previously been shown to degrade cellulose (Table 4.2). We compared the number of reads of the 50 most abundant OTUs to the percentage of cellulose degraded by each sample; the abundances of OTU1 (*Acidovorax*;  $r^2 = 0.55$ ; ANOVA,  $F = 29.38$ ,  $df = 1$  and  $25$ ,  $p < 0.0001$ ) and OTU9 (*Ferruginibacter*;  $r^2 = 0.27$ ; ANOVA,  $F = 9.12$ ,  $df = 1$  and  $25$ ,  $p = 0.0059$ ) were significantly correlated with the percentage of cellulose degraded by the communities (Figure 4.5).

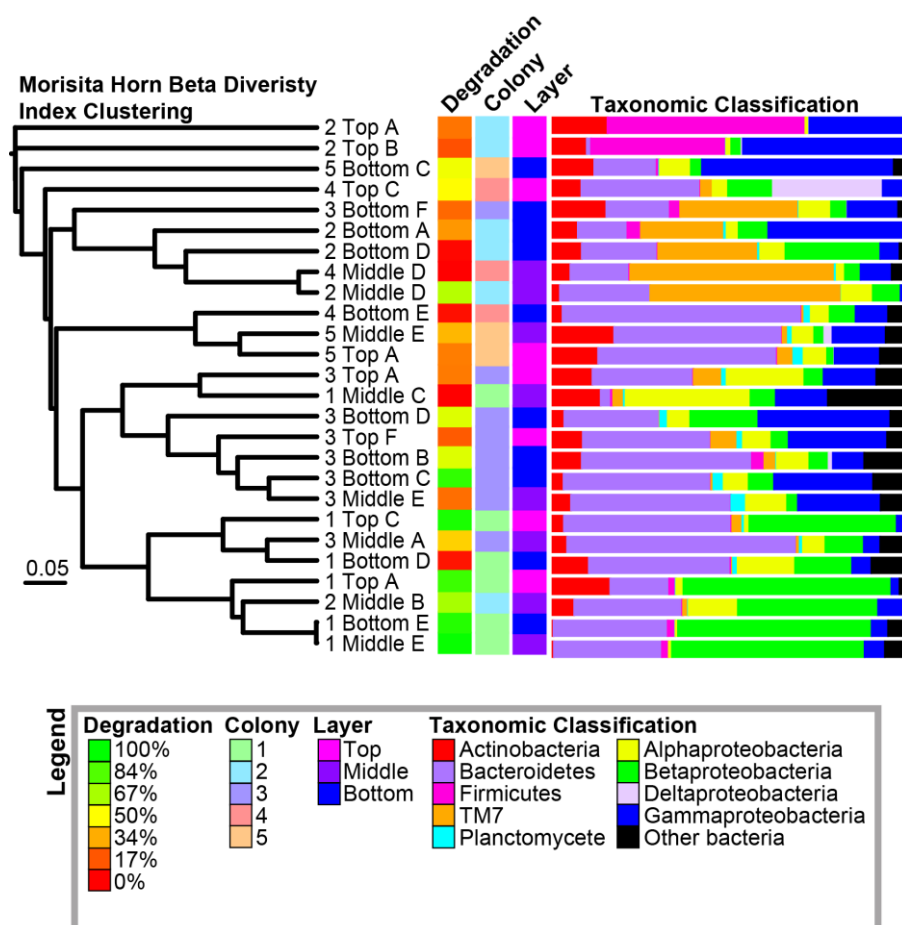


**Figure 4.4. Morisita-Horn Beta Diversity Clustering of Samples.** The corresponding percentage of cellulose degradation, colony, layer, and taxonomic classification of OTUs are shown for each sample.

**Table 4.2.** Taxonomy of top 30 OTUs across all samples and reports of cellulose degradation.<sup>a</sup>

| OTU   | Number of Reads | Taxonomy <sup>b</sup> |                          |                          |                           |                               | Report in Literature of Cellulose Degradation | Reference of Degradation |
|-------|-----------------|-----------------------|--------------------------|--------------------------|---------------------------|-------------------------------|-----------------------------------------------|--------------------------|
|       |                 | Phylum                | Class                    | Order                    | Family                    | Genus                         |                                               |                          |
| OTU1  | 10803           | Proteobacteria        | Betaproteobacteria       | Burkholderiales          | Comamonadaceae            | <i>Acidovorax</i>             | Yes                                           | [52,53]                  |
| OTU2  | 8536            | Bacteroidetes         | Sphingobacteria          | Sphingobacteriales       | Cytophagaceae             | <i>Leadbetterella</i>         | No                                            |                          |
| OTU3  | 7917            | Bacteroidetes         | Flavobacteria            | Flavobacteriales         | Flavobacteriaceae         | unclassified                  | nd <sup>c</sup>                               |                          |
| OTU4  | 6971            | Proteobacteria        | Gammaproteobacteria      | Xanthomonadales          | Xanthomonadaceae          | <i>Dokdonella</i>             | No                                            | [54]                     |
| OTU5  | 6392            | TM7                   | TM7 class incertae sedis | TM7 order incertae sedis | TM7 family incertae sedis | TM7 genus incertae sedis      | nd                                            |                          |
| OTU6  | 6200            | Actinobacteria        | Actinobacteria           | Actinomycetales          | Microbacteriaceae (78)    | unclassified                  | nd                                            |                          |
| OTU7  | 4587            | Proteobacteria        | Alphaproteobacteria      | Caulobacteriales         | Caulobacteraceae          | <i>Asticcacaulis</i> (77)     | Yes                                           | [55,56]                  |
| OTU8  | 3167            | Proteobacteria        | Gammaproteobacteria      | Xanthomonadales          | Xanthomonadaceae          | <i>Pseudoxanthomonas</i> (99) | No                                            |                          |
| OTU9  | 3106            | Bacteroidetes         | Sphingobacteria          | Sphingobacteriales       | Chitinophagaceae (99)     | <i>Ferruginibacter</i> (77)   | No                                            |                          |
| OTU10 | 3039            | Proteobacteria        | unclassified             | unclassified             | unclassified              | unclassified                  | nd                                            | [57,58]                  |
| OTU11 | 2974            | Bacteroidetes         | Sphingobacteria          | Sphingobacteriales       | Cytophagaceae             | <i>Sporocytophaga</i>         | Yes                                           |                          |
| OTU12 | 2875            | Proteobacteria        | Betaproteobacteria       | Burkholderiales          | Alcaligenaceae            | <i>Castellaniella</i>         | No                                            |                          |
| OTU13 | 2719            | Proteobacteria        | Gammaproteobacteria      | Xanthomonadales          | Xanthomonadaceae          | <i>Dyella</i> (64)            | Yes                                           | [59]                     |
| OTU14 | 2290            | Firmicutes            | Bacilli                  | Bacillales               | Paenibacillaceae 1        | <i>Cohnella</i>               | Yes                                           |                          |
| OTU15 | 2228            | Proteobacteria        | Gammaproteobacteria      | Xanthomonadales          | Xanthomonadaceae          | unclassified                  | nd                                            |                          |
| OTU16 | 1707            | Bacteroidetes         | Sphingobacteria          | Sphingobacteriales       | Saprospiraceae            | <i>Haliscomenobacter</i>      | No                                            | [60]                     |
| OTU17 | 1270            | Proteobacteria        | Deltaproteobacteria      | Myxococcales             | Polyangiaceae             | <i>Sorangium</i>              | Yes                                           |                          |
| OTU18 | 1083            | Proteobacteria        | Gammaproteobacteria      | Xanthomonadales          | Xanthomonadaceae          | <i>Stenotrophomonas</i> (91)  | Yes                                           |                          |
| OTU19 | 1027            | TM7                   | TM7 class incertae sedis | TM7 order incertae sedis | TM7 family incertae sedis | TM7 genus incertae sedis      | nd                                            | [61]                     |
| OTU20 | 1005            | Proteobacteria        | Betaproteobacteria       | Burkholderiales          | Comamonadaceae            | unclassified                  | nd                                            |                          |
| OTU21 | 881             | Firmicutes            | Bacilli                  | Bacillales               | Paenibacillaceae 1        | <i>Paenibacillus</i>          | Yes                                           |                          |
| OTU22 | 867             | Proteobacteria        | Betaproteobacteria       | Burkholderiales          | Alcaligenaceae            | <i>Achromobacter</i> (65)     | No                                            | [62,63]                  |
| OTU23 | 850             | Deinococcus-Thermus   | Deinococci               | Deinococcales            | Trueperaceae              | <i>Truepera</i>               | No                                            |                          |
| OTU24 | 813             | Verrucomicrobia       | Opitutae                 | Opitutales               | Opitutaceae               | <i>Opitutus</i>               | Yes                                           |                          |
| OTU25 | 792             | Proteobacteria        | Alphaproteobacteria      | Rhizobiales              | Bradyrhizobiaceae         | unclassified (74)             | nd                                            | [64]                     |
| OTU26 | 792             | Firmicutes            | Bacilli                  | Bacillales               | Paenibacillaceae 1        | <i>Paenibacillus</i>          | Yes                                           |                          |
| OTU27 | 720             | unclassified          | unclassified             | unclassified             | unclassified              | unclassified                  | nd                                            |                          |
| OTU28 | 647             | Bacteroidetes         | Sphingobacteria          | Sphingobacteriales       | Sphingobacteriaceae       | unclassified                  | nd                                            | [62,63]                  |
| OTU29 | 626             | Planctomycetes        | Planctomycetacia         | Planctomycetales         | Planctomycetaceae         | <i>Singulisphaera</i> (91)    | No                                            |                          |
| OTU30 | 601             | Bacteroidetes         | Flavobacteria            | Flavobacteriales         | Flavobacteriaceae         | <i>Ornithobacterium</i> (69)  | No                                            |                          |

<sup>a</sup>See Dataset A4.1 for full taxonomy.<sup>b</sup>Number in parentheses indicates confidence score of taxonomic assignment. No number is shown if confidence is 100.<sup>c</sup>nd = Not determined because taxonomy not identified to genus level

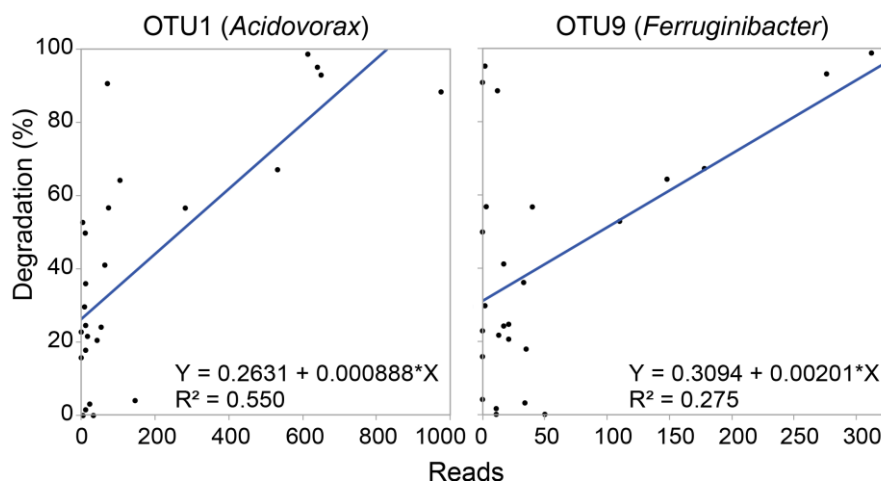


**Figure 4.5. Correlation between number of reads and cellulose degradation by the community for OTU1 (*Acidovorax*) and OTU9 (*Ferruginibacter*) across sequenced samples.**

#### 4.4.4 Alpha Diversity of Enriched Communities

All samples had high coverage based on Good's estimator (average =  $98.9 \pm 0.07\%$  [SE]; range = 98.3% - 99.7%; probability that an additional sequence obtained would already be represented in the dataset). However, the Chao1 richness estimator indicated that 17% to 39% of the OTUs in each sample were not identified by our sequencing (average =  $27.5 \pm 1.0\%$  [SE]; Table 4.1).

There were slight but significant differences in these diversity indices among some colonies. For example, the samples from colony 2 had significantly lower observed OTUs than those of colonies 3 and 5 (Tukey-Kramer HSD Test,  $p = 0.0102$  and  $p = 0.0109$ , respectively), and the samples from colony 2 had significantly lower predicted OTUs (Chao1 richness estimator) than those from colony 5 (Tukey-Kramer HSD Test,  $p = 0.043$ ). There were no significant differences in diversity metrics among dump layers. Also, there was no correlation between the percentage of cellulose degradation and the observed number of OTUs, the estimated number of OTUs (Chao1 richness estimator), the Inverse Simpson's Diversity Index, or the Berger Parker Index (dominance measure; Table 4.1). However, while medium and low degradation samples

contained a large range of diversities (Inverse Simpson's Diversity Index range = 3.5 - 14.6, average =  $8.4 \pm 1.3$  [SE] and range = 2.9 - 14.3, average =  $8.4 \pm 1.3$  [SE], respectively), high degradation samples only had low diversity (range = 2.8 - 7.4, average =  $4.7 \pm 0.5$  [SE]; Figure A4.3, Table 4.1).

#### ***4.4.5 Beta Diversity of Enriched Communities***

We clustered samples based on their similarity in community structure using the Morisita-Horn Index, which compares the overlap among samples based on the abundance of each OTU and the total number of OTUs (Figure 4.4) [65]. Parsimony analysis of the resulting cladogram indicated that the samples clustered significantly based on their source colony ( $p = 0.005$ ), but not based on their degradation level (high, medium, or low, see Table 4.1) or layer. Additionally, we mapped the major lineages (phyla or subphyla) found in each sample onto the cladogram (Figure 4.4). The classification of lineages in each sample matched the clustering patterns of the samples. For example, samples with a high proportion of Betaproteobacteria or TM7 generally clustered together.

In the cladogram, there are samples that cluster but have vastly different cellulolytic abilities (ex. 1 Top C and 1 Bottom D; 4 Middle D and 2 Middle D; 3 Middle E and 3 Bottom D) (Figure 4.4). To provide insight into relationship between OTU abundance and cellulolytic ability, we performed comparisons on each of these pairs of samples, graphing the abundance of each OTU in the highly cellulolytic sample and the non-cellulolytic sample (Figure A4.4). This analysis indicated that in each pair of samples, at least one OTU was identified at high abundance in both

communities, likely driving their clustering. However, most of the non-dominant OTUs identified in one sample were absent in the other sample.

To further analyze the factors that correlated with community diversity, we visualized the Morisita-Horn distance matrix using a PCoA (Figure 4.6). The resulting matrix significantly clustered by colony (AMOVA,  $df = 4$ ,  $F = 2.73$ ,  $p < 0.001$ ) and degradation level (AMOVA,  $df = 2$ ,  $F = 2.39$ ,  $p < 0.001$ ; Figure 4.6a). Specifically, highly cellulolytic communities significantly clustered separately from communities with medium or low levels of degradation (AMOVA,  $df = 1$ ,  $F = 2.94$ ,  $p = 0.008$  and AMOVA,  $df = 1$ ,  $F = 3.39$ ,  $p = 0.007$ , respectively). However, there was no significant clustering based on layer ( $p = 0.44$ ; Figure 4.6b). We confirmed these patterns by measuring beta diversity using a weighted Unifrac analysis, which quantitatively groups samples by the similarity of their phylogenetic structure (Figure A4.5) [66]. Samples clustered significantly when grouped by colony (AMOVA,  $df = 25$ ,  $F = 2.4$ ,  $p < 0.001$ ), as in the Morisita-Horn based clustering. Additionally, when a Bonferroni correction was applied for multiple comparisons, high-degrading samples were significantly different from the low-degrading samples (AMOVA,  $df = 16$ ,  $F = 2.33$ ,  $p = 0.014$ ), but not from medium-degrading samples ( $df = 16$ ,  $F = 2.37$ ,  $p = 0.019$ ).

#### **4.4.6 Correlation Analysis**

We determined the Pearson correlation of the principal coordinates of the PCoA plot with each OTU and with the percentage of cellulose degradation (Figure 4.6c, Dataset A4.2). The percentage of cellulose degradation correlated positively with coordinate two ( $p = 0.016$ ). The abundances of OTU1 (*Acidovorax*) and OTU4 (*Dokdonella*) also correlated positively with



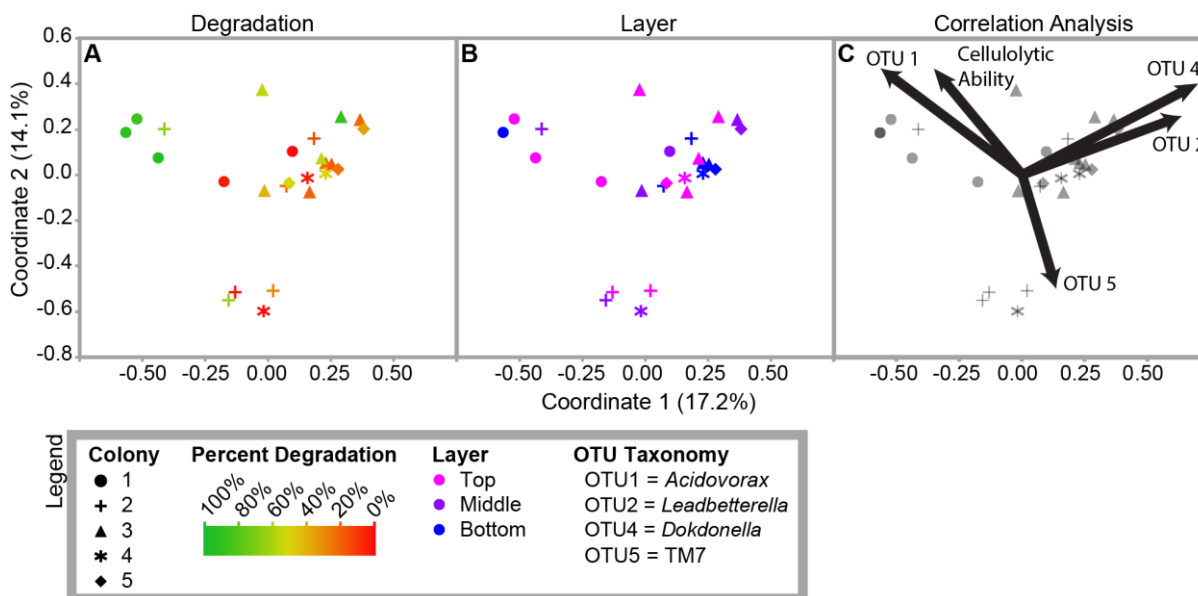


**Figure 4.6. PCoA clustering of Morisita-Horn Diversity Index.** Sample shape indicates colony. Sample color indicates degradation (A) or layer (B). Panel C shows the correlation analysis. The vectors indicate the correlation of each OTU and the percentage of cellulose degradation with the principal coordinates shown.

coordinate two ( $p = 0.015$  and  $p = 0.04$ , respectively). In contrast, OTU5's (TM7) abundances significantly correlated negatively with coordinate two ( $p = 0.009$ ). Additionally, OTU1's abundances significantly correlated negatively with coordinate one ( $p = 0.0028$ ), while OTU4 and OTU2 (*Leadbetterella*) correlated positively with coordinate one ( $p = 0.000066$  and  $p = 0.00044$ , respectively). Neither OTU3 (unclassified *Flavobacteriaceae*) or any of the other 30-most abundant OTUs significantly correlated with either coordinate one or two (Figure A4.3b).

#### 4.4.7 *Acidovorax*

To understand the role of dominant microbes in the enrichment communities, we isolated 20 strains from highly cellulolytic enrichments on a range of selective and rich media (Table A4.2). Of these isolates, the only strain that represented a dominant community member was *Acidovorax* strain AcolKP-3D from the 1 Top A enrichment community, which was isolated on

AO agar. Using the Ribosomal Database Project's seqmatch program, the almost full-length 16S rRNA gene sequence for our strain matched best to *Acidovorax caeni* R-24608, with 90.6% of unique 7-base oligomers shared between the two strains (S\_ab score; [67]). Additionally, the 16S rRNA gene sequence of AcolKP-3D was identical to the majority (7391/10803; 68%) of the sequences that comprise the *Acidovorax* OTU1 (Figure A4.6).

Because *Acidovorax* sequences were highly detected in our samples and correlated with cellulolytic ability (Table 4.2, Figure 4.5, Dataset A4.2), we hypothesized that *Acidovorax* contributes to cellulase production in the community. To test if strain AcolKP-3D could degrade cellulose in isolation, we grew it in the same liquid minimal media as the enrichments with cellobiose, filter paper, PASC, or crystalline cellulose as the sole carbon source, but no growth was observed in any of these cultures after two weeks (Figure A4.6). However, the strain showed growth with glucose as the sole carbon source after 4 days and in rich LB liquid media after 1 day (Figure A4.6). All growth observations were confirmed using plating onto LB.

## 4.5 Discussion

Exploring how microbial communities break down plant biomass has important implications for carbon cycling, climate change, and bioenergy research. However, analyses of natural cellulose-degrading communities have proven to be challenging due to their high levels of microbial diversity and complexity. By employing enrichment techniques, we linked community membership and diversity with degradation ability in simplified communities from *A. colombica* leaf-cutter ant refuse dumps. Our enrichment strategy allowed us to select for the portion of the community that contributes to cellulose degradation while preserving as many critical inter-

species interactions as possible. We observed a significant correlation between the microbial community structure of enriched communities from leaf-cutter ant refuse dumps and their cellulolytic ability. Additionally, community structure correlated significantly with colony of origin.

SEM and TEM images indicated that communities of diverse rod-shaped and cocci bacteria are present on refuse dump material (Figure 4.2). Additionally, the mixture of intact and degraded plant tissue supports the results of Moreira-Soto et al. [68], who showed that plant cell degradation and the abundance of bacteria increase in the refuse dump relative to the fungus garden. Based on the proximity of the microbes in the images, it is possible that they are interacting, either positively or negatively, as they degrade the plant material.

Our enrichments successfully reduced the diversity of the refuse dump communities while allowing for interactions between organisms. Native refuse dump material has a high level of microbial diversity (average Inverse Simpson's Diversity Index = 56.6) [25], but in our enriched samples, the average diversity was eight times lower (7.2; Table 4.1). Community simplification has also been observed in other enrichment-based studies on plant biomass components [10,69]. Additionally, as the communities simplified, certain OTUs began to dominate, as indicated by the increased Berger-Parker Index (average  $d = 0.16$  in native refuse dump material; average  $d = 0.33$  in enriched samples; Table 4.1) [25]. This decrease in diversity and increase in dominance allowed us to start to understand the ecology of cellulose degradation in communities isolated from refuse dumps.

Our results suggest that there may be a negative correlation between diversity and cellulolytic ability within our samples since high cellulolytic ability was only recorded in low-diversity samples (Table 4.1, Figure A4.3). However, these differences were not significant because low cellulolytic ability was found across a range of high- and low-diversity samples. The correlation between diversity and community function has been a topic of much debate, both in bacteria and eukaryotes [70–72]. In our simplified enrichments, we propose two explanations for the pattern found between diversity and cellulose degradation. Potentially, high cellulolytic ability in low diversity samples was the result of an increased abundance of a few key organisms. This idea is supported by the correlations between the abundances of *Acidovorax* (OTU1) and *Ferruginibacter* (OTU9) with cellulolytic ability (Figure 4.5). Alternatively, in high diversity samples, negative interactions between organisms could have decreased the level of degradation. Non-cellulolytic organisms may compete with cellulolytic microbes for nutrients including the small oligosaccharides that are released as cellulose is degraded extracellularly. This competition would decrease the growth rates of the cellulolytic microbes and therefore the overall extent of cellulose breakdown.

The taxonomic composition of our enrichments was unique compared with native leaf-cutter ant refuse dumps or fungus gardens and compared with previous enrichment experiments. Different families of Gammaproteobacteria are abundant in the leaf-cutter ant fungus garden (*Enterobacteriaceae*) than in these enrichment communities (*Xanthomonadaceae*) [29]. Our communities also show little overlap in abundant OTUs with other plant biomass degradation enrichment experiments [10,17–19], likely because of differences in the carbon sources, growth conditions, and inoculums across these experiments. Similar to native refuse dumps, the majority

of our enrichments were dominated by Bacterioidetes, and dominant families including *Comamonadaceae*, *Flavobacteriaceae*, and *Xanthomonadaceae* are abundant in both enriched and native refuse dumps [25]. However, at the genus level, abundant OTUs in these enrichments were detected at very low levels in native refuse dumps, with the exception of *Paenibacillus* [25]. This difference emphasizes that cellulose is not the only carbon source in the refuse dumps, and therefore sequencing of complex natural environments cannot always identify the portion of the community responsible for cellulose degradation.

In this study, there were no significant differences in cellulolytic ability when we grouped samples by layer. In contrast, previous culture-independent analyses of leaf-cutter ant refuse dumps showed that the microbial communities have a predictable pattern of succession between layers [25]. It is possible that the factors that led to the community shift between layers in natural refuse dumps do not vary in our enrichment cultures. For example, in native refuse dumps, there are more aerobes in the upper strata and more anaerobes in lower strata. Since our enrichments were aerobic, we could not detect these differences. Additionally, since cellulose is present in all layers of the dump, cellulolytic microbes may not vary across strata.

We did, however, observe differences in cellulolytic ability and microbial community structure between refuse dumps. While the microbial communities from some dumps could degrade almost all detectable cellulose in 10 days, the communities from other dumps rarely showed signs of cellulose break down (Figure 4.3). Furthermore, there were differences in microbial community structure among refuse dumps (Figures 4.4 and 4.6). Possible factors driving the differences among refuse dumps could include the age of the colony, variation in substrate input,

or the temperature, pH, or moisture level in the refuse dump. Also, some refuse dumps may be dominated by cellulolytic fungi instead of bacteria, and since we did not detect any fungi in the enrichments, their contributions would not be represented. Because of the large importance of leaf-cutter ant refuse dumps for carbon turnover, it would be useful to understand if these differences in degradation abilities are steady over time and ecologically relevant.

Interestingly, our results identified multiple communities that were highly similar based on beta diversity metrics but had large differences in cellulolytic ability (Figures 4.4 and 4.6, Figure A4.4). This observation is counter to the view that similar phenotypic activity correlates with similar community structure. Our analyses indicated that these communities are dominated by a small number of OTUs that are highly abundant in both samples (Figure A4.4). Therefore, we predict that the OTU definitions at 97% identity are not always specific enough to differentiate between cellulolytic and non-cellulolytic organisms.

The results from our 16S rRNA gene sequencing suggest that *Acidovorax* spp. may be important for the ability of our enriched communities to degrade cellulose (Table 4.2, Figure 4.5), but an *Acidovorax* isolate was not able to grow on cellulose in isolation. Although the strain grew overnight in rich media, it took four days to show growth in glucose media. Furthermore, it did not show signs of growth on cellobiose, filter paper, crystalline cellulose, or the less recalcitrant PASC in the liquid minimal medium used for the enrichments (Figure A4.6). These results suggest that *Acidovorax* relies on other member(s) of the community to grow in the enrichments. Possibly, *Acidovorax* is a secondary consumer or scavenger. Alternatively, *Acidovorax* may

receive essential nutrients or stimuli for growth and cellulase production from other members of the community.

Many of the other abundant OTUs in our samples were classified in genera that contain known cellulose-degrading strains (Table 4.2). We hypothesize that these organisms contribute to cellulose degradation in our communities. Although the ability of one species in a genus to degrade cellulose does not mean that other closely-related strains are cellulolytic, cellulases are generally conserved phylogenetically [73]. Therefore, this finding indicates that cellulolytic microbes are enriched for in our experimental samples. Furthermore, the ability to degrade cellulose is a rare trait [1], so it is notable that half of the abundant OTUs were classified in genera that can degrade cellulose. Interestingly, the potentially cellulolytic OTUs were not from well-studied groups of cellulose-degrading microbes. The abundance of poorly studied bacteria in our samples emphasizes the need to explore the diversity of cellulose-degrading microbes that are functionally important in leaf-cutter ant refuse dumps and other natural ecosystems.

A better understanding of the diversity of microbes that contribute to cellulose degradation in the environment is critical for both analyzing the microbial contribution to carbon turnover and for identifying novel enzymes capable of breaking down plant material to produce sustainable cellulosic biofuels. This study shows that microbial communities and their abilities to degrade cellulose can vary significantly, even in seemingly similar environments such as leaf-cutter ant refuse dumps. Therefore, it is necessary to understand the complexities of communities to predict the rate of cellulose degradation and carbon turnover in any specific environment. Cellulose degradation by a small number of model organisms has been extensively studied in laboratory

environments, but our understanding of plant biomass decomposition by microbial communities in nature is still limited. This study is an important step in linking the ability to degrade cellulose with the membership and diversity of microbial communities from the leaf-cutter ant refuse dump, and detailed analysis of refuse dumps can serve as a model of community-driven plant biomass degradation in other nutrient-rich environments.

#### 4.6 References

1. Wilson D. Microbial diversity of cellulose hydrolysis. *Curr Opin Microbiol.* 2011;14: 259–263. doi:10.1016/j.mib.2011.04.004
2. Pauly M, Keegstra K. Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J.* 2008;54: 559–68. doi:10.1111/j.1365-313X.2008.03463.x
3. Rubin EM. Genomics of cellulosic biofuels. *Nature.* 2008;454: 841–5. doi:10.1038/nature07190
4. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002;66: 506–577. doi:10.1128/MMBR.66.3.506
5. Brumm PJ. Bacterial genomes: what they teach us about cellulose degradation. *Biofuels.* 2013;4: 669–681. doi:10.4155/bfs.13.44
6. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva I a., et al. Revealing nature’s cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science.* 2013;342: 1513–1516. doi:10.1126/science.1244273
7. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome



- sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol. 2008;26: 553–60. doi:10.1038/nbt1403
8. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. Sci Rep. 2013;3: 1–10. doi:10.1038/srep01030
  9. Riederer A, Takasuka TE, Makino SI, Stevenson DM, Bukhman YV, Elsen NL, et al. Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. Appl Environ Microbiol. 2011;77: 1243–1253. doi:10.1128/AEM.02008-10
  10. DeAngelis KM, Gladden JM, Allgaier M, D’haeseleer P, Fortney JL, Reddy A, et al. Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. BioEnergy Res. 2010;3: 146–158. doi:10.1007/s12155-010-9089-z
  11. Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, et al. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science. 2011;331: 463–7. doi:10.1126/science.1200387
  12. Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A. 2006;103: 626–31. doi:10.1073/pnas.0507535103
  13. Pope PB, Denman SE, Jones M, Tringe SG, Barry K, Malfatti SA, et al. Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores. Proc Natl Acad Sci. 2010;107: 14793–8. doi:10.1073/pnas.1005297107

14. Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, et al. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc Natl Acad Sci.* 2014;111: 14500–5. doi:10.1073/pnas.1319718111
15. Zhou Y, Pope PB, Li S, Wen B, Tan F, Cheng S, et al. Omics-based interpretation of synergism in a soil-derived cellulose-degrading microbial community. *Sci Rep.* 2014;4: 1–6. doi:10.1038/srep05288
16. Kato S, Haruta S, Cui ZJ, Ishii M, Igarashi Y. Network relationships of bacteria in a stable mixed culture. *Microb Ecol.* 2008;56: 403–11. doi:10.1007/s00248-007-9357-4
17. Gladden JM, Allgaier M, Miller CS, Hazen TC, VanderGheynst JS, Hugenholtz P, et al. Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass. *Appl Environ Microbiol.* 2011;77: 5804–12. doi:10.1128/AEM.00032-11
18. DeAngelis KM, Fortney JL, Borglin S, Silver WL, Simmons BA, Hazen TC. Anaerobic decomposition of switchgrass by tropical soil-derived feedstock-adapted consortia. *MBio.* 2012;3: e00249–11. doi:10.1128/mBio.00249-11
19. Eichorst SA, Varanasi P, Stavila V, Zemla M, Auer M, Singh S, et al. Community dynamics of cellulose-adapted thermophilic bacterial consortia. *Environ Microbiol.* 2013;15: 2573-87. doi:10.1111/1462-2920.12159
20. Hart AG, Ratnieks FLW, Sheffield S. Waste management in the leaf-cutting ant *Atta colombica*. *Behav Ecol.* 2002;13: 224–231.
21. Costa AN, Vasconcelos HL, Vieira-Neto EHM, Bruna EM. Do herbivores exert top-down effects in Neotropical savannas? Estimates of biomass consumption by leaf-cutter ants. *J Veg Sci.* 2008;19: 849–854. doi:10.3170/2008-8-18461

22. Weber NA. Fungus-Growing Ants. *Science*. 1966;153: 587–604.
23. Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás A, et al. An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet*. 2010;6. doi:10.1371/journal.pgen.1001129
24. Scott JJ, Budsberg KJ, Suen G, Wixon DL, Balser TC, Currie CR. Microbial community structure of leaf-cutter ant fungus gardens and refuse dumps. *PLoS One*. 2010;5: e9922. doi:10.1371/journal.pone.0009922
25. Scott JJ. Microbial community ecology of leaf-cutter ant refuse dumps. 2011.
26. Haines BL. Element and Energy Flows through Colonies of the Leaf-Cutting Ant, *Atta colombica*, in Panama. *Biotropica*. 1978;10: 270–277.
27. Herz H, Beyschlag W, Hölldobler B. Assessing herbivory rates of leaf-cutting ant (*Atta colombica*) colonies through short-term refuse deposition counts. *Biotropica*. 2007;39: 476–481. doi:10.1111/j.1744-7429.2007.00283.x
28. Hudson TM, Turner BL, Herz H, Robinson JS. Temporal patterns of nutrient availability around nests of leaf-cutting ants (*Atta colombica*) in secondary moist tropical forest. *Soil Biol Biochem*. 2009;41: 1088–1093. doi:10.1016/j.soilbio.2009.02.014
29. Aylward FO, Burnum KE, Scott JJ, Suen G, Tringe SG, Adams SM, et al. Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME J*. 2012;6 1688–701. doi:10.1038/ismej.2012.10
30. da C Jesus E, Marsh TL, Tiedje JM, de S Moreira FM. Changes in land use alter the structure of bacterial communities in Western Amazon soils. *ISME J*. 2009;3: 1004–1011.

doi:10.1038/ismej.2009.98

31. Mendes LW, Tsai SM, Navarrete AA, de Hollander M, van Veen JA, Kuramae EE. Soil-Borne Microbiome: Linking Diversity to Function. *Microb Ecol.* 2015;70: 255–65.  
doi:10.1007/s00248-014-0559-2
32. Voelskow H. Extrazelluläre mikrobielle Polysaccharide: Polymere mit interessanten Anwendungsmöglichkeiten. *Forum Mikrobiol.* 1988;6: 273–283.
33. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, et al. Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically-linked capacity for the degradation of lignocellulose. *Appl Environ Microbiol.* 2014;80: 4692–4701. doi:10.1128/AEM.01133-14
34. Weimer PJ, Lopez-Guisa JM, French AD. Effect of cellulose fine structure on kinetics of its digestion by mixed ruminal microorganisms in vitro. *Appl Environ Microbiol.* 1990;56: 2421–9.
35. Hanshew AS, Mason CJ, Raffa KF, Currie CR. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *J Microbiol Methods.* 2013;95: 149–55. doi:10.1016/j.mimet.2013.08.007
36. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75: 7537–41. doi:10.1128/AEM.01541-09
37. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence

- data compatible with ARB. *Nucleic Acids Res.* 2007;35: 7188–7196.  
doi:10.1093/nar/gkm864
38. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* 2011;27: 2194–2200.  
doi:10.1093/bioinformatics/btr381
39. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 2009;37: 141–145. doi:10.1093/nar/gkn879
40. Reichenbach H. The Order Cytophagales. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. *Prokaryotes*. 3rd ed. New York: Springer; 2006. pp. 7:549–590. doi:10.1007/0-387-30747-8
41. Reasoner DJ, Geldreich EE. A new medium for the enumeration and subculture of bacteria from potable water. *Appl Environ Microbiol.* 1985;49: 1–7.
42. El-Nakeeb MA, Lechevalier HA. Selective isolation of aerobic Actinomycetes. *Appl Microbiol.* 1963;11: 75–77.
43. Lane D. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. *Nucleic acid techniques in bacterial systematics*. New York: Wiley; 1991. pp. 115–175.
44. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73: 5261–7. doi:10.1128/AEM.00062-07
45. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein

- or nucleotide sequences. *Bioinformatics*. 2006;22: 1658–1659.  
doi:10.1093/bioinformatics/btl158
46. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28: 3150–3152. doi:10.1093/bioinformatics/bts565
  47. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26: 680–682.  
doi:10.1093/bioinformatics/btq003
  48. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30: 2725–2729.  
doi:10.1093/molbev/mst197
  49. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Gatew Comput Environ Work*, 2010. 2010.  
doi:10.1109/GCE.2010.5676129
  50. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–1313.  
doi:10.1093/bioinformatics/btu033
  51. Schülein M. Enzymatic properties of cellulases from *Humicola insolens*. *J Biotechnol*. 1997;57: 71–81. doi:10.1016/S0168-1656(97)00090-4
  52. Willems A, Gillis M. Genus II. *Acidovorax*. In: Garrity G, Boone DR, Castenholz RW, editors. *Bergey's Manual of Determinative Bacteriology*. 2nd ed. New York: Springer-Verlag; 2001. pp. 696–704.

53. Choi J-H, Kim M-S, Roh SW, Bae J-W. *Acidovorax soli* sp. nov., isolated from landfill soil. *Int J Syst Evol Microbiol*. 2010;60: 2715–8. doi:10.1099/ijms.0.019661-0
54. Zhu L, Long M, Si M, Wei L, Li C, Zhao L, et al. *Asticcacaulis endophyticus* sp. nov., a prosthecate bacterium isolated from the root of *Geum aleppicum*. *Int J Syst Evol Microbiol*. 2014;64: 3964–3969. doi:10.1099/ijms.0.066605-0
55. Lembeck WJ, Colmer AR. Effect of herbicides on cellulose decomposition by *Sporocytophaga myxococcoide*. *Appl Microbiol*. 1967;15: 300–303.
56. Vance I, Topham CM, Blayden SL, Tampion J. Extracellular Cellulase Production by *Sporocytophaga myxococcoides* NCIB 8639. *Microbiology*. 1980;117: 235–241. doi:10.1099/00221287-117-1-235
57. An DS, Im WT, Yang HC, Yang DC, Lee ST. *Dyella koreensis* sp. nov., a B-glucosidase-producing bacterium. *Int J Syst Evol Microbiol*. 2005;55: 1625–1628. doi:10.1099/ijms.0.63695-0
58. Weon HY, Anandham R, Kim BY, Hong SB, Jeon YA, Kwon SW. *Dyella soli* sp. nov. and *Dyella terrae* sp. nov., isolated from soil. *Int J Syst Evol Microbiol*. 2009;59: 1685–1690. doi:10.1099/ijms.0.004838-0
59. Khianggam S, Tanasupawat S, Akaracharanya A, Kim KK, Lee KC, Lee JS. *Cohnella cellulosilytica* sp. nov., isolated from buffalo faeces. *Int J Syst Evol Microbiol*. 2012;62: 1921–1925. doi:10.1099/ijms.0.032607-0
60. Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, et al. Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat Biotechnol*. 2007;25: 1281–1289. doi:10.1038/nbt1354

61. Ping L, Yanxin W, Kun L, Lei T. Construction of a microbial system for efficient degradation of cellulose. 2008 Int Work Educ Technol Train 2008 Int Work Geosci Remote Sens. 2008; 344–347. doi:10.1109/ETTandGRS.2008.117
62. Rivas R, García-Fraile P, Mateos PF, Martínez-Molina E, Velázquez E. *Paenibacillus cellulosilyticus* sp. nov., a cellulolytic and xylanolytic bacterium isolated from the bract phyllosphere of *Phoenix dactylifera*. Int J Syst Evol Microbiol. 2006;56: 2777–2781. doi:10.1099/ijs.0.64480-0
63. Wang CM, Shyu CL, Ho SP, Chiou SH. Characterization of a novel thermophilic, cellulose-degrading bacterium *Paenibacillus* sp. strain B39. Lett Appl Microbiol. 2008;47: 46–53. doi:10.1111/j.1472-765X.2008.02385.x
64. Schlesner H, Jenkins C, Staley JT. The Phylum Verrucomicrobia: A Phylogenetically Heterogeneous Bacterial Group. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. Prokaryotes. 3rd ed. New York: Springer; 2006. pp. 7:881–896. doi:10.1007/0-387-30747-8
65. Magurran AE. Measuring Biological Diversity. Malden: Blackwell Science; 2004.
66. Hamady M, Lozupone C, Knight R. Fast UniFrac: Facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. ISME J. 2010;4: 17–27. doi:10.1038/ismej.2009.97
67. Cole JR, Wang Q, Fish JA., Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014;42: 633–642. doi:10.1093/nar/gkt1244
68. Moreira-Soto RD, Sanchez E, Currie CR, Pinto-Tomás A. Ultrastructural and microbial



analyses of cellulose degradation in leaf-cutter ant colonies.

69. Allgaier M, Reddy A, Park JI, Ivanova N, D'haeseleer P, Lowry S, et al. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One*. 2010;5: e8812. doi:10.1371/journal.pone.0008812
70. Hooper DU, Chapin III FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, et al. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecol Monogr*. 2005;75: 3–35. doi:10.1890/04-0922
71. Nielsen UN, Ayres E, Wall DH, Bardgett RD. Soil biodiversity and carbon cycling: A review and synthesis of studies examining diversity-function relationships. *Eur J Soil Sci*. 2011;62: 105–116. doi:10.1111/j.1365-2389.2010.01314.x
72. Hättenschwiler S, Fromin N, Barantal S. Functional diversity of terrestrial microbial decomposers and their substrates. *Comptes Rendus - Biol. Academie des sciences*; 2011;334: 393–402. doi:10.1016/j.crv.2011.03.001
73. Berlemont R, Martiny AC. Phylogenetic distribution of potential cellulases in bacteria. *Appl Environ Microbiol*. 2012;79: 1545–1554. doi:10.1128/AEM.03305-12

## **Chapter 5: Long-term cellulose enrichment selects for highly cellulolytic taxa and competition for public goods**

**In preparation:** Gina R. Lewin, Nicole Davis, Bradon R. McDonald, Adam J. Book, Marc G. Chevrette, Steven Suh, Ardina Boll, Cameron R. Currie. Long-term cellulose enrichment selects for highly cellulolytic taxa and competition for public goods.

**Co-authors contributed in the following ways:** Conceived and designed the experiments: GRL, AJB, ND, CRC. Performed experiments and analyzed data for Figure 5.1: GRL; Figure 5.2: GRL, ND; Figure 5.3: GRL; Figure 5.4: GRL, ND; Figure 5.5: GRL, AJB, BRM; Figure 2.6: GRL, AJB, BRM, MGC. Contributed strains: SS, AB. Wrote the paper: GRL, CRC.

### **5.1 Abstract**

There is an important but complex relationship between a community's taxonomic diversity and its ability to degrade recalcitrant but energy rich plant biomass. This relationship is mediated through both positive and negative interactions among community members, but the complexity of many natural environments prevents clear identification of how interactions can shape a community and its ability to degrade plant biomass. Here, we enriched six complex microbial communities from biomass-enriched leaf-cutter ant refuse dumps for up to 60 transfers using cellulose, the dominant polymer within the plant cell wall, as the sole source of carbon and energy. The cellulolytic ability of four enrichment lines increased then stabilized, while cellulolytic ability varied in two lines throughout the sampling period. Over time, the cellulose-degrading organisms *Cellvibrio* and/or *Cellulomonas* increased in abundance and then dominated in each enrichment line, and a diversity of taxa were maintained at low abundance. Cellulose degradation likely was slowed through competition, as exhibited by a higher expression of genes

that allow use of cellulose breakdown products in slower degrading enrichment lines and high expression of genes for localization and the production of antagonistic compounds across all communities. The results of this study demonstrate that even simple media conditions can support a fairly complex community and that not only general taxonomic diversity but also specific functional diversity and competition affect the stability and function of communities.

## **5.2 Introduction**

The microbial decomposition of plant biomass is a critical step in the global carbon cycle and a major pathway for the release of carbon dioxide into the atmosphere. A select few microbes possess the full suite of enzymes necessary to degrade plant biomass, and these organisms are often found within diverse communities in soil, leaf litter, and the guts of herbivores. A variety of field-based and microcosm experiments have revealed that species diversity has a complex and varied relationship with the rate of decomposition of plant biomass [1–8]. In part, the variation in this relationship occurs because the functional ability of a community is not simply the sum of the traits of its constituent taxa [9,10]. Microbes alter the abundance and physiology of their neighbors through cooperation, antagonism, and a variety of other mechanisms. These interactions have the potential to shape the ability of a community to degrade plant biomass.

Understanding plant biomass degradation within a community framework is important for dissecting the effects of interactions and taxonomic diversity on degradation rates. While detailed studies have made progress in understanding the community dynamics of decomposition in anaerobic environments such as the termite gut [11], the high diversity of aerobic plant biomass-degrading communities in soil and leaf litter often inhibits a detailed, molecular

characterization of the degradation and the associated interactions. In model laboratory isolates, aerobic organisms secrete a suite of enzymes to break down the polymers within the plant cell wall [12]. These insoluble polymers, including cellulose, a crystal of  $\beta$ -1,4-linked glucose molecules and the most abundant component of the plant cell wall, are degraded into soluble oligosaccharides, which are imported into the cell and catabolized into their constituent monomers. In communities, there are a myriad of conceivable interactions that would influence degradation rates. Organisms may degrade cellulose alone or cooperatively. The soluble cellooligosaccharides can be thought of as public goods that could be metabolized by the cellulase producers or by noncellulolytic cheaters. Cross-feeding could stabilize the community while toxin production, or even predation and phage, could negatively influence community dynamics and decomposition rates.

Experimental enrichments have proven to be a useful technique to dissect complex microbial communities [13–15]. By repeatedly transferring natural communities on a selective substrate, in this case, plant biomass or its constituent polymers, it is possible to lower the diversity of communities [13,16]. The decreased diversity, then, can allow for the identification of the organisms responsible for specific functions and the identification of the interactions between organisms.

Here we used leaf-cutter ant refuse dumps as inoculum for experimental enrichments. In Neotropical forests and savannahs, a significant percentage of plant material is degraded in refuse dumps of the dominant herbivore, leaf-cutter ants (Figure 5.1a,b). Refuse dumps are created by ants as they discard leaf material that has already been partially degraded by the ants’

fungal cultivar. This discarded leaf material is enriched in cellulose and lignin, and a complex microbial community develops in refuse dumps distinct from that found in tropical soils or in ant colonies [17–21]. We chose this refuse dump material as our inoculum source because it is a defined, recalcitrant substrate. Also, metagenomic analyses have indicated that this community contains known cellulose degrading genera, including *Cellvibrio*, *Cellulomonas*, *Sorangium*, and *Xanthomonas*, but have not been able to discern the roles of these organisms or their interactions with other microbes due to the community's high diversity [18]. Finally, refuse dump communities have been shown to have high cellulolytic ability when grown on pure cellulose [17].

In this project, we aimed to study the interplay between community diversity and growth on cellulose and to identify microbe-microbe interactions that impact cellulose degradation. We generated then used simplified microbial communities using an enrichment procedure, selecting for communities capable of growing on cellulose in otherwise carbon-free minimal media. The six enrichment lines presented here were inoculated with material from leaf-cutter ant refuse dumps from Costa Rica and then selected to degrade cellulose filter paper for up to 60 transfers (~275 generations; Figure 5.1a,c). We correlated the ability of each enrichment line to degrade cellulose over time with changes in its taxonomic composition. Then, using metagenomic and metatranscriptomic sequencing combined with physiological analysis of isolates, we analyzed the dominant cellulolytic organisms and their interactions with other community members.

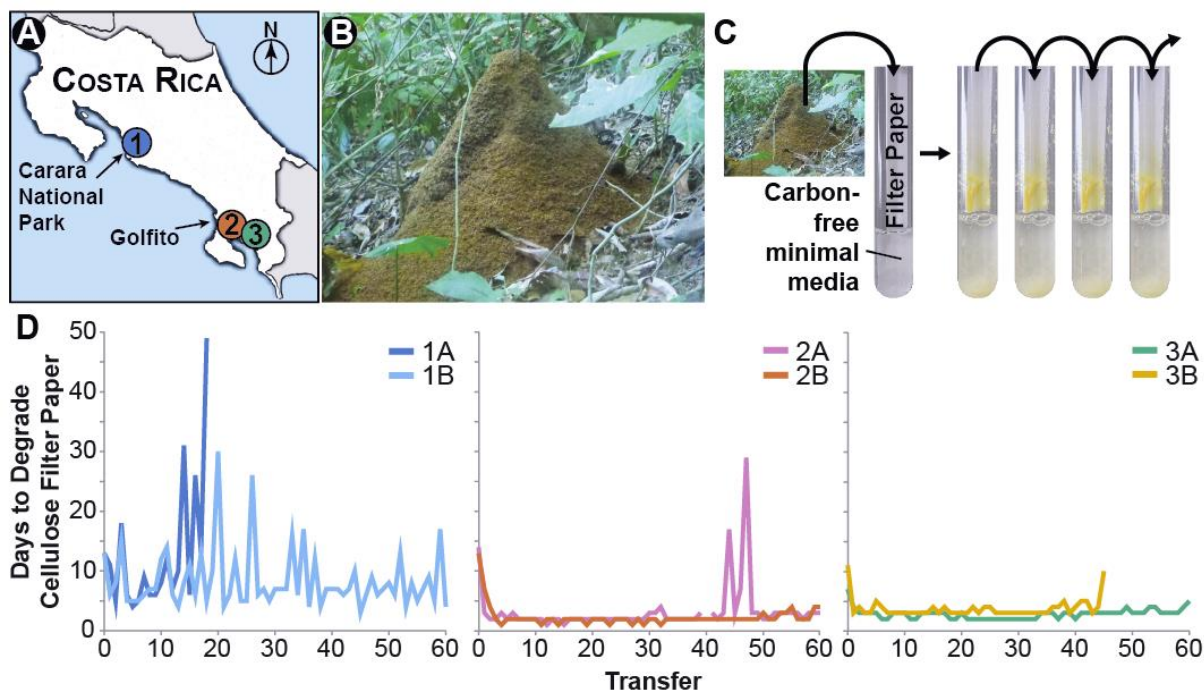


**Figure 5.1. Cellulose enrichment methods and results.** (A) Locations of leaf-cutter ant refuse dumps 1, 2, and 3 (B) Leaf-cutter ant refuse dumps are large piles of cellulose-enriched leaf material. (C). Refuse material was inoculated into test tubes with a strip of cellulose filter paper as the sole carbon source. Once the filter paper broke in half (always at the air-liquid interface), the microbial community in the corresponding test tube was transferred to a fresh filter paper test tube. (D) Ability of six enrichment lines including two independent inoculations from each refuse dump to degrade filter paper over time, measured by the time to degrade the cellulose filter paper.

## 5.3 Results

### 5.3.1 Cellulose degradation by enrichment lines

We tracked the ability of six enrichment lines from three leaf-cutter ant (*Atta colombica*) refuse dumps to degrade a strip of cellulose filter paper over sixty transfers (Figure 5.1d). Initially, communities took between 7 and 14 days to each degrade the cellulose filter paper. Over the 60 transfers, four of the six enrichment lines remained capable of degrading cellulose, and at

transfer 60, the remaining communities took between 3 and 5 days to degrade the cellulose filter paper. The filter paper-degrading ability of lines 1A and 1B from refuse dump 1 were more variable than those from dumps 2 and 3. The enrichment lines from refuse dump 1 oscillated in their abilities to degrade cellulose over time, ranging from degradation in 4 days to 30 days, and transfer 19 of enrichment line 1A did not break apart the cellulose filter paper and therefore was not continued. In contrast, during the first five transfers, the enrichment lines from refuse dumps 2 and 3 (2A, 2B, 3A, and 3B) increased their cellulose degradation rate and then stabilized, consistently breaking the filter paper between 1 and 4 days. However, the ability of enrichment line 2A to degrade cellulose slowed multiple times between transfers 40 and 50, and enrichment line 3B was stopped after transfer 46 because the community did not break the cellulose filter paper.

The variability in cellulolysis of enrichment lines from refuse dump 1 and the stability of enrichment lines from refuse dumps 2 and 3 are supported by replicates of each enrichment line started at the first transfer and continued until transfer 20 (Figure A5.1). Quantitative cellulose degradation assays also supported the pattern of high degradation in lines 2A, 2B, 3A, and 3B and variable, slower degradation in lines 1A and 1B (Figure A5.2). Communities assayed from across the sixty transfers from enrichment lines 2A, 2B, 3A, and 3B degraded almost all detectable cellulose in 10 days (average:  $94.5 \pm 4.3\%$  cellulose degradation), while communities from enrichment lines 1A and 1B degraded between 38% and 68% of cellulose across different transfers (average:  $50.9 \pm 10.8\%$  cellulose degradation) (Figure A5.2).

### ***5.3.2 Taxonomic shifts in community structure across 60 transfers***

To understand taxonomic shifts in these communities during the enrichments, we analyzed the community structure using 16S rRNA gene sequencing of representative enrichment lines at five time points for line 1A and 10 time points for lines 2B and 3A. We found that the predicted number of OTUs (Chao1 Metric) decreased then stabilized in all three communities and significantly linearly correlated with the number of days to degrade the cellulose filter paper for lines 2B and 3A (Figure 5.2a). Similarly, the Inverse Simpson's Diversity Index decreased and then stabilized, and it only significantly linearly correlated with days to degrade cellulose for enrichment line 2B (Figure 5.2b). The Berger Parker Index (dominance measurement) stabilized in community 2B but was variable over time in communities 1A and 3A (Figure 5.2c), and no significant correlation was found between this Index and the number of days to degrade cellulose.

We found that the community structure significantly differed between all three enrichment lines using an Analysis of Molecular Variance (AMOVA;  $p < 0.001$  for all pairs using the Bray-Curtis Dissimilarity Metric). Additionally, using non-metric multidimensional scaling (NMDS) with three dimensions (stress = 0.16;  $R^2 = 0.82$ ) to represent the differences between transfers and lines, we observed that enrichment line 1A shifted positively along axis 2 and negatively along axis 3 during the first 5 transfers (Figure 5.2d). In contrast, enrichment lines 2B and 3A shifted in the opposite directions, negatively along axis 2 and positively along axis 3 during the first ~5 transfers. Also, the number of days to degrade filter paper correlated positively with axis 2 and negatively with axis 3 (Spearman correlation;  $p = 0.0023$  and  $p = 0.0223$ , respectively; axis 1 not significant). After approximately 5 transfers, we observed that the community structure of the 2B and 3A lines stabilized, reflecting the stable and rapid cellulose degradation from transfer 5 to 60



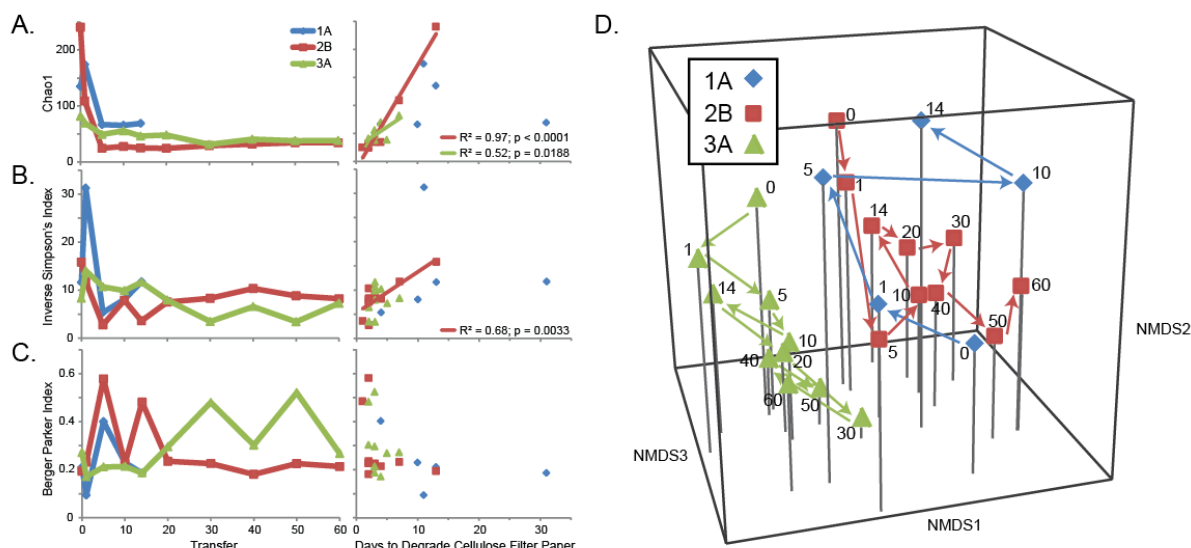


**Figure 5.2. Diversity changes in enrichment lines over time.** Shifts in (A) Chao1 (Estimated number of OTUs), (B) Inverse Simpson's Index, and (C) Berger Parker Index of dominance over time and relative to the cellulolytic ability of the community. A best fit linear line is shown for the significant linear relationships between each diversity index and the days to degrade cellulose filter paper. (D) NMDS representation (Stress = 0.16,  $R^2 = 0.82$ ) of Bray-Curtis dissimilarity metric between samples. The transfer number of each sample is indicated next to the corresponding point.

in these lines. In contrast, no stabilization was observed in the 1A enrichment line before it stopped degrading the cellulose filter paper.

Similar patterns are illustrated by focusing on changes in OTU relative abundance over time and among samples (Figure 5.3a). In enrichment line 1A, the most abundantly detected OTU varied over time, and only a few OTUs were consistently abundant across transfers. In contrast, in the fast degradation enrichment lines (2B and 3A), community abundance was relatively stable and sequences for putative cellulose degrading OTUs *Cellulosimicrobium* (OTU2) and *Cellvibrio* (OTU4) were highly abundant [22,23]. When the *Cellvibrio* OTU was clustered at 100% identity, the population structure appeared to shift over time in enrichment lines 1A and 2B. For

example, in line 2B, initially the 2B-II amplicon was more abundant, but at later transfer points, the 2B-I amplicon sequence was dominant (Figure A5.3a).

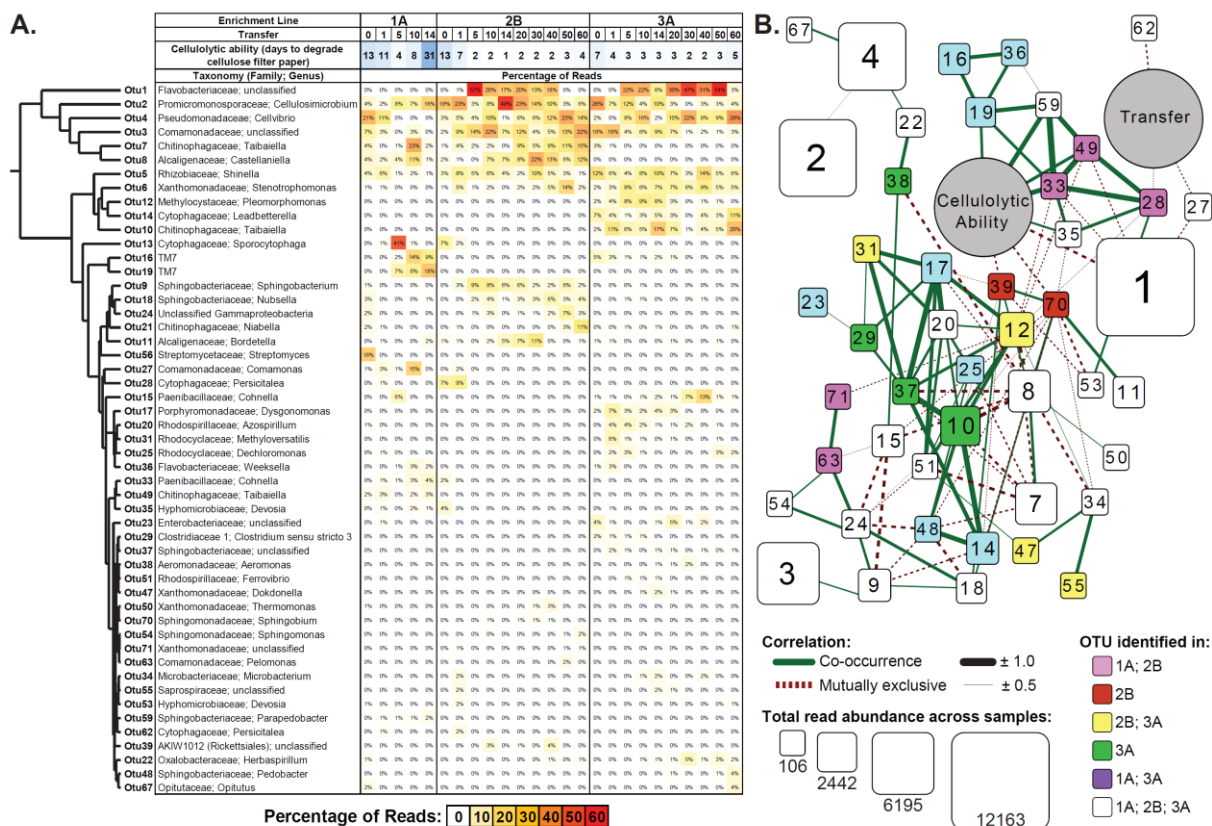


**Figure 5.3. OTU patterns and correlations among enrichment lines and across transfers.** (A) Change in OTU abundance over time and between enrichment lines. OTUs are clustered using Ward's method, and their assigned taxonomy is indicated. (B) Network of correlating OTUs. OTUs are indicated by the number in each box. Correlations were determined using Pearson, Spearman, and Kendall correlation metrics, and a correlation (edge) was designed between 2 OTUs (nodes) if significant correlations of their abundances across samples were found using at least two of these methods to be greater than 0.5 (co-occurrence; green solid line) or less than -0.5 (mutually exclusive; red dashed line) after statistical correction. OTU abundance was also correlated with transfer time point and cellulolytic ability (days to degrade cellulose filter paper). Line thickness indicates average correlation strength. OTUs are colored based on the enrichment lines they were identified in; no OTUs included in this network were found exclusively in line 1A. OTU size indicates the total read abundance across samples.

Co-occurrence network analysis demonstrated that *Cellulosimicrobium* (OTU2) and *Cellvibrio* (OTU4) were negatively correlated with each other (Figure 5.3b). Interestingly, the abundance of the *Cellvibrio* (OTU4) was positively correlated with the putative nitrogen-fixing *Herbaspirillum* (OTU22) [24]. The co-occurrence network and heat map analyses also identified OTUs that were initially present in all three enrichment lines but then decreased in abundance during the enrichment process including OTU27 (*Comamonas*), OTU28 (*Persicitalea*), and OTU62 (*Persicitalea*). Additionally, certain OTUs including OTU1 (unclassified *Flavobacteriaceae*) were negatively correlated with cellulolytic ability, while other OTUs correlated positively with cellulolytic activity including OTU19 (TM7), OTU33 (*Cohnella*), and OTU35 (*Devosia*), likely driven by their shifts in abundance in enrichment line 1A.

### **5.3.3 Taxonomic shifts in community structure during one transfer**

To further investigate community succession and provide insight into community dynamics, we tracked OTU abundance over 7 days in replicate tubes of enrichment line 2B at transfer 70. These samples broke the filter paper between 48 and 72 hours, and correspondingly, cellulose degradation was detected using quantitative methods starting at the 72-hour time point (Figure 5.4a). Using 16S rRNA gene sequencing of this community at 0, 1, 8, 24, 48, 120, and 168 hours, no significant differences were observed in the number of OTUs detected or estimated number of OTUs. However, there was a significant difference overall in the Inverse Simpson's Index between time points (Analysis of Variance [ANOVA]:  $df = 7$ ,  $F = 2.9001$ ,  $p = 0.0263$ ; Figure 5.4b). Specifically, we observed a >2-fold increase in this Index between the 48 and 72-hour time points ( $3.57 \pm 0.57$  [Standard Error, SE] and  $7.29 \pm 0.81$  [SE], respectively; Tukey-

Kramer Honest Significant Difference [HSD] Test:  $p = 0.0207$ ), which corresponds with the interval in which the cellulose filter paper broke in half. A change in diversity during this interval is also supported by the Berger-Parker Dominance Index, which differed across samples (ANOVA:  $df = 7$ ,  $F = 3.4507$ ,  $p = 0.0120$ ) and specifically differed between the 48-hour and 72-hour

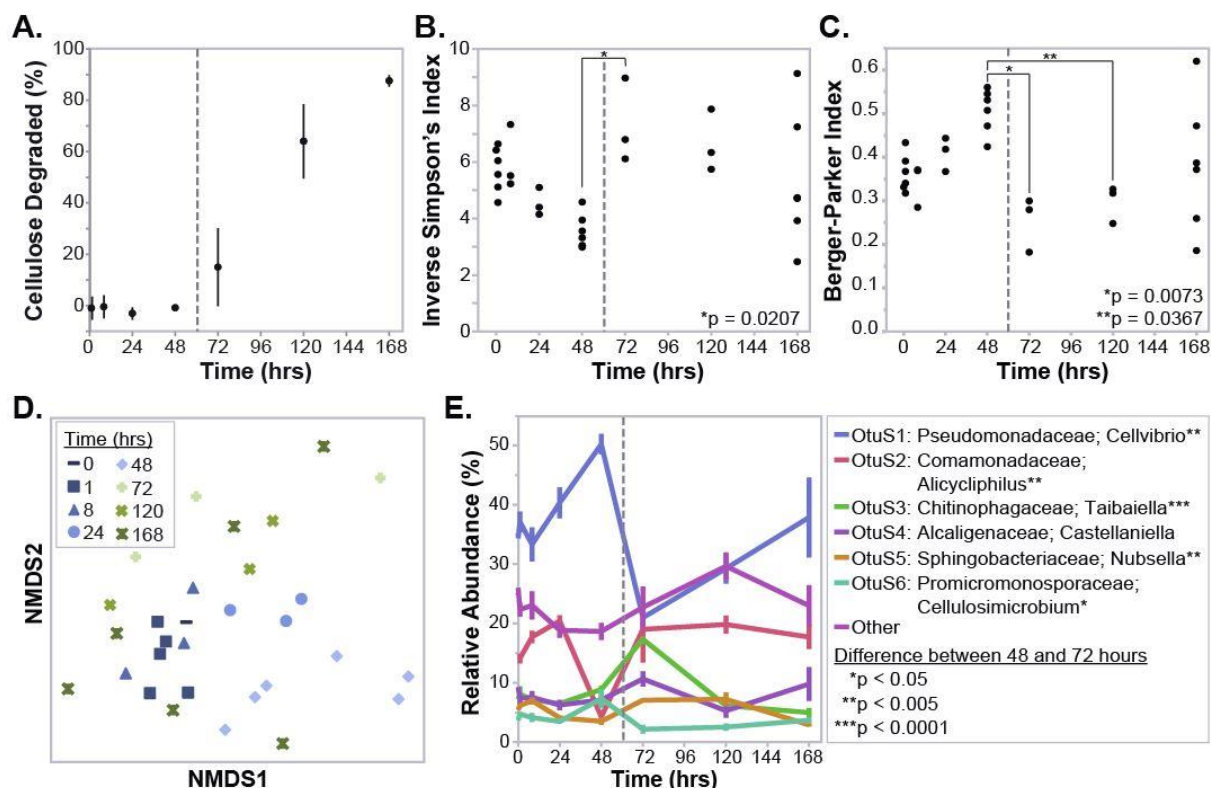


**Figure 5.4. Decomposition and taxonomic shifts for enrichment line 2B across seven days.** Replicate cultures of enrichment line 2B at transfer 70 were monitored over 7 days. In all panels, the dashed line indicates when filter paper broke in half in cultures used for community analysis. Diversity analyses were performed using OTUs clustered at 97% identity, and statistical analyses were performed using the Tukey-Kramer HSD Test. (A) Quantitative measurement of cellulose degradation. (B) Change in alpha diversity (Inverse Simpson's Index) over time. (C) Change in dominance index over time. (D) NMDS representation (Stress = 0.17;  $R^2 = 0.89$ ) of the Bray-Curtis dissimilarity metric over time. (E) Change in relative abundance of the 6 most abundant OTUs over time. Significant p-values are indicated for differences in relative read abundance directly before and after the cellulose filter paper broke (between 48 and 72 hours).

hour time points (Tukey-Kramer HSD Test:  $p = 0.0073$ ) and between the 48-hour and 120-hour time points (Tukey-Kramer HSD Test:  $p = 0.0367$ ; Figure 5.4c). NMDS representation of the Bray-Curtis dissimilarity metric indicates a large shift in community structure between 48 hours and 72 hours (Figure 5.4d). This shift correlates with significant decreases in the relative abundances of the putative cellulolytic *Cellvibrio* and *Cellulosimicrobium* OTUs during this time period and significant increases in the abundances of the Comamonadaceae, Chitinophagaceae, and Sphingobacteriaceae OTUs (Figure 5.4e).

### 5.3.4 Functional analysis of cellulose degradation

To further understand the roles of the microbes in each enrichment line and their interactions, we sequenced and assembled the metagenomes of enrichment lines 1A and 1B at transfer 14 and enrichments lines 2A, 2B, 3A, and 3B at transfer 20 and the metatranscriptomes of enrichment lines 1A, 2B, and 3A at corresponding time points. The majority of cellulase genes present and expressed in the communities coded for endoglucanases from glycoside hydrolase [GH] families 5 and 9, but exocellulases and lytic polysaccharide monooxygenase [LPMO] genes were also detected (Figure 5.5b and Figure A5.4). *Cellvibrio* and/or *Cellulomonas* species encoded and expressed the majority of putative cellulases in each enrichment line, except in enrichment line 1B, where the taxonomy of many of the cellulase genes was unassigned (Figure 5.5a; the *Cellulosimicrobium* OTU in from the 16S rRNA gene analysis includes *Cellulomonas* sequences). However, in both the metagenomic and metatranscriptomic data, we identified differences in the relative quantity and transcription levels of cellulase genes, which were more abundant in the highly cellulolytic communities (2A, 2B, 3A, and 3B) than in the communities that were more variable in their ability to degrade cellulose (1A and 1B).

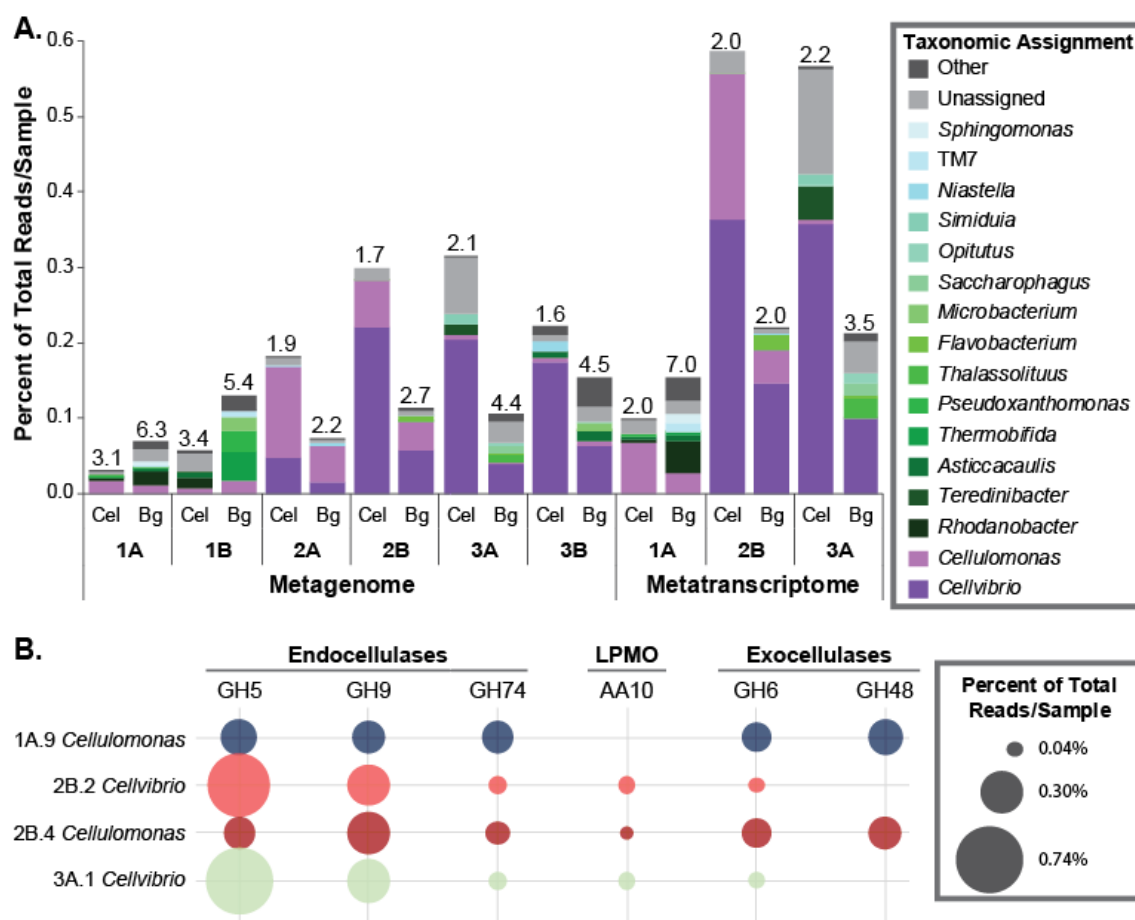


**Figure 5.5. Relative levels of cellulase (Cel) and  $\beta$ -glucosidase (Bg) genes and transcripts. (A)**

Endocellulase, exocellulase, and lytic polysaccharide monooxygenase CAZy classes GH5, GH6, GH7, GH9, GH12, GH48, GH74 and AA10 were included as cellulases. CAZy classes GH1 and GH3 were included as  $\beta$ -glucosidases. For metagenomic counts, the coverage of genes that mapped to each taxonomy was summed and normalized to the total coverage in each metagenome. For metatranscriptome counts, RNA read counts mapping to genes assigned to each taxonomy were summed and normalized to total read counts of each sample. Genera in purple (*Cellvibrio* and *Cellulomonas*) are dominant cellulose degraders across communities. The Inverse Simpson's Index of cellulase or  $\beta$ -glucosidase producers in each sample is indicated above the corresponding bar. (B) Relative gene expression of endocellulase, lytic polysaccharide monooxygenase (LPMO), non-reducing end exocellulase GH6, and reducing end exocellulase GH48 genes in *Cellvibrio* and *Cellulomonas* composite genomes. Read counts are normalized to total mapped metatranscriptomic reads in each sample.

We further investigated the contributions of *Cellvibrio* and *Cellulomonas* to cellulase production by binning the metagenomic contigs to create composite genomes estimated to be 95% and 99% complete for the *Cellvibrio* populations in enrichment lines 2B and 3A, respectively, and 59% and 54% complete for the *Cellulomonas* populations in enrichment lines 1A and 2B, respectively (Table A5.1). In these composite genomes, cellulases were highly expressed (Figure 5.5b). In both *Cellulomonas* composite genomes, the full suite of genes necessary to deconstruct cellulose including multiple endocellulase genes (GH5 and GH9), a reducing end exocellulase gene (GH48), and two non-reducing exocellulase genes (GH6) were expressed. Also, in the enrichment line 2B *Cellulomonas* populations, GH10 and GH11 putative xylanases were the 18<sup>th</sup> and 25<sup>th</sup> most highly expressed genes, and a GH10 was the 56<sup>th</sup> most highly expressed genes in the 1A *Cellulomonas* population. In contrast, the *Cellvibrio* composite genomes did not encode a GH48 reducing end exocellulase, but GH9 and GH5 endocellulases and an AA10 LPMO were highly expressed (Figure 5.5b).

We isolated one *Cellvibrio* strain from enrichment line 3A at transfer 50. The 16S rRNA gene sequences of this isolate (*Cellvibrio* sp. 3A-T50a) closely matched the 16S rRNA genes identified in the 3A.1 *Cellvibrio* composite genomes and in OTU4 in the amplicon analysis (Figure A5.3a). Additionally, the isolate degraded a strip of cellulose filter paper in 3 days and degraded ~100% of available cellulose in 10 days, similar to the enrichment line that it was isolated from (Figure A5.3b,c). *Cellvibrio japonicus*, a well-studied cellulolytic strain isolated from soil, only degraded 51% of available cellulose in 10 days [25].

### **5.3.5 Indications of competition within communities**

We examined the diversity of organisms that contained and expressed  $\beta$ -glucosidase genes, which encode enzymes to break down cellobiose into glucose (Figure 5.5a). As with cellulase production, *Cellvibrio* and *Cellulomonas* species contained and expressed the majority of  $\beta$ -glucosidase genes in most samples. We quantified the taxonomic diversity of organisms containing and expressing cellulases and  $\beta$ -glucosidases across samples using the Inverse Simpson's Index (Figure 5.5a). While the diversity of  $\beta$ -glucosidase producers was higher than or equal to the diversity of cellulase producers across all samples, the largest differences between  $\beta$ -glucosidase producer diversity and cellulase producer diversity was observed in lines 1A and 1B, the communities that were not as efficient at degrading cellulose. In line 1A, *Cellulomonas* dominated cellulase expression, and the diversity of cellulase expressing organisms was 2.0; however, *Rhodanobacter* was the most abundant  $\beta$ -glucosidase-expressing organism in this sample, and the overall diversity was 7.0. Furthermore, in the cellulolytic communities, over two times as many metagenomics and metatranscriptomic reads mapped to cellulase genes than  $\beta$ -glucosidase genes. In contrast, in enrichment lines 1A and 1B,  $\beta$ -glucosidase reads were approximately twice as abundant as cellulase reads.

When we functionally classified the 25 most highly expressed genes in select composite genomes, we found further signatures of competition (Figure 5.6a). Twenty-seven of the two hundred total genes analyzed are predicted to encode RHS repeat domains, toxins, and secondary metabolite production proteins. Genes predicted to encode putative defensive functions such as efflux pumps were highly expressed by the 1A.8 *Rhodanobacter* population and the 3A.7 *Opitutaceae* population, and the gene encoding a putative penicillin binding protein was highly expressed in the 2B.4 *Cellulomonas* composite genome. CRISPR genes and those predicted to



encode restriction enzymes were also identified, as were mobile elements and phage-related genes. Additionally, in composite genomes from all three enrichment lines analyzed, genes related to twitching (Type IV pili), gliding, or flagella-based motility were highly expressed. Also, in enrichment line 2B, putative adhesion genes were expressed by the *Cellvibrio* population and a biofilm-related gene was highly expressed by the *Cellulomonas* population.

We further investigated the highly expressed secondary metabolism genes in the *Cellvibrio* population 3A.1, which comprised parts of two separate gene clusters. Highly expressed genes 3A.1\_02159 and 3A.1\_02160 were sequential on the chromosome and were predicted to function as polyketide synthetases (PKSs) within a hybrid type 1 PKS-non ribosomal protein synthetase (NRPS) cluster (Figure 5.6b). Much of the rest of this cluster was also highly expressed including the NRPS domain, which was the 31<sup>st</sup> most highly expressed gene in the composite genome. 3A.1\_03151 was predicted to encode an NRPS protein within a cluster with high similarity to a cluster known to make the triscatecholate siderophore turnerbactin in the cellulolytic bacterium *Teredinibacter turnerae* T7901 (Figure 5.6c) [26]. Homologs of every turnerbactin gene are present in the 3A.1 *Cellvibrio* cluster with e-values ranging from  $6 \times 10^{-27}$  to 0.0. A notable difference from the previously described turnerbactin cluster is a rearrangement of a gene block containing a conserved hypothetical open reading frame and a putative transmembrane peptidase. Also, three genes are present this cluster that are not found in turnerbactin, one of which is an MbtH-like protein and two of which likely function as transporters.

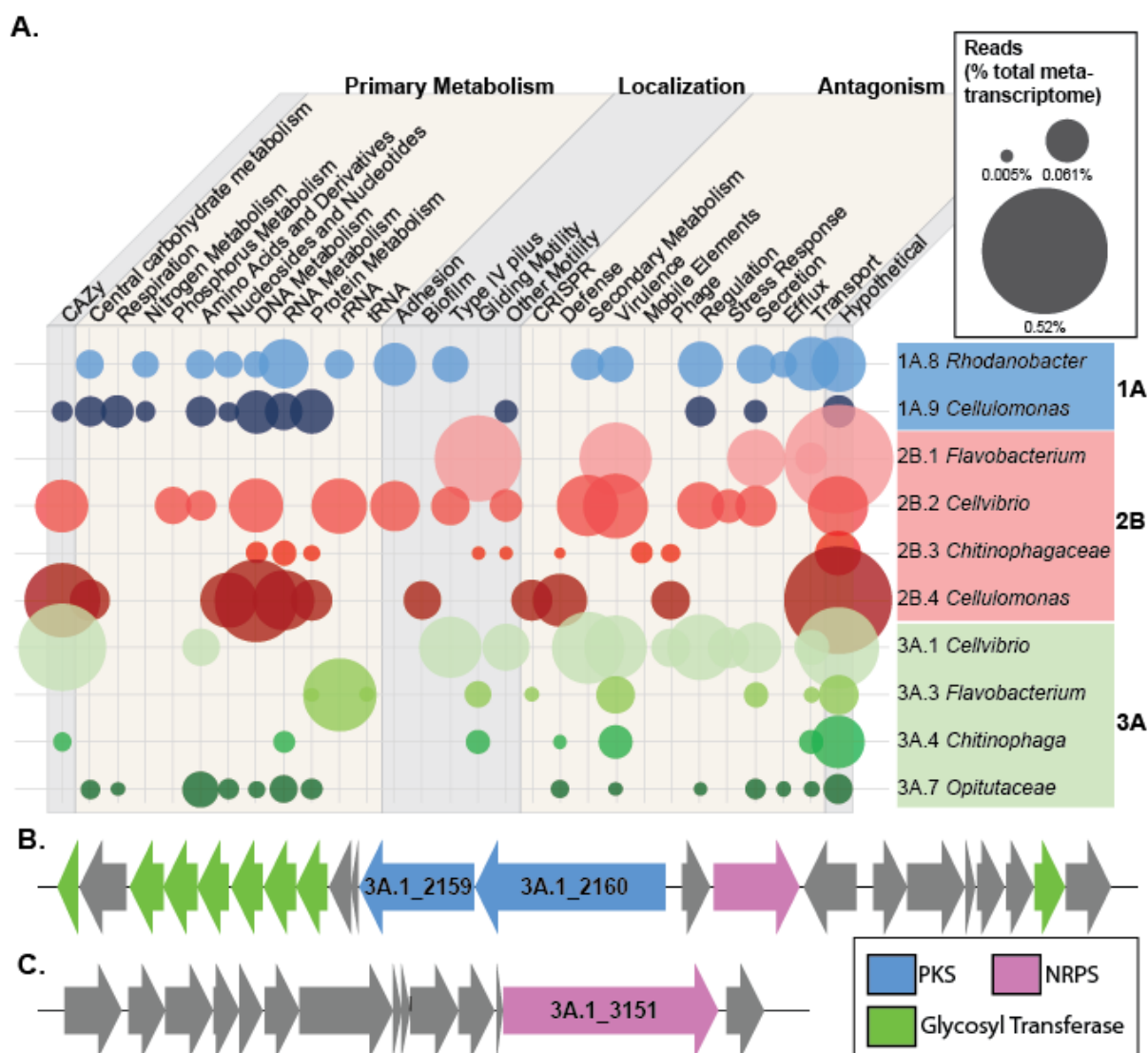


**Figure 5.6. Functional annotation and expression level of the 25 most highly expressed genes in select composite genomes.** (A) Functional annotations were hand curated using SEED, TIGRFAM, PFAM, and KEGG annotations, and similar functional classes were grouped. The size of each bubble represents the sum of transcriptomic reads mapping to each functional class for the 25 most highly expressed genes in each composite genome, relative to the total number of reads in the metatranscriptome. (B) Hybrid PKS-NRPS biosynthetic gene cluster including two highly expressed PKS genes in the 3A.1 *Cellvibrio* composite genome. (C) NRPS gene cluster in 3A.1 *Cellvibrio* composite genome including a highly expressed NRPS gene predicted to make the siderophore turnerbactin.

## 5.4 Discussion

Many recent studies have made strides in studying the natural history of microbial communities across space and time and have established that community diversity can correlate with community function [27–31]. While the underlying mechanisms that link the structure and function of communities are often elusive, they likely include interactions with abiotic and biotic factors. In this study, we created and analyzed enrichment lines of cellulose degrading communities to understand how interactions between bacteria can structure a community and affect its function. By studying the taxonomic changes over both short and long time periods, we tracked the succession of the communities, providing insight into the relationship between cellulolytic ability and diversity. Then, we used metagenomic and metatranscriptomic sequencing and isolate analysis to identify signatures of cooperation, competition, and antagonism that shape the communities.

### 5.4.1 Cellulose degradation in enriched communities

In this study, we observed remarkably high cellulolytic ability in many of our enriched communities and in our *Cellvibrio* isolate. The cellulolytic ability of the enrichment lines from refuse dumps 2 and 3 initially decreased drastically and then stabilized, breaking the filter paper in the enrichment test tubes in approximately three days and degrading almost 100% of cellulose in quantitative assays in 10 days (Figure 5.1d and Figure A5.1). Likewise, the *C. sp.* 3A-T50a isolate was able to degrade all detectable cellulose in 10 days (Figure A5.3). In a previous study using bacterial strains isolated from a variety of environments including soil, eukaryotic hosts, and leaf-cutter ant refuse dumps, the fastest filter paper degradation measured was six days and the maximum cellulose degradation in 10 days was 58% [32]. In another study using enrichment

of leaf-cutter ant refuse dumps material only for one transfer, a few communities had similarly high degradation abilities, but many were not able to degrade the cellulose filter paper [17].

In contrast, the enrichment lines from refuse dumps 1A and 1B showed much more variability in the time to degrade the cellulose filter paper (Figure 5.1d and Figure A5.1). While these communities were capable at times of degrading cellulose in as little as 4 days, at other times, these communities degraded the cellulose in over 20 days. The replicate enrichment lines for these communities were also variable (Figure A5.1). We hypothesize that this inconsistency is a reflection of an instability in the communities. While these lines contained cellulase-producing organisms including *Cellulomonas* (Figure 5.3), the communities as a whole were not able to degrade the cellulose as quickly as in the enrichment lines from refuse dumps 2 and 3. The slower cellulose degradation correlates with the lower relative number of DNA and RNA reads mapping to cellulase genes, suggesting fewer cellulases in these enrichment lines (Figure 5.5 and Figure A5.4). Additionally, as detailed below, competition for the breakdown products of cellulose likely contributed to the slower degradation rate, and also, possibly to the instability of the 1A and 1B enrichment lines.

Across all enrichment lines, *Cellvibrio* and *Cellulomonas* populations were the major cellulase producing organisms (Figure 5.5). The *Cellvibrio* populations identified here appear to be highly cellulolytic and specific to refuse dumps. *Cellvibrio* strain 3A-T50a, isolated from the 3A community, degraded more cellulose in 3 days than the well-characterized strain *Cellvibrio japonicus* degraded in 10 days (Figure A5.3). Additionally, although there were multiple *Cellvibrio* populations in each enrichment line based on clustering of 16S rRNA gene sequences

at 100%, strains with identical 16S rRNA gene sequences were dominant over time in both the 2B and 3A enrichment lines (Figure A5.3). Further, mapping of transcriptomic reads onto the 2B.2 *Cellvibrio* composite genome and 3A.1 *Cellvibrio* composite genome indicated that the same classes of cellulases were expressed at similar levels by both populations in the two different enrichment lines (Figure 5.5b). Since in enrichment line 2B, another dominant cellulose degrader, *Cellulomonas*, was present, while in enrichment line 3A, *Cellvibrio* produced the majority of cellulases, *Cellvibrio* did not appear to change its cellulase expression in response to the presence of another cellulolytic organism. Similarly, *Cellulomonas* expressed a similar ratio of cellulase genes whether it was the dominant cellulase producer, as in enrichment line 1A, or one of two cellulase producers as in enrichment line 2B (Figure 5.5b).

#### **5.4.2 Changes in taxonomic diversity over time**

While native refuse dumps are highly diverse, the enrichment process used in this study successfully reduced community diversity [18]. Much of this reduction occurred over the first five transfers, where both the estimated number of OTUs and the Inverse Simpson's Index decreased then stabilized in the majority of enrichment lines (Figure 5.2a,b). In some enrichment lines, this decrease in diversity was significantly correlated with an increased cellulose degradation rate (Figure 5.2). This negative relationship between taxonomic diversity and cellulose degradation is partially explained by the increased dominance of cellulose degrading OTUs *Cellvibrio* and *Cellulomonas* (*Cellulosimicrobium*) (Figure 5.2c and Figure 5.3a). In addition, a decreased diversity may have decreased competition, allowing the cellulose-degrading organisms to use more cellular resources for growth and cellulase production instead of for energy-demanding antagonistic phenotypes such as secondary metabolite production and

biofilm formation. (Figure 5.6a). However, in other lines, no correlation was found between taxonomic diversity and cellulose degradation (Figure 5.2), potentially because, as discussed in detail below, functional diversity is important in addition to taxonomic diversity.

Given the simplistic conditions of these enrichment cultures, consisting of minimal medium with cellulose filter paper as the sole carbon source, it is somewhat surprising that competition did not drive these communities to simplify further. In this “paradox of the plankton”, there were a limited number of niches in the cultures including (1) cellulose degradation and (2) use of cellobiose sugars released by cellulose degradation and (3) use of secreted waste products, yet the communities stabilized to approximately 30 OTUs (Figure 5.2a) [33]. Perhaps, these communities did not fully stabilize. Small fluctuations occurred in beta-diversity and taxonomic abundance across the 60 transfers (Figure 5.2d and Figure 5.3a), and competition appeared to be a driving force in community dynamics (Figure 5.6). However, the persistence of a diversity of organisms across 60 transfers indicates that niche space in these communities was complex enough to support over 30 populations.

The pattern of mutual exclusion between cellulose-degrading organisms *Cellulomonas* (*Cellulosimicrobium*) and *Cellvibrio* (Figure 5.3) provides evidence that competition for niche space was occurring. These taxa could likely directly compete for access to cellulose and its degradation products for carbon and energy. Initially, these organisms were present in all three enrichment lines analyzed over time, reflecting their presence in natural refuse dumps (Figure 5.3a) [18]. In enrichment line 1A and 1B, *Cellulomonas* (*Cellulosimicrobium*) eventually dominated, and in enrichment line 3A and 3B *Cellvibrio* eventually dominated (Figure 5.3 and

Figure 5.5a). In enrichment lines 2A and 2B, both remained abundant, although the dominant genus fluctuated over time, potentially indicating competition for niche space. The congruence in the dominant cellulolytic OTU(s) between enrichment lines from the same refuse dumps indicates that potentially, deterministic factors such as interactions with other community members mediated the fluctuations in abundance of *Cellvibrio* and *Cellulomonas* populations. Analysis of replicate lines could help further determine if abundance fluctuations are stochastic or deterministic (Figure A5.1).

The niches filled by other populations in these enrichment lines are more difficult to determine. Some organisms coded for and expressed  $\beta$ -glucosidases but not cellulases, indicating that these organisms were competing with cellulose degrading *Cellvibrio* and *Cellulomonas* for cellobiose sugars (Figure 5.5a). However, many relatively abundant community members did not code for or express cellulases or  $\beta$ -glucosidases. These organisms are likely secondary consumers in the community. Dead cell debris and secreted waste compounds could provide a number of carbon sources and a variety of niches.

Analysis of the taxonomic shifts in the 2B community across a single transfer supports the hypothesis that the majority of community members used energy released by cellulose-degrading populations (Figure 5.4). Over the time period when the filter paper broke in half and significant cellulose degradation was first detected (between 48 and 72 hours), there was a significant increase in the Inverse Simpson's Index and decrease in the Berger-Parker Dominance index, as well as the Bray-Curtis Dissimilarity Index. While reads mapping to *Cellvibrio* and *Cellulomonas* (*Cellulosimicrobium*) OTUs decreased in relative abundance during this time

period, reads mapping to many of the other dominant community members significantly increased in proportion between 48 and 72 hours (Figure 5.4e). This abrupt shift in community structure once more nutrients are available to non-cellulolytic organisms further supports that most community members were secondary consumers.

### 5.4.3 Interactions within enriched communities

Competition for cellulose and its breakdown products appears to be an extremely important interaction within these enrichment lines. Cellulose degrading *Cellvibrio* and *Cellulomonas* populations likely were in competition, as demonstrated by the pattern of mutual exclusion between these OTUs (Figure 5.3b). While, the mechanism of this competition is unclear, data indicate that access to cellobiose is important for the stability of the communities. The enrichment lines from refuse dump 1, which were more variable in their ability to degrade cellulose, had a higher diversity of organisms coding for and expressing  $\beta$ -glucosidase genes and more  $\beta$ -glucosidase genes relative to cellulase genes (Figure 5.5a). Competition for cellobiose by noncellulolytic microbes would likely slow the growth of cellulolytic organisms. Similarly, competition for cellobiose have been one factor leading to mutual exclusion between *Cellvibrio* and *Cellulomonas*.

Our analysis of the metatranscriptomic data also indicated that a number of other negative interactions were occurring in these communities. We hypothesize that competition for localization is indicated by the high expression of adhesion-, biofilm-, and motility-related genes in composite genomes (Figure 5.6a). Any organism competing for cellulose breakdown products must localize as close as possible to the insoluble cellulose, increasing the effective



concentration of cellulases and cellobiose. Additionally, the putative production of the siderophore turnerbactin by *Cellvibrio* indicates competition for iron, and the putative production of an unknown NRPS secondary metabolite by *Cellvibrio* and a range of highly expressed toxin- and defense-associated genes by a number of community members indicates further antagonism (Figure 5.6b). In a small number of community members, CRISPR-related genes were also highly expressed, indicating that phage are likely present and active in these communities as well (Figure 5.6a).

Overall, we detected more indications of negative interactions in these communities than positive interactions. These findings show that although the community structure of the enrichment lines appeared stable (Figure 5.2), antagonism is dominant force in the communities. Competition for resources dominated gene expression, possibly because the communities were only growing on a single carbon source (Figure 5.6). It is likely that on complex plant biomass, as opposed to purified cellulose, cooperation may have a more dominant role in substrate degradation due to the diversity of carbon-rich polymers within the plant cell wall [2]. However, these results do not exclude the possibility that a number of positive interactions are occurring as well. Positive interactions such as metabolic cross-feeding or syntrophy are difficult to detect within our dataset and subtler than many negative interactions [34]. Further investigation of these communities and others combined with detailed analysis of isolates and their metabolism will allow a broader understanding of the interplay of both positive and negative interactions in microbial communities.

## 5.5 Conclusion

We have presented a model in which upon enrichment on cellulose, one or two dominant organisms are responsible for cellulose degradation and competition for sugars is common among community members. By tracking these communities over time and analyzing their ecological dynamics, we linked diversity and community function with underlying signatures of cooperation and competition. Understanding how the combination of interactions within each community together influences its stability and cellulolytic ability will further increase our understanding of how communities assemble and function. Additionally, future investigations of these enrichment lines can illuminate what evolutionary changes may occur over the almost 300 generations per lineage in each community.

## **5.6 Materials and Methods**

### ***5.6.1 Refuse dump collection***

Samples were aseptically collected from the middle layer of refuse dumps and stored at 4°C until inoculation. Material from refuse dump 1 was collected in Carara National Park, Costa Rica on protected land, permitted by Resolution ACOPAC-INV-RES-003-13 from the Área de Conservación Pacífico Central, Sistema de Áreas de Conservación of the Ministerio del Ambiente, Energía y Telecomunicaciones. Refuse dump 2 was collected at GPS coordinates 8.653888, -83.183334, and refuse dump 3 was collected at GPS coordinates 8.644873, -83.187111, both in Golfito, Costa Rica. Collections on these protected land were permitted by Resolution ACOSA-INV-009-13 from the Área de Conservación Osa, Sistema de Áreas de Conservación of the Ministerio del Ambiente, Energía y Telecomunicaciones (Figure 5.1a).

### ***5.6.2 Enrichment procedure***

Two enrichment lines were started from each refuse dump sample (ex. Dump 2 inoculation A = enrichment line 2A). Enrichments were performed in test tubes containing 5 mL of carbon-free M63 minimal media and a 1 x 10 cm strip of Whatman #1 cellulose filter paper (GE Healthcare Life Sciences, Pittsburgh, PA) pressed against the side of the tube [17]. Cellulose was the only carbon source in these test tubes. For each enrichment line, we inoculated an approximately 3 mg (~2 mm diameter) piece of refuse dump material into a test tube and grew the cultures shaking at 30°C. If the microbial community was capable of degrading cellulose, the microbes grew directly on the cellulose filter paper and eventually broke the filter paper in half (Figure 5.1c). We checked each tube daily, and as soon as we observed that the filter paper was broken, the test tube was vortexed to break up the filter paper, and 200 µl of culture was transferred into three fresh test tubes using a wide orifice p200 pipette tip. At this first transfer, each of these replicate tubes was run as a separate enrichment line (ex. 2AA, 2AB, 2AC). At the second and all subsequent transfers, three replicate filter paper tubes were inoculated from each enrichment line, and the fastest tube to degrade the filter paper was transferred. At transfer 20, one of the remaining replicate lines from each original inoculation was chosen to be continued, and all other lines were discontinued (Figure A5.1).

### ***5.6.3 Quantification of cellulose degradation***

Quantification of cellulose degradation was performed every 5-10 transfers. During the transfer procedure, communities were also inoculated into tubes containing 8 mL of M63 minimal media and two pre-weighed 1 x 4 cm strips of cellulose filter paper (~70 mg). These tubes were grown shaking for ten days, after which the filter paper degradation was quantified using a previous published acid-detergent method [35].

#### ***5.6.4 Seven-day analysis of enrichment line 2B at transfer 70***

During the passage of enrichment line 2B for transfer 70, an additional 30 filter paper tubes were inoculated. Six tubes each were collected after 1, 48, and 168 hours, and three tubes each were collected after 8, 24, 72, and 120 hours. Samples were frozen for DNA extraction and sequencing, as detailed below. Tubes were also inoculated to quantify cellulose degradation, and three experimental and four control (uninoculated) tubes were collected at each time point.

#### ***5.6.5 Collection of cultures for sequencing analyses***

At each transfer or time point, 20% glycerol freezer stocks were stored at -80°C. Also, samples were collected for DNA analysis by centrifuging 2-3 mL of culture for 12 minutes at 16,100 xg in a benchtop centrifuge, removing the supernatant, and storing cell pellets at -20°C. At transfer 14 for enrichment lines 1A and 1B and transfer 20 for lines 2A, 2B, 3A, and 2B, larger cultures were also inoculated to obtain enough material for both RNA and DNA sequencing. To create these cultures, 3 mL from the three replicate tubes of an enrichment line were pooled, and 2 mL of the mixed material was inoculated into three flasks containing 50 mL of M63 minimal media and 12 1 x 4 cm pieces of filter paper. Flasks were grown shaking at 30°C. Pooled communities were also inoculated into three filter paper test tubes. When the filter paper broke in at least one of these tubes, the flasks were harvested. For DNA extraction, 3 microcentrifuge tubes, each with 1.5 mL of culture were centrifuged for 12 minutes at 16,100 xg in a benchtop centrifuge, the supernatant was removed, and samples were stored at -20°C. For RNA extraction, the remaining material in each flask was centrifuged in a conical tube for 15 minutes at 5000 xg at 4°C, washed three times with ultrapure water, and stored at -80°C.

### **5.6.6 DNA extraction and sequencing**

To extract DNA for 16S rRNA gene amplicon and metagenomic sequencing, the cell pellets collected as described above were thawed and combined with 450  $\mu$ L of Tris/Sucrose/EDTA (TSE) buffer with 5 mg/ml lysozyme. This mixture was vortexed gently then incubated for 45 min at 37°C. 13  $\mu$ L of 20 mg/ml Proteinase K was then added, and cells were incubated at 50°C for 10 minutes. To fully lyse cells, 225  $\mu$ L of 2% sodium dodecyl sulfate (SDS) was added; tubes were gently mixed for 1-2 min and then left at room temperature for 5 min. Nucleic acids were purified by adding 100  $\mu$ L neutral phenol followed by 50  $\mu$ L chloroform. Tubes were shaken then centrifuged at 16,100 xg for 15 minutes. The aqueous layer was combined with 60  $\mu$ L 3M sodium acetate and 700  $\mu$ L isopropanol, mixed gently, and stored overnight at -20°C. Tubes were spun for 5 min, the supernatant was removed, and cell pellets were redissolved in 500  $\mu$ L Tris/EDTA buffer (TE). To remove RNA, 10  $\mu$ L RNase (10 mg/mL) was added to each sample and gently rocked overnight. DNA was again purified using a phenol:chloroform extraction at a ratio of 2:1, followed by a chloroform extraction. DNA was precipitated in 50  $\mu$ L 3M sodium acetate and 350  $\mu$ L isopropanol. Tubes were mixed gently, left on ice for 30 minutes, and then centrifuged at 16,100 xg for 5 minutes. The supernatant was removed and pellets were resuspended in 13  $\mu$ L of water. DNA concentrations were determined using a Qubit Fluorometer (Thermo Fisher Scientific, Wilmington, DE) and purity was determined using a NanoDrop Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE).

For 16S rRNA gene-based community profiling, the V6 to V8 region of the 16S rRNA gene was amplified using bacteria-specific primers and purified according to previously published

protocols [17]. DNA was sequenced using a 454 GS Junior sequencer with FLX Titanium chemistry and long-read modifications [36]. For metagenomic sequencing, we confirmed DNA was not sheared by analyzing 200 ng using gel electrophoresis. Genomic DNA libraries were prepared, indexed, and sequenced together on one chip on an Illumina MiSeq Sequencer with 2x300 bp read chemistry by the University of Wisconsin-Madison Biotechnology Center.

#### ***5.6.7 RNA extraction and sequencing***

Total community RNA was extracted from large volume cultures of enrichment line 1A at transfer 14 and lines 2B and 3A at transfer 20 using previously published protocols [35]. The Ribo-Zero Gold rRNA Removal kit (Epidemiology) was used to remove rRNA (Epicenter, Madison, WI). cDNA libraries were constructed and Illumina HiSeq2500 Rapid Run 2x150 bp technology was used to sequence all three samples on one lane at the University of Wisconsin-Madison Biotechnology Center.

#### ***5.6.8 16S rRNA gene sequencing analysis***

Reads were analyzed using mothur version 1.35.1 [37]. The data for enrichment lines across transfers (long-term time series) were analyzed separately from the data for changes in enrichment line 2B transfer 70 across 7 days (short-term time series). For each dataset, reads were filtered to remove those without exact matches to barcode or primers, denoised, and trimmed to a maximum length of 200 bp. Reads were aligned to the Silva database v.119 using the default kmer-based search methods, and reads were removed that did not overlap the region of interest [38]. In the long-term time series data, 641 potential chimeras out of 3169 unique sequences were identified using default settings in UCHIME v.4.2.40 and removed in mothur

[38,39]; in the short-term time series 610 potential chimeras out of 2406 unique sequences were removed. Sequences were classified using a mothur-formatted version of Ribosomal Database Project (RDP) training set 9 with a cut-off of 60% identity [40]. All OTUs were constructed at a 97% similarity threshold, and OTUs from the short-term time series are differentiated with an “S” (e.g. “OTUS1”). Alpha and beta diversity analyses were performed in mothur using data subsampled to 1018 reads in the long-term time series and 1435 reads in the short-term time series. The statistical analyses of alpha diversity indices and OTU abundance with cellulolytic ability and time were performed in JMP. For the long-term time series, the Spearman correlation of cellulolytic ability with the NMDS axes was determined using the `corr.axis` command in mothur.

#### ***5.6.9 Analysis of OTU succession over time***

Changes in OTU abundance between enrichment lines was represented as a heat map, using samples subsampled to 1018 reads. OTUs were included if they were present in the co-occurrence network (see below) and/or represented at least 10% of reads in at least one sample. OTUs were sorted based on their abundance profiles using Ward’s hierarchical clustering of rarefied data in JMP Pro 11 (SAS, Cary, NC). The shifts within OTU4 (*Cellvibrio*) were determined by extracting all reads clustered into this OTU using the `bin.seqs` command in mothur. Reads were sorted based on enrichment line then clustered at 100% identity with CD-HIT-est [41–43].

#### ***5.6.10 Co-occurrence analysis among enrichment lines across transfers***

We used CoNet beta [44] and Cytoscape v.3.3.0 [45] to produce co-occurrence networks of the

using 97% OTU shared abundance tables from mothur. OTUs were only included if their abundance across samples was equal to or greater than 100 and if they were present in at least 5 samples. OTU abundance in each sample was normalized proportionally. The transfer number and days to degrade the cellulose filter paper were matched to each sample as features. Pearson, Spearman, and Kendall correlation analyses were performed with a threshold of  $\pm 0.5$ . Correlations were retained if they were significant in at least two of the methods, and the mean correlation was used for downstream analyses. For each correlation analysis method, p-values were calculated using the “edgeScores” routine, which performs 1000 re-shufflings of rows (OTUs) to determine a random distribution of edge scores. The minimum p-value was maintained across methods (Simes method), and renormalization was used to correct for compositionality bias [44]. Multiple test correction was performed using the Benjamini–Hochberg method with a p-value threshold of 0.05.

#### **5.6.11 Phylogenetic analysis of *Cellvibrio* populations**

Clusters from CD-HIT-est analysis (see above) with >3 representatives were aligned with the 16S rRNA gene sequences for known *Cellvibrio* type strains, the 16S rRNA gene sequence for isolate *C. sp.* 3A-T50a, and the 16S rRNA gene sequences of *Cellvibrio* in the metagenomes produced by EMIRGE (see below) in MEGA6 using Muscle with default parameters [46]. A trimmed alignment was used to build a maximum likelihood tree using RAxML-HPC2 on XSEDE through Cipres with rapid bootstrapping [47,48].

#### **5.6.12 Metagenome assembly**

The metAMOS pipeline was used to remove adapters and trim reads using ea-utils, to assemble



reads using IDBA-UD, and to call open reading frames using MetaGeneMark [49–53]. Each sample was assembled independently. To assign taxonomy, scaffolds were analyzed using PhymmBL, v4.0 [54,55]. Because many of our scaffolds were relatively long, we only used BLAST results from PhymmBL. Taxonomic assignments with a BLAST score  $<1 \times 10^{-100}$  were retained; otherwise, taxonomy was marked as “undefined”.

### ***5.6.13 Composite genome assembly and taxonomic analysis***

Composite genomes were constructed using VizBin [56] clustering of assembled contigs from each metagenome individually; default settings were used, except the minimum contig length was 500 bp and contigs were annotated with their coverage. CheckM [57] was used to evaluate the completeness and contamination of each composite genome. Composite genomes were discarded if they were less than 50% complete or more than 5% contaminated. The trimmed reads comprising each composite genome were extracted using a custom python script, and these reads were reassembled using SPAdes with parameters `–careful` and `–only-assembler` [58]. Quast v3.1 was used to compare original and reassembled genomes, and the version with higher n50 was chosen for further analyses [59]. Taxonomy of composite genomes was assigned by annotating ribosomal proteins using phylosift v.1.0.1 and viewed using Archaeopteryx v.0.9901 beta [60,61]. Open reading frames were assigned using Prodigal [62].

### ***5.6.14 Classification of 16S rRNA genes in metagenomic data***

To classify 16S rRNA genes and their abundance in the metagenomes, we removed adapters paired-end metagenome reads using command `ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:1:true` and trimmed reads using commands `LEADING:3 TRAILING:3 MAXINFO:40:0.25` and

MINLEN:36 in Trimmomatic v0.32 [63]. Between 96% and 98% of reads survived this trimming. These trimmed paired-end reads were analyzed using EMIRGE with sixty iterations using the default Silva database [38,64,65].

#### ***5.6.15 Metatranscriptome mapping***

Using Trimmomatic v0.32, adapters were removed and paired-end metatranscriptome reads were trimmed using the same commands as above, but with MINLEN:70 [63]. BWA was used to map trimmed reads to assembled metagenomes and to composite genomes using default parameters [66]. Then, samtools was used to produce a bam file and produce alignment statistics [67]. Lastly, R packages Rsamtools, GenomicFeatures, and GenomicAlignments were used to count reads [68,69].

#### ***5.6.16 Annotation of metagenomes and composite genomes***

Hidden Markov models for TIGRfam, resfam, antiSMASH, NRPSsp, and Pfam families were used to assign taxonomy to open reading frames [70–74]. Protein sequences were assigned CAZy annotations with a custom script that combines BLAST searching and Pfam searching against the CAZy database [55,75,76], KEGG annotations using BlastKOALA for composite genomes and GhostKOALA for metagenomes searched against the genus\_prokaryotes database [77], and SEED annotations using the desktop version of myRAST with default parameters [78].

Relative abundance of cellulases and  $\beta$ -glucosidases in the metagenomes was determined by summing the coverage of each gene's corresponding scaffold in the metagenome assembly and normalizing sums to the total coverage of all scaffolds in each metagenome. Relative abundance

of cellulases and  $\beta$ -glucosidases in the metatranscriptomes was determined by summing the number of reads mapping to each gene, normalized to the total number of reads mapped to each sample. Functional classifications of highly expressed genes in each composite genome was hand curated and normalized to total number of mapped reads in each metatranscriptome. Analysis of biosynthetic gene clusters (BGCs) in the 3A.1 *Cellvibrio* composite genome was performed by predicting secondary metabolic content using antiSMASH v.3.0.5 and the following pipeline [72]. Biosynthetic signatures of known BGCs were identified in the composite genome by tblastn and putative BGCs were called as +/- 30kb from each group of signatures [79]. Putative BGCs were annotated against NCBI's conserved domain database and BGC open reading frames were compared via blastp to the MIBiG database of known BGCs [79–81].

#### **5.6.17 *Cellvibrio* isolation and 16S rRNA gene sequencing**

The freezer stock from the 3A enrichment line at transfer 50 was grown in an M63 filter paper test tube and dilution plated onto plates containing carbon-free M63 minimal media with 15% agar (w/v) overlaid with CEL1 media consisting of 1 g ammonium sulfate, 5 g Sigmacell-20 (crystalline cellulose; Sigma-Aldrich, St. Louis, MO), and 10 g of agar in 1 L, pH 7.2 [17]. Colonies capable of clearing the cellulose overlay were isolated onto 15% agar plates containing M63 minimal media with 5 g/L glucose and taxonomically classified using Sanger sequencing of the 16S rRNA gene [17].

## **5.7 References**

1. Matulich KL, Martiny JBH. Microbial composition alters the response of litter

- decomposition to environmental change. *Ecology*. 2015;96: 154–163. doi:10.1890/14-0357.1
2. Deng YJ, Wang SY. Synergistic growth in bacteria depends on substrate complexity. *J Microbiol*. 2016;54: 23–30. doi:10.1007/s12275-016-5461-9
  3. Kato S, Haruta S, Cui ZJ, Ishii M, Igarashi Y. Effective cellulose degradation by a mixed-culture system composed of a cellulolytic *Clostridium* and aerobic non-cellulolytic bacteria. *FEMS Microbiol Ecol*. 2004;51: 133–42. doi:10.1016/j.femsec.2004.07.015
  4. Tiunov A V., Scheu S. Facilitative interactions rather than resource partitioning drive diversity-functioning relationships in laboratory fungal communities. *Ecol Lett*. 2005;8: 618–625. doi:10.1111/j.1461-0248.2005.00757.x
  5. Hattenschwiler S, Tiunov A, Scheu S. Biodiversity and litter decomposition in terrestrial ecosystems. *Annu Rev Ecol Evol*. 2005;36: 191–218. doi:10.1146/annurev.ecolsys.36.112904.151932
  6. Strickland MS, Lauber C, Fierer N, Mark A. Bradford CL. Testing the functional significance of microbial community composition. *Ecology*. 2009;90: 441–451. doi:10.1890/08-0296.1
  7. Gessner MO, Swan CM, Dang CK, McKie BG, Bardgett RD, Wall DH, et al. Diversity meets decomposition. *Trends Ecol Evol*. Elsevier Ltd; 2010;25: 372–380. doi:10.1016/j.tree.2010.01.010
  8. Peter H, Beier S, Bertilsson S, Lindström ES, Langenheder S, Tranvik LJ. Function-specific response to depletion of microbial diversity. *ISME J*. 2011;5: 351–61. doi:10.1038/ismej.2010.119
  9. Murray JL, Connell JL, Stacy A, Turner KH, Whiteley M. Mechanisms of synergy in

- polymicrobial infections. *J Microbiol.* 2014;52: 188–199. doi:10.1007/s12275-014-4067-3
10. Kaiser C, Franklin O, Dieckmann U, Richter A. Microbial community dynamics alleviate stoichiometric constraints during litter decay. *Ecol Lett.* 2014;17: 680–690. doi:10.1111/ele.12269
  11. Brune A. Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol.* Nature Publishing Group; 2014; doi:10.1038/nrmicro3182
  12. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev.* 2002;66: 506–577. doi:10.1128/MMBR.66.3.506
  13. DeAngelis KM, Gladden JM, Allgaier M, D’haeseleer P, Fortney JL, Reddy A, et al. Strategies for enhancing the effectiveness of metagenomic-based enzyme discovery in lignocellulolytic microbial communities. *BioEnergy Res.* 2010;3: 146–158. doi:10.1007/s12155-010-9089-z
  14. DeAngelis KM, Allgaier M, Chavarria Y, Fortney JL, Hugenholtz P, Simmons B, et al. Characterization of trapped lignin-degrading microbes in tropical forest soil. *PLoS One.* 2011;6: e19306. doi:10.1371/journal.pone.0019306
  15. Allgaier M, Reddy A, Park JI, Ivanova N, D’haeseleer P, Lowry S, et al. Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One.* 2010;5: e8812. doi:10.1371/journal.pone.0008812
  16. Reddy AP, Simmons CW, D’haeseleer P, Khudyakov J, Burd H, Hadi M, et al. Discovery of Microorganisms and Enzymes Involved in High-Solids Decomposition of Rice Straw Using Metagenomic Analyses. Liles MR, editor. *PLoS One.* 2013;8: e77985. doi:10.1371/journal.pone.0077985

17. Lewin GR, Johnson AL, Moreira Soto RD, Perry K, Book AJ, Horn HA, et al. Cellulose-Enriched Microbial Communities from Leaf-Cutter Ant (*Atta colombica*) Refuse Dumps Vary in Taxonomic Composition and Degradation Ability. PLoS One. 2016;11: e0151840. doi:10.1371/journal.pone.0151840
18. Scott JJ. Microbial community ecology of leaf-cutter ant refuse dumps. 2011.
19. Haines BL. Element and Energy Flows through Colonies of the Leaf-Cutting Ant, *Atta colombica*, in Panama. Biotropica. 1978;10: 270–277.
20. Hudson TM, Turner BL, Herz H, Robinson JS. Temporal patterns of nutrient availability around nests of leaf-cutting ants (*Atta colombica*) in secondary moist tropical forest. Soil Biol Biochem. Elsevier Ltd; 2009;41: 1088–1093. doi:10.1016/j.soilbio.2009.02.014
21. Scott JJ, Budsberg KJ, Suen G, Wixon DL, Balser TC, Currie CR. Microbial community structure of leaf-cutter ant fungus gardens and refuse dumps. PLoS One. 2010;5: e9922. doi:10.1371/journal.pone.0009922
22. Gardner JG, Keating DH. Requirement of the type II secretion system for utilization of cellulosic substrates by *Cellvibrio japonicus*. Appl Environ Microbiol. 2010;76: 5079–87. doi:10.1128/AEM.00454-10
23. Bakalidou A, Kampfer P, Berchtold M, Kuhnigk T, Wenzel M, König H. *Cellulosimicrobium variabile* sp. nov., a cellulolytic bacterium from the hindgut of the termite *Mastotermes darwiniensis*. Int J Syst Evol Microbiol. 2002;52: 1185–1192. doi:10.1099/ijs.0.01904-0.Abbreviations
24. Schmid M, Baldani JJ, Hartmann A. The genus *Herbaspirillum*. In: Dworkin M, Falkow S, Schleifer K-H, Rosenberg E, Stackebrandt E, editors. The Prokaryotes. 3rd ed. New York: Springer; 2006. pp. 141–150. doi:10.1007/0-387-30745-1\_7

25. Humphry DR, Black GW, Cummings SP. Reclassification of “*Pseudomonas fluorescens* subsp. *cellulosa*” NCIMB 10462 (Ueda *et al.* 1952) as *Cellvibrio japonicus* sp. nov. and revival of *Cellvibrio vulgaris* sp. nov., nom. rev. and *Cellvibrio fulvus* sp. nov., Int J Syst Evol Microbiol. 2003;53: 393–400. doi:10.1099/ijs.0.02271-0
26. Han AW, Sandy M, Fishman B, Trindade-Silva AE, Soares CAG, Distel DL, et al. Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont *Teredinibacter turnerae* T7901. PLoS One. 2013;8. doi:10.1371/journal.pone.0076151
27. Hooper DU, Chapin III FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, et al. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. Ecol Monogr. 2005;75: 3–35. doi:10.1890/04-0922
28. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. Nature Publishing Group; 2012;486: 207–14. doi:10.1038/nature11234
29. Reed HE, Martiny JBH. Testing the functional significance of microbial composition in natural communities. FEMS Microbiol Ecol. 2007;62: 161–170. doi:10.1111/j.1574-6941.2007.00386.x
30. Salles JF, Poly F, Schmid B, Le Roux X. Community niche predicts the functioning of denitrifying bacterial assemblages. Ecology. 2009;90: 3324–3332. doi:10.1890/09-0188.1
31. Bell T, Newman JA, Silverman BW, Turner SL, Lilley AK. The contribution of species richness and composition to bacterial services. Nature. 2005;436: 1157–1160. doi:10.1038/nature03891
32. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Wendt-Pienkowski E, Doering DT, et al. Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of

- expanded gene content and coordinated gene expression. PLOS Biol.
33. Hutchinson G. The paradox of the plankton. Am Nat. 1961;95: 137–145. Available: <http://www.jstor.org/stable/2458386>
  34. Seth EC, Taga ME. Nutrient cross-feeding in the microbial world. Front Microbiol. 2014;5: 1–6. doi:10.3389/fmicb.2014.00350
  35. Book AJ, Lewin GR, McDonald BR, Takasuka TE, Doering DT, Adams AS, et al. Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically-linked capacity for the degradation of lignocellulose. Appl Environ Microbiol. 2014;80: 4692–4701. doi:10.1128/AEM.01133-14
  36. Hanshew AS, Mason CJ, Raffa KF, Currie CR. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. J Microbiol Methods. Elsevier B.V.; 2013;95: 149–55. doi:10.1016/j.mimet.2013.08.007
  37. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75: 7537–41. doi:10.1128/AEM.01541-09
  38. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35: 7188–7196. doi:10.1093/nar/gkm864
  39. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27: 2194–2200.



- doi:10.1093/bioinformatics/btr381
40. Cole JR, Wang Q, Fish J a., Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 2014;42: 633–642. doi:10.1093/nar/gkt1244
  41. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;26: 680–682.  
doi:10.1093/bioinformatics/btq003
  42. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28: 3150–3152. doi:10.1093/bioinformatics/bts565
  43. Li W, Godzik A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22: 1658–1659.  
doi:10.1093/bioinformatics/btl158
  44. Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol.* 2012;8: e1002606. doi:10.1371/journal.pcbi.1002606
  45. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 2498–2503. doi:10.1101/gr.1239303.metabolite
  46. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30: 2725–2729.  
doi:10.1093/molbev/mst197
  47. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gatew Comput Environ Work GCE 2010. 2010;

- doi:10.1109/GCE.2010.5676129
48. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–1313.  
doi:10.1093/bioinformatics/btu033
  49. Treangen TJ, Koren S, Sommer DD, Liu B, Astrovskaya I, Ondov B, et al. MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biol. BioMed Central Ltd*; 2013;14: R2. doi:10.1186/gb-2013-14-1-r2
  50. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28: 1420–8. doi:10.1093/bioinformatics/bts174
  51. Besemer J, Borodovsky M. Heuristic approach to deriving models for gene finding. *Nucleic Acids Res*. 1999;27: 3911–3920. doi:10.1093/nar/27.19.3911
  52. Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res*. 2010;38: 1–15. doi:10.1093/nar/gkq275
  53. Aronesty E. ea-utils: “Command-line tools for processing biological sequencing data.” 2011.
  54. Brady A, Salzberg S. PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nat Methods. Nature Publishing Group*; 2011;8: 367.  
doi:10.1038/nmeth0511-367
  55. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10: 421. doi:10.1186/1471-2105-10-421
  56. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian H, et al. VizBin

- an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome*. 2015; 1–7. doi:10.1186/s40168-014-0066-1
- 57. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. 2015; 1–13.
- 58. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol*. 2012;19: 455–477. doi:10.1089/cmb.2012.0021
- 59. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29: 1072–1075. doi:10.1093/bioinformatics/btt086
- 60. Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*. 2014;2: e243. doi:10.7717/peerj.243
- 61. Han M V, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*. 2009;10: 356. doi:10.1186/1471-2105-10-356
- 62. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11: 119. doi:10.1186/1471-2105-11-119
- 63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
- 64. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol*. 2011;12: R44. doi:10.1186/gb-2011-12-5-r44

65. Miller CS, Handley KM, Wrighton KC, Frischkorn KR, Thomas BC, Banfield JF. Short-Read Assembly of Full-Length 16S Amplicons Reveals Bacterial Diversity in Subsurface Sediments. *PLoS One*. 2013;8. doi:10.1371/journal.pone.0056018
66. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324
67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352
68. Morgan M, Pagès H, Obenchain V, Hayden N. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. 2016.
69. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9: 1–10. doi:10.1371/journal.pcbi.1003118
70. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res*. 2013;41: 387–395. doi:10.1093/nar/gks1234
71. Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J. Nature Publishing Group*; 2014;9: 1–10. doi:10.1038/ismej.2014.106
72. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, et al. antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res*. 2013;41: 204–212. doi:10.1093/nar/gkt449
73. Prieto C, Garcia-Estrada C, Lorenzana D, Martin JF. NRPSp: Non-ribosomal peptide

- synthase substrate predictor. *Bioinformatics*. 2012;28: 426–427.  
doi:10.1093/bioinformatics/btr659
74. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2015;44: gkv1344. doi:10.1093/nar/gkv1344
  75. Takasuka TE, Book AJ, Lewin GR, Currie CR, Fox BG. Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Sci Rep*. 2013;3: 1–10.  
doi:10.1038/srep01030
  76. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42: D490–5. doi:10.1093/nar/gkt1178
  77. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol. The Authors*; 2015;428: 726–731. doi:10.1016/j.jmb.2015.11.006
  78. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 2014;42: 206–214. doi:10.1093/nar/gkt1226
  79. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. *J Mol Biol*. 1990;215: 403–410. doi:10.1016/S0022-2836(05)80360-2
  80. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, et al. The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification. *Nat Chem Biol*. 2015;11: 625–631. doi:10.1038/nchembio.1890
  81. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD:

NCBI's conserved domain database. *Nucleic Acids Res.* 2015;43: D222–D226.

doi:10.1093/nar/gku1221

## Chapter 6: Summary and Future Outlook

In this dissertation, I explored the effects of interactions on the ecology and evolution of cellulose degradation. This work helps demonstrate that associations between microbes and eukaryotic hosts can influence the evolution of cellulolytic ability and that interactions within bacterial communities can influence the rate of cellulose degradation. These results are important in understanding and predicting the rate of plant biomass degradation in the environment, which directly influences the rate carbon dioxide release to the atmosphere. Additionally, the highly cellulolytic communities and isolates discussed here contribute towards the production of sustainable biofuels and bioproducts from cellulosic biomass.

In Chapter 1, I argued that the evolutionary history and ecological importance of plant biomass degradation and secondary metabolite production in Actinobacteria makes this phylum amenable to produce valuable biofuels and bioproducts. Then, in Chapters 2 and 3, I presented data showing that two clades of insect-associated *Streptomyces* have evolved the ability to rapidly deconstruct biomass through selectively acquiring and then regulating key cellulase genes. I hypothesize that herbivorous insect hosts selected for these phylogenetically related strains to be able to degrade cellulose in isolation, while soil associated *Streptomyces*, which cannot rapidly degrade cellulose in isolation, were potentially selected to contribute to plant biomass degradation within a larger community of microbes. To investigate this hypothesis, future work must establish the role of cellulolytic *Streptomyces* with their eukaryotic hosts and determine if soil associated strains express cellulases when interacting with other organisms. Furthermore, a more comprehensive understanding of the evolution of regulatory pathways, in this system or others, is important for understanding the processes that shape cellulolytic ability in these taxa

over evolutionary timescales. Additionally, to connect with Chapter 1, future work should focus on engineering these highly cellulolytic *Streptomyces* for the cellulosic biofuels industry.

This work with *Streptomyces* emphasized the importance of studying cellulolytic organisms in the context of a larger community. However, natural cellulose degrading communities are extremely complex, preventing the analysis of interactions between microbes. In Chapters 4 and 5, I used an enrichment method to simplify microbial communities from leaf-cutter ant refuse dumps, with the goal to understand how community structure and community interactions influence cellulose degradation. Specifically, in Chapter 4, I found that cellulose-degrading communities vary between refuse dumps, but that some communities are extremely capable at degrading cellulose. In Chapter 5, I used long-term enrichments to demonstrate that cellulolytic ability inversely correlates with community diversity and that competitive interactions are dominant in enriched communities. While I only studied the ecological changes in these communities over sixty transfers in Chapter 5, these enrichment lines have now been transferred approximately 200 times, representing over 900 generations. Future work should characterize the interplay between evolutionary and ecological changes in these communities. I am especially interested in how community interactions influence the accumulation of mutations over time in these enrichments. To explore this question, it would be exciting to perform experimental evolution of select community members in isolation and then compare differences in the evolutionary trajectories of these populations with the corresponding populations within the larger community. I found that within communities, a significant portion of cellular resources was used to mediate interactions. Therefore, I hypothesize that populations within the larger community would acquire unique beneficial mutations that facilitate community interactions.



While I have provided a few specific experiments above to address the questions raised by the data chapters of my dissertation, there is still much to be learned in general about how microbes and microbial communities degrade plant biomass. Future work should focus on understanding this process at the level of the microbial cell. The abiotic and biotic interactions of cellulose degrading organisms occur on a nanometer scale, but most work, including the work presented here, does not address how spatial distribution and scale influence plant biomass degradation and the interactions driving this process. Additionally, microbial plant biomass degradation is currently considered a “black box” in most climate models. The factors controlling the rate of plant biomass degradation must be more comprehensively studied to allow for accurate modeling of climate change.

Beyond plant biomass degradation, a broader range of work is necessary to understand the influence of interspecies interactions on community structure and function and on the evolution of organisms. This work fits with a need to analyze the function and roles of environmental microbes, not just their biogeographic distributions. Further, this research can help answer questions in microbial ecology and evolution from the origins of microbial biodiversity to how microbial communities are so diverse in the midst of competition for niche space. The challenges of this work are large, but as the importance of microbes and microbial communities is becoming increasingly recognized with the renaissance of the microbiome, microbiologists have an increasing diversity of tools to tackle these fundamental questions. Hypothesis-driven studies that test theories and work to develop ecological and evolutionary models in microbes have the potential to revolutionize biology.

**Appendix 1: Supplementary Material for Chapter 1**

**In review as Supplementary Material:** Gina R. Lewin, Camila Carlos, Marc G. Chevrete, Heidi A. Horn, Bradon R. McDonald, Robert J. Stankey, Brian G. Fox, and Cameron R. Currie. Evolution and Ecology of Actinobacteria and their Bioenergy Applications. *Annual Reviews of Microbiology*.

**Author contributions are the same as in Chapter 1.**

**Table A1.1.** MG-RAST datasets and genus distributions used to calculate average abundances in Figure 1.1.

| MG-RAST ID | Type            | Project name                                                                                                                | Habitat                                           | Sequencing type | bp count  |
|------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|-----------------|-----------|
| 4455879.3  | Freshwater      | mladenov_bangladesh_arsenic                                                                                                 | Large river delta biome                           |                 | 364362    |
| 4462963.3  | Freshwater      | McMahon Temperate Bog Lake Timeseries                                                                                       | Temperate Freshwater                              |                 | 2280000   |
| 4462964.3  | Freshwater      | McMahon Temperate Bog Lake Timeseries                                                                                       | Temperate Freshwater                              |                 | 5046900   |
| 4463801.3  | Freshwater      | McMahon Temperate Bog Lake Timeseries                                                                                       | Temperate Freshwater                              |                 | 2494500   |
| 4490049.3  | Freshwater      | Three sediment comparison                                                                                                   | large lake biome                                  | illumina        | 4392464   |
| 4508183.3  | Freshwater      | Lake water, lake sediments, soils, biofilms and subsurface environments in China                                            | freshwater biome                                  | 454             | 2748771   |
| 4508184.3  | Freshwater      | Lake water, lake sediments, soils, biofilms and subsurface environments in China                                            | freshwater biome                                  | 454             | 3110812   |
| 4537448.3  | Freshwater      | Biofoam                                                                                                                     | Small lake biome                                  | 454             | 35499564  |
| 4556435.3  | Freshwater      | Sendefo : Impact of tebuconazole on freshwater bacteria diversity                                                           | Small river biome                                 | 454             | 19174743  |
| 4444448.3  | Host-associated | Oral Metagenome - Human                                                                                                     | animal-associated habitat                         | illumina        | 73001821  |
| 4449917.3  | Host-associated | animal-1                                                                                                                    | animal-associated habitat                         | 454             | 14734838  |
| 4456342.3  | Host-associated | frierer_undergraduate_palms                                                                                                 | human-associated habitat                          |                 | 718281    |
| 4471952.3  | Host-associated | Bacterial biogeography of the human digestive tract                                                                         | terrestrial biome                                 | illumina        | 85207442  |
| 4479832.3  | Host-associated | Sponge symbiont microbial communities (16S)                                                                                 | animal-associated habitat                         | 454             | 8040706   |
| 4479834.3  | Host-associated | Sponge symbiont microbial communities (16S)                                                                                 | animal-associated habitat                         | 454             | 8904126   |
| 4483775.3  | Host-associated | Project for: 16S rumen bacterial community 16 holstein dairy cows                                                           | animal-associated habitat                         | 454             | 66488823  |
| 4489673.3  | Host-associated | Human gut microbiome viewed across age and geography (16S)                                                                  | human-associated habitat                          | 454             | 378182380 |
| 4489818.3  | Host-associated | Human gut microbiome viewed across age and geography (16S)                                                                  | human-associated habitat                          | 454             | 133429476 |
| 4490253.3  | Host-associated | Fat-induced taurocholic acid mouse gut study                                                                                | terrestrial biome                                 | 454             | 518928    |
| 4514864.3  | Host-associated | bovine rumen colonization assay                                                                                             | Warm deserts and semideserts                      | 454             | 22327498  |
| 4514865.3  | Host-associated | bovine rumen colonization assay                                                                                             | Warm deserts and semideserts                      | 454             | 20716594  |
| 4533580.3  | Host-associated | Deep_Sea_Sponge                                                                                                             | marine bathyal zone biome                         | 454             | 2239527   |
| 4536055.3  | Host-associated | F.G.Termite.Guts.Otani.et.al.                                                                                               | flooded savanna biome                             | 454             | 4329628   |
| 4536057.3  | Host-associated | F.G.Termite.Guts.Otani.et.al.                                                                                               | flooded savanna biome                             | 454             | 10335360  |
| 4536060.3  | Host-associated | F.G.Termite.Guts.Otani.et.al.                                                                                               | flooded savanna biome                             | 454             | 8439802   |
| 4536061.3  | Host-associated | F.G.Termite.Guts.Otani.et.al.                                                                                               | flooded savanna biome                             | 454             | 11167314  |
| 4452734.3  | Marine          | Western English Channel Time Series                                                                                         |                                                   |                 | 1230568   |
| 4452737.3  | Marine          | Western English Channel Time Series                                                                                         |                                                   |                 | 365530    |
| 4468742.3  | Marine          | Ocean acidification mesocosm 2006                                                                                           | Marine Biome                                      |                 | 10024950  |
| 4477060.3  | Marine          | CaporasilluminaPNAS2011_3prime                                                                                              | marine water body                                 | 454             | 18925570  |
| 4490056.3  | Marine          | Three sediment comparison                                                                                                   | mangrove biome                                    | illumina        | 3828783   |
| 4497999.3  | Marine          | Microbial Diversity Of Deep Sea Sponges                                                                                     | oceanic bathypelagic zone biome                   | 454             | 2271141   |
| 4509242.3  | Marine          | Analysis of composition and structure of coastal to mesopelagic bacterioplankton communities in the northern Gulf of Mexico | marine biome                                      | 454             | 3520221   |
| 4509245.3  | Marine          | Analysis of composition and structure of coastal to mesopelagic bacterioplankton communities in the northern Gulf of Mexico | marine biome                                      | 454             | 3599025   |
| 4509247.3  | Marine          | Analysis of composition and structure of coastal to mesopelagic bacterioplankton communities in the northern Gulf of Mexico | marine biome                                      | 454             | 3766971   |
| 4533576.3  | Marine          | Deep_Sea_Sponge                                                                                                             | marine bathyal zone biome                         | 454             | 4987917   |
| 4455655.3  | Soil            | rousk_ph_gradient                                                                                                           | soil                                              |                 | 420728    |
| 4455738.3  | Soil            | lauber_88_soils                                                                                                             | Temperate broadleaf and mixed forest biome, soil  | 454             | 395504    |
| 4455781.3  | Soil            | lauber_88_soils                                                                                                             | Temperate grasslands, soil                        | 454             | 336574    |
| 4455782.3  | Soil            | lauber_88_soils                                                                                                             | soil, forest                                      | 454             | 374839    |
| 4455783.3  | Soil            | lauber_88_soils                                                                                                             | Tropical humid forests, soil                      | 454             | 415490    |
| 4468483.3  | Soil            | Temple_TX_Native_Exotic_Holmoeckel                                                                                          | temperate grassland                               |                 | 22397250  |
| 4468693.3  | Soil            | Russian Tundra Permafrost                                                                                                   | Tundra                                            | illumina        | 8883150   |
| 4488319.3  | Soil            | Project for: NTS_023_16S_ELE                                                                                                | soil                                              | 454             | 3972470   |
| 4501999.3  | Soil            | Biogeography of biocrust soil bacteria                                                                                      | Cold-winter (continental) deserts and semideserts | 454             | 231361    |
| 4505429.3  | Soil            | Icelandic critical zone                                                                                                     | Temperate grasslands                              | 454             | 1087551   |

**Table A1.2.** Genomes used for phylogentic tree construction and molecular clock analysis for Figure 1.1.

| Actinobacteria Genomes                | Outgroup Genomes                             |
|---------------------------------------|----------------------------------------------|
| Rubrobacter_xylophilus_DSM_9941       | Aquifex_aeolicus_VF5                         |
| Corynebacterium_diphtheriae_CDCE_8392 | Thermotoga_maritima_MSB8                     |
| Streptomyces_coelicolor_A3_2_         | Bacillus_subtilis_subtilis_168               |
| Geodermatophilus_obscurus_DSM_43160   | Lactobacillus_fermentum_IFO_3956             |
| Eggerthella_lenta_DSM_2243            | Enterococcus_faecium_NRRLB-2354              |
| Atopobium_parvulum_DSM_20469          | Fibrobacter_succinogenes_succinogenes_S85    |
| Kineococcus_radiotolerans_SRS30216    | Prevotella_denticola_F0289                   |
| Pseudonocardia_dioxanivorans_CB1190   | Bacteroides_vulgatus_ATCC8482                |
| Mycobacterium_tuberculosis_CDC1551    | Gloeobacter_violaceus_PCC_7421               |
| Rothia_dentocariosa_ATCC_17931        | Synechococcus_elongatus_PCC_6301             |
| Conexibacter_woesei_DSM_14684         | Prochlorococcus_marinus_MIT_9312             |
| Propionibacterium_propionicum_F0230a  | Microcystis_aeruginosa_NIES-843              |
| Amycolatopsis_orientalis_HCCB10007    | Sinorhizobium_fredii_NGR234                  |
| Acidothermus_cellulolyticus_11B       | Burkholderia_cenocepacia_HI2424              |
| Nocardioides_JS614                    | Vibrio_cholerae_O1_El_Tor_N16961             |
| Micromonospora_aurantiaca_ATCC_27029  | Escherichia_coli_K-12_subMG1655              |
| Kribbella_flavida_DSM_17836           | Salmonella_enterica_enterica_Typhimurium_LT2 |
| Salinispora_tropica_CNB_440           | Deinococcus_radiodurans_R1                   |
| Microbacterium_testaceum_StLB037      |                                              |
| Nocardia_brasiliensis_ATCC_700358     |                                              |
| Rhodococcus_equi_103S                 |                                              |
| Nakamurella_multipartita_DSM_44233    |                                              |
| Cellulomonas_fimi_ATCC_484            |                                              |
| Kocuria_rhizophila_DC2201             |                                              |
| Sanguibacter_keddiei_DSM_10542        |                                              |
| Brachybacterium_faecium_DSM_4810      |                                              |
| Acidimicrobium_ferrooxidans_DSM_10331 |                                              |
| Micrococcus_luteus_NCTC_2665          |                                              |
| Microthricus_phosphovorus_NM_1        |                                              |
| Leifsonia_xyli_CTCB07                 |                                              |
| Arthrobacter_aurescens_TC1            |                                              |
| Frankia_EAN1pec                       |                                              |
| Bifidobacterium_animalis_ATCC_25527   |                                              |

**Table A1.3.** CAZy and biosynthetic gene cluster abundance used to calculate data represented in Figure 1.3. Excel table available on request.

**Table A1.4.** Titters of biofuels and bioproducts produced by Actinobacteria compared with other microbes.<sup>a</sup>

| Product               | Organism                                                       | Titer in Actinobacteria                                   | Titer in other microbes                                                           |
|-----------------------|----------------------------------------------------------------|-----------------------------------------------------------|-----------------------------------------------------------------------------------|
| 1-propanol            | <i>Thermobifida fusca</i><br><i>Corynebacterium glutamicum</i> | 0.48 g/L on switchgrass (5,6)<br>0.72 g/L on glucose (23) | 2.78 g/L ( <i>Escherichia coli</i> ) (2)                                          |
| isobutanol            | <i>C. glutamicum</i>                                           | 4.9 g/L on glucose (3, 24)                                | 22 g/L ( <i>E. coli</i> ) (1)<br>1.6 g/L ( <i>Saccharomyces cerevisiae</i> ) (13) |
| fatty acid acyl ester | <i>Streptomyces</i> sp. S161                                   | 0.2 g/L on starch (12)                                    | 1.5 g/L ( <i>E. coli</i> ) (17)<br>0.03 g/L ( <i>S. cerevisiae</i> ) (22)         |
| triacylglycerol       | <i>Rhodococcus opacus</i>                                      | 15.9 g/L on corn stover (10, 11)                          | 0.2–1.0 g/L ( <i>S. cerevisiae</i> ) (8, 20)<br>0.5 g/L ( <i>E. coli</i> ) (7)    |
| bisabolene            | <i>Streptomyces venezuelae</i>                                 | 10 mg/L on glucose (18)                                   | 912 mg/L ( <i>E. coli</i> )<br>994 mg/L ( <i>S. cerevisiae</i> ) (16)             |
| pinene                | <i>C. glutamicum</i>                                           | 176 µg/L on glucose (9)                                   | 32 mg/L ( <i>E. coli</i> ) (21, 26)                                               |
| 4-vinylphenol         | <i>Streptomyces lividans</i>                                   | 0.39 g/L on glucose<br>0.25 g/L on PASC <sup>b</sup> (14) | 0.5 g/L ( <i>Pseudomonas putida</i> ) (25)<br>0.39 g/L ( <i>E. coli</i> ) (19)    |
| benzoate              | <i>Streptomyces maritimus</i>                                  | 0.46 g/L on starch<br>0.13 g/L on PASC <sup>b</sup> (15)  | None reported                                                                     |
| 1,2-propanediol       | <i>C. glutamicum</i>                                           | 4.6 g/L on glucose (23)                                   | 5.6 g/L ( <i>E. coli</i> ) (4)<br>2.19 g/L ( <i>S. cerevisiae</i> )               |

<sup>a</sup> Comparisons made across similar experiments

<sup>b</sup> Phosphoric Acid Swollen Cellulose (amorphous cellulose)

#### References for Table A1.4

1. Atsumi S, Hanai T, Liao JC. 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*. 451(7174):86–89
2. Atsumi S, Liao JC. 2008. Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. *Appl. Environ. Microbiol.* 74(24):7802–8
3. Blombach B, Riester T, Wieschalka S, Ziert C, Youn J-W, et al. 2011. *Corynebacterium glutamicum* tailored for efficient isobutanol production. *Appl. Environ. Microbiol.* 77(10):3300–3310
4. Clomburg JM, Gonzalez R. 2011. Metabolic engineering of *Escherichia coli* for the production of 1,2-propanediol from glycerol. *Biotechnol. Bioeng.* 108(4):867–79
5. Deng Y, Fisher AB, Fong SS. 2015. Systematic analysis of intracellular mechanisms of propanol production in the engineered *Thermobifida fusca* B6 strain. *Appl Microbiol Biotechnol.* 99(19):8089–8100
6. Deng Y, Fong SS. 2011. Metabolic engineering of *Thermobifida fusca* for direct aerobic bioconversion of untreated lignocellulosic biomass to 1-propanol. *Metab. Eng.* 13(5):570–77

7. Janßen HJ, Steinbüchel A. 2014. Production of triacylglycerols in *Escherichia coli* by deletion of the diacylglycerol kinase gene and heterologous overexpression of atfA from *Acinetobacter baylyi* ADP1. *Appl Microbiol Biotechnol.* 98(4):1913–24
8. Kamisaka Y, Kimura K, Uemura H, Yamaoka M. 2013. Overexpression of the active diacylglycerol acyltransferase variant transforms *Saccharomyces cerevisiae* into an oleaginous yeast. *Appl Microbiol Biotechnol.* 97(16):7345–55
9. Kang M-K, Eom J-H, Kim Y, Um Y, Woo HM. 2014. Biosynthesis of pinene from glucose using metabolically-engineered *Corynebacterium glutamicum*. *Biotechnol. Lett.* 36(10):2069–77
10. Kurosawa K, Laser J, Sinskey AJ. 2015. Tolerance and adaptive evolution of triacylglycerol-producing *Rhodococcus opacus* to lignocellulose-derived inhibitors. *Biotechnol Biofuels.* 8:76
11. Kurosawa K, Wewetzer SJ, Sinskey AJ. 2013. Engineering xylose metabolism in triacylglycerol-producing *Rhodococcus opacus* for lignocellulosic fuel production. *Biotechnol Biofuels.* 6(1):134
12. Lu Y, Wang J, Deng Z, Wu H, Deng Q, et al. 2013. Isolation and characterization of fatty acid methyl ester (FAME)-producing *Streptomyces* sp. S161 from sheep (*Ovis aries*) faeces. *Letters in Applied Microbiology.* 57(3):200–205
13. Matsuda F, Ishii J, Kondo T, Ida K, Tezuka H, Kondo A. 2013. Increased isobutanol production in *Saccharomyces cerevisiae* by eliminating competing pathways and resolving cofactor imbalance. *Microb. Cell Fact.* 12:119
14. Noda S, Kawai Y, Tanaka T, Kondo A. 2015. 4-Vinylphenol biosynthesis from cellulose as the sole carbon source using phenolic acid decarboxylase- and tyrosine ammonia lyase-expressing *Streptomyces lividans*. *Bioresour. Technol.* 180:59–65
15. Noda S, Kitazono E, Tanaka T, Ogino C, Kondo A. 2012. Benzoic acid fermentation from starch and cellulose via a plant-like  $\beta$ -oxidation pathway in *Streptomyces maritimus*. *Microb. Cell Fact.* 11:49
16. Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS. 2011. Identification and microbial production of a terpene-based advanced biofuel. *Nat Commun.* 2:483
17. Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. 2012. Microbial engineering for the production of advanced biofuels. *Nature.* 488(7411):320–28

18. Phelan RM, Sekurova ON, Keasling JD, Zotchev SB. 2015. Engineering terpene biosynthesis in *Streptomyces* for production of the advanced biofuel precursor bisabolene. *ACS Synth Biol.* 4(4):393–99
19. Qi WW, Vannelli T, Breinig S, Ben-Bassat A, Gatenby AA, et al. 2007. Functional expression of prokaryotic and eukaryotic genes in *Escherichia coli* for conversion of glucose to p-hydroxystyrene. *Metab. Eng.* 9(3):268–76
20. Rungtaphan W, Keasling JD. 2014. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.* 21:103–13
21. Sarria S, Wong B, García Martín H, Keasling JD, Peralta-Yahya P. 2014. Microbial synthesis of pinene. *ACS Synth Biol.* 3(7):466–75
22. Shi S, Valle Rodríguez JO, Siewers V, Nielsen J. 2014. Engineering of chromosomal wax ester synthase integrated *Saccharomyces cerevisiae* mutants for improved biosynthesis of fatty acid ethyl esters. *Biotechnol. Bioeng.* 111(9):1740–47
23. Siebert D, Wendisch VF. 2015. Metabolic pathway engineering for production of 1,2-propanediol and 1-propanol by *Corynebacterium glutamicum*. *Biotechnol Biofuels.* 8:91
24. Smith KM, Cho K-M, Liao JC. 2010. Engineering *Corynebacterium glutamicum* for isobutanol production. *Appl Microbiol Biotechnol.* 87(3):1045–55
25. Verhoef S, Wierckx N, Westerhof RGM, de Winde JH, Ruijsenaars HJ. 2009. Bioproduction of p-hydroxystyrene from glucose by the solvent-tolerant bacterium *Pseudomonas putida* S12 in a two-phase water-decanol fermentation. *Appl. Environ. Microbiol.* 75(4):931–36
26. Yang J, Nie Q, Ren M, Feng H, Jiang X, et al. 2013. Metabolic engineering of *Escherichia coli* for the biosynthesis of alpha-pinene. *Biotechnol Biofuels.* 6(1):60

**Appendix 2: Supplementary Material for Chapter 2**

**Previously published as Supplementary Material:** Adam J. Book\*, Gina R. Lewin\*, Bradon R. McDonald, Taichi E. Takasuka, Drew T. Doering, Aaron S. Adams, Joshua A.V. Blodgett, Jon Clardy, Kenneth F. Raffa, Brian G. Fox, and Cameron R. Currie; \*authors contributed equally (2014). Cellulolytic *Streptomyces* strains associated with herbivorous insects share a phylogenetically linked capacity to degrade lignocellulose. *Applied and Environmental Microbiology*, 80(15):4692-4701.

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



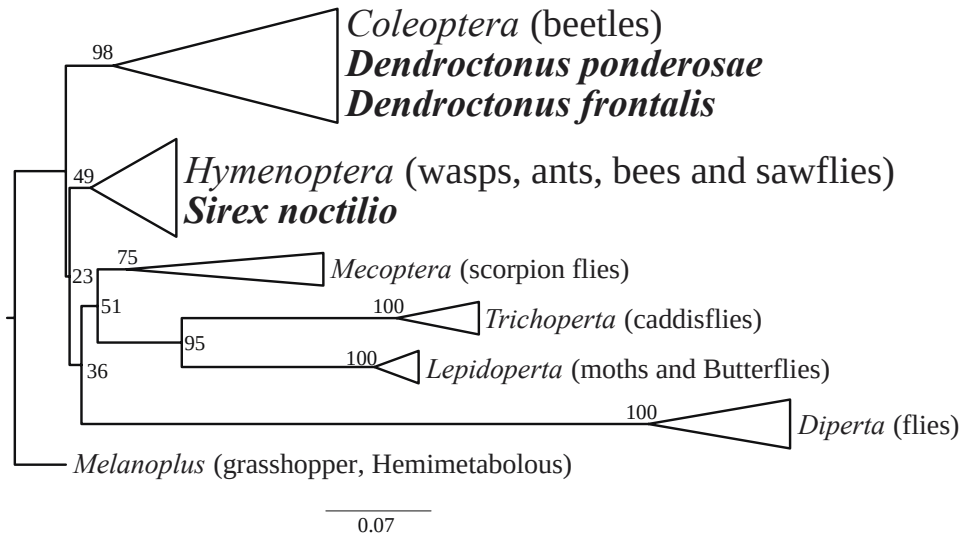


**Figure A2.1. 18S phylogenetic tree of select holometabolous insect phyla.** The maximum likelihood tree was generated from an alignment containing greater than three species per phyla and the hemimetabolous outgroup. Phyla clades were collapsed for clarity. Bootstrap support for the maximum-likelihood analysis is given at each node.

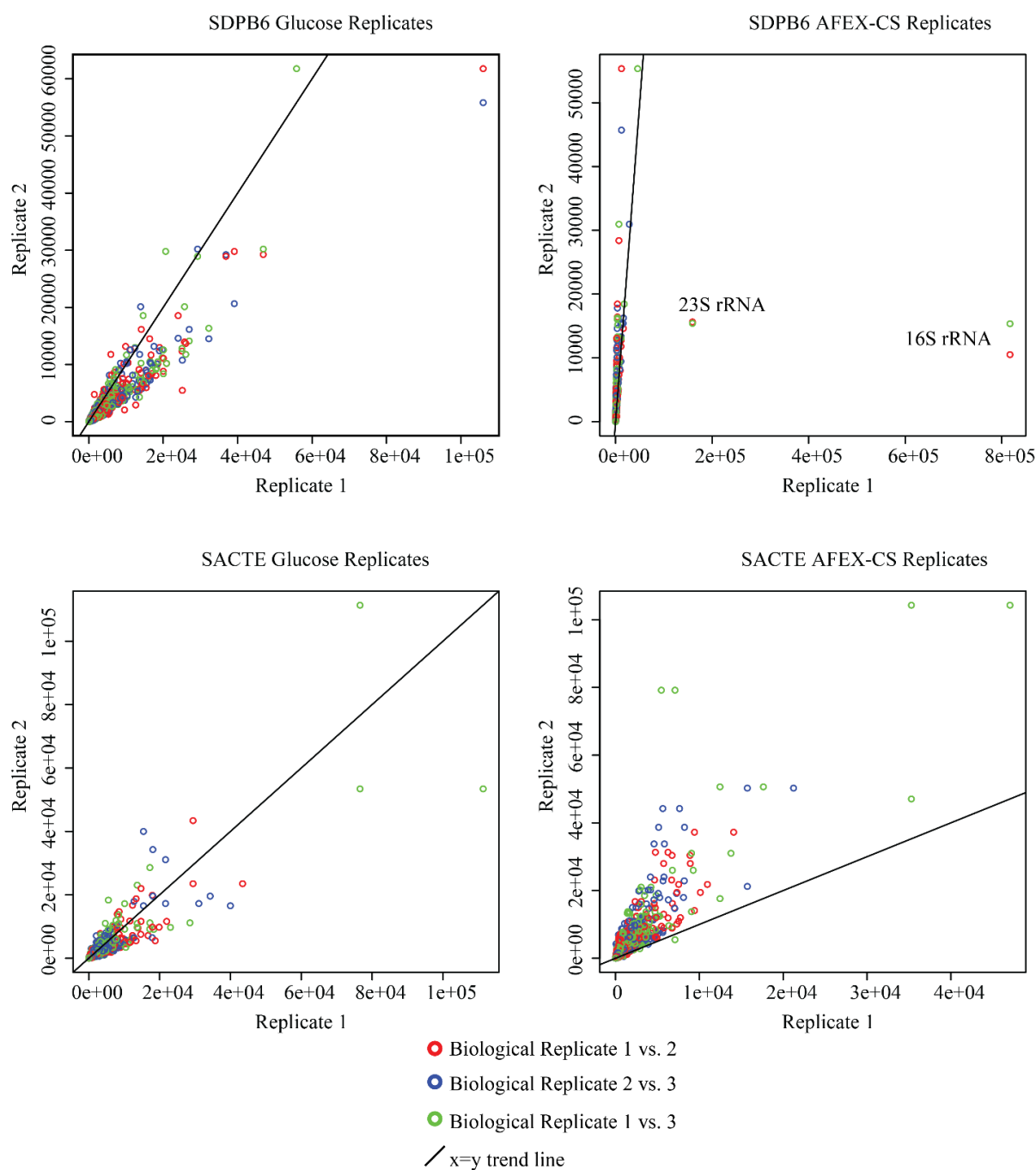
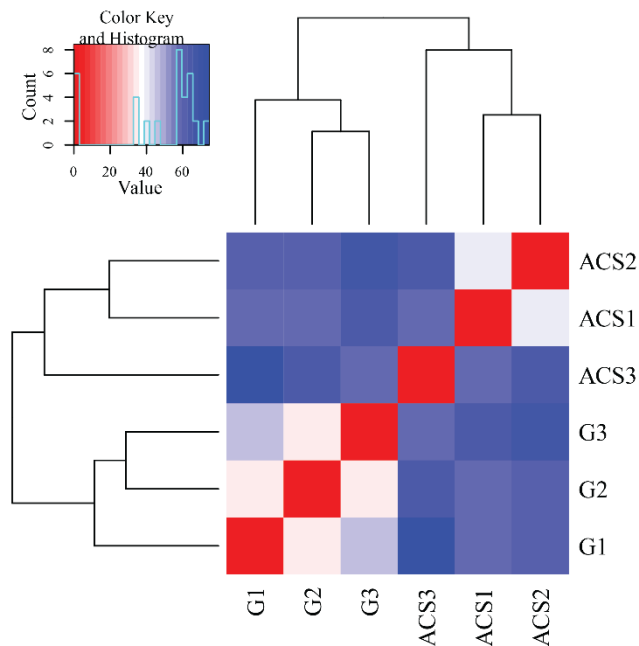


**Figure A2.2. Pairwise comparison of raw number of reads for each biological replicate in each condition.**

Red circles represent replicate 1 versus replicate 2; Blue, replicate 2 versus replicate 3; Green, replicate 1 versus replicate 3. Trend line indicates  $y=x$ , the theoretical line for perfect match between replicates.

SACTE



SDPB6

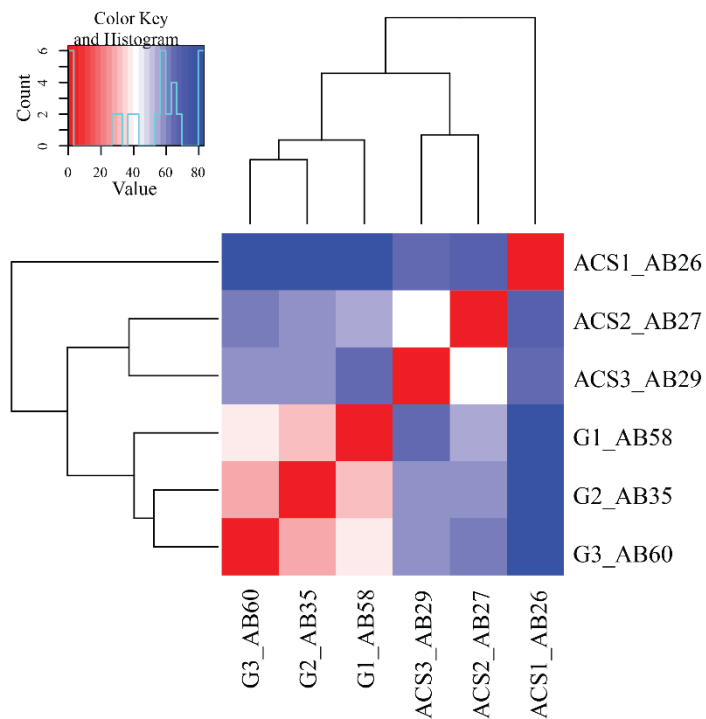


**Figure A2.3. Sample to sample variance of RNA-Seq data.** Heatmap showing the Euclidean distances between samples as calculated from the variance-stabilizing transformation of the RNA-Seq count data. Red indicates strong correlation. Cluster-tree indicates relatedness of samples.

**Table A2.1.** RNA-Seq statistics

| Strain | Growth Condition | Sample ID | Total Reads | Mapped Reads | % Mapped | Genome Coverage (reads per ORF) |
|--------|------------------|-----------|-------------|--------------|----------|---------------------------------|
| SACTE  | Glucose          | AB-70     | 15154581    | 7954329      | 52%      | 1197                            |
|        |                  | AB-32     | 16291889    | 10289877     | 63%      | 1548                            |
|        |                  | AB-18     | 8213541     | 4849260      | 59%      | 730                             |
|        | AFEX-CS          | AB-39     | 7777610     | 3273638      | 42%      | 492                             |
|        |                  | AB-57     | 21935958    | 13832930     | 63%      | 2081                            |
|        |                  | AB-44     | 11087724    | 4933583      | 44%      | 742                             |
| SDPB6  | Glucose          | AB-58     | 26068114    | 10134764     | 39%      | 1593                            |
|        |                  | AB-35     | 17600584    | 5195048      | 30%      | 817                             |
|        |                  | AB-60     | 17661709    | 5757524      | 33%      | 905                             |
|        | AFEX-CS          | AB-26     | 11205491    | 3246117      | 29%      | 510                             |
|        |                  | AB-29     | 18020392    | 4908959      | 27%      | 772                             |
|        |                  | AB-27     | 11538238    | 4930499      | 43%      | 775                             |

**Dataset A2.1.** TIGERFAM protein families used to generate the multilocus phylogenetic tree in Figure 2.3.

Available online at [http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-](http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd2.xlsx)

14.DCSupplemental/zam999105528sd2.xlsx.

**Dataset A2.2.** CAZy annotations for SACTE, SDPB6, SACTG, and SPB74. Available online at

<http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd3.xlsx>.

**Dataset A2.3.** LC-MS/MS of SACTE, S2a protein report, and S2b peptide report. Available online at

<http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd4.xlsx>.

**Dataset A2.4.** LC-MS/MS of SDPB6, S2a protein report, and S2b peptide report. Available online at

<http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd5.xlsx>.

**Dataset A2.5.** Differential expression analysis of SACTE (ActE) RNA-seq data for cultures grown on glucose or

AFEX-CS. Available online at [http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-](http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd6.xlsx)

14.DCSupplemental/zam999105528sd6.xlsx.

**Dataset A2.6.** Differential expression analysis of SDPB6 RNA-seq data for cultures grown on glucose or AFEX-CS. Available online at <http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd7.xlsx>.

**Dataset A2.7.** Differential expression of KEGG and CAZy categories for SDPB6 and SACTE based on RNA-seq differential expression (DE) analysis. Available online at <http://aem.asm.org/content/suppl/2014/07/03/AEM.01133-14.DCSupplemental/zam999105528sd8.xlsx>.

**Appendix 3: Supplementary Material for Chapter 3**

**Previously Published as Supplementary Material:** Adam J. Book\*, Gina R. Lewin\*, Bradon R. McDonald, Taichi E. Takasuka, Evelyn Wendt- Pienkowski, Drew T. Doering, Steven Suh, Kenneth F. Raffa, Brian G. Fox, and Cameron R. Currie; \*authors contributed equally (2016). Evolution of high cellulolytic activity in symbiotic *Streptomyces* through selection of expanded gene content and coordinated gene expression. *PLoS Biology*. 14(3): e1002475.

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

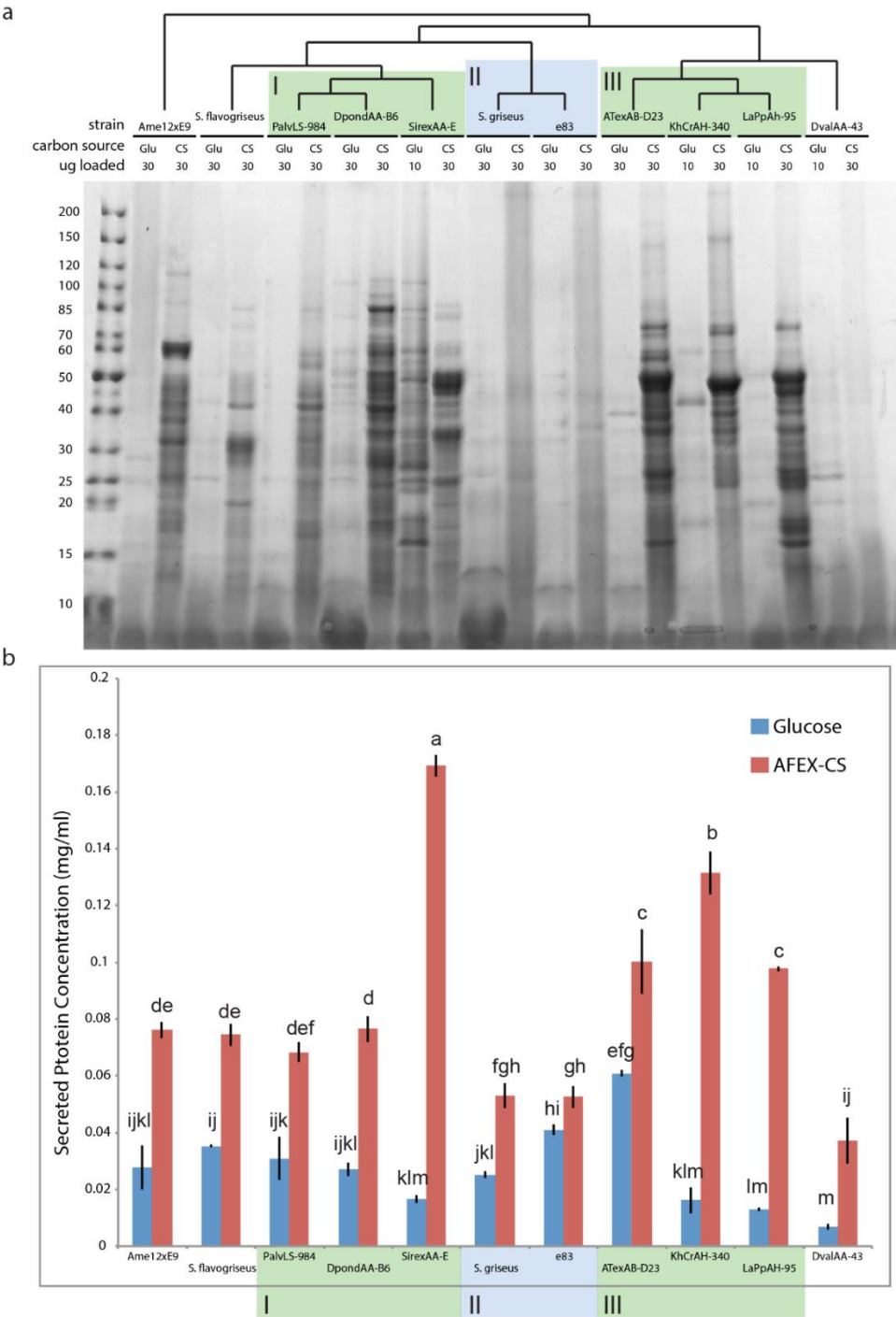


**Figure A3.1. Secreted protein profiles of *Streptomyces* strains from clades I-III. (a)** SDS-PAGE of total secreted proteins from each strain grown on either glucose (glu) or AFEX-CS (CS) as the sole carbon source. Phylogenetic relationship of strains indicated by cladogram. **(b)** Total secreted protein concentration of strains grown with glucose or AFEX-CS as the carbon source.

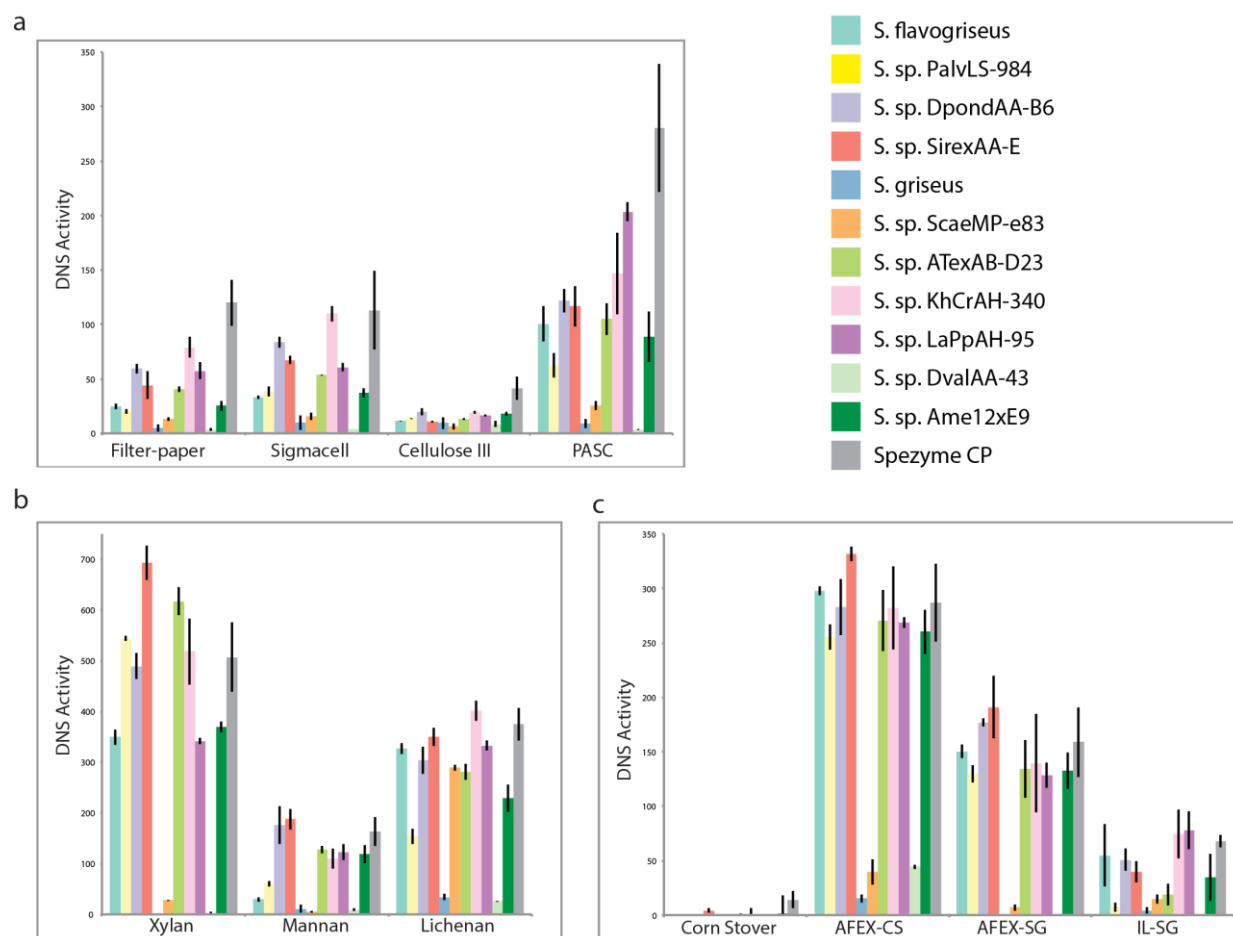


**Figure A3.2. Polysaccharide degrading enzyme activities (DNS activities) of *Streptomyces* secretomes. (a)**

Cellulose degrading activity of secretomes isolated from cellulolytic and clade II strains using filter paper, Sigmacell (Avicel), Cellulose III, and Phosphoric acid swollen cellulose (PASC). **(b)** Hemicellulose degrading activity with xylan, mannan, and lichenan substrates. **(c)** Plant biomass degrading activity of untreated and pre-treated substrates; corn stover (untreated), AFEX pre-treated corn stover (AFEX-CS) and switchgrass (AFEX-SG), and ionic liquid pretreated switchgrass (IL-SG).





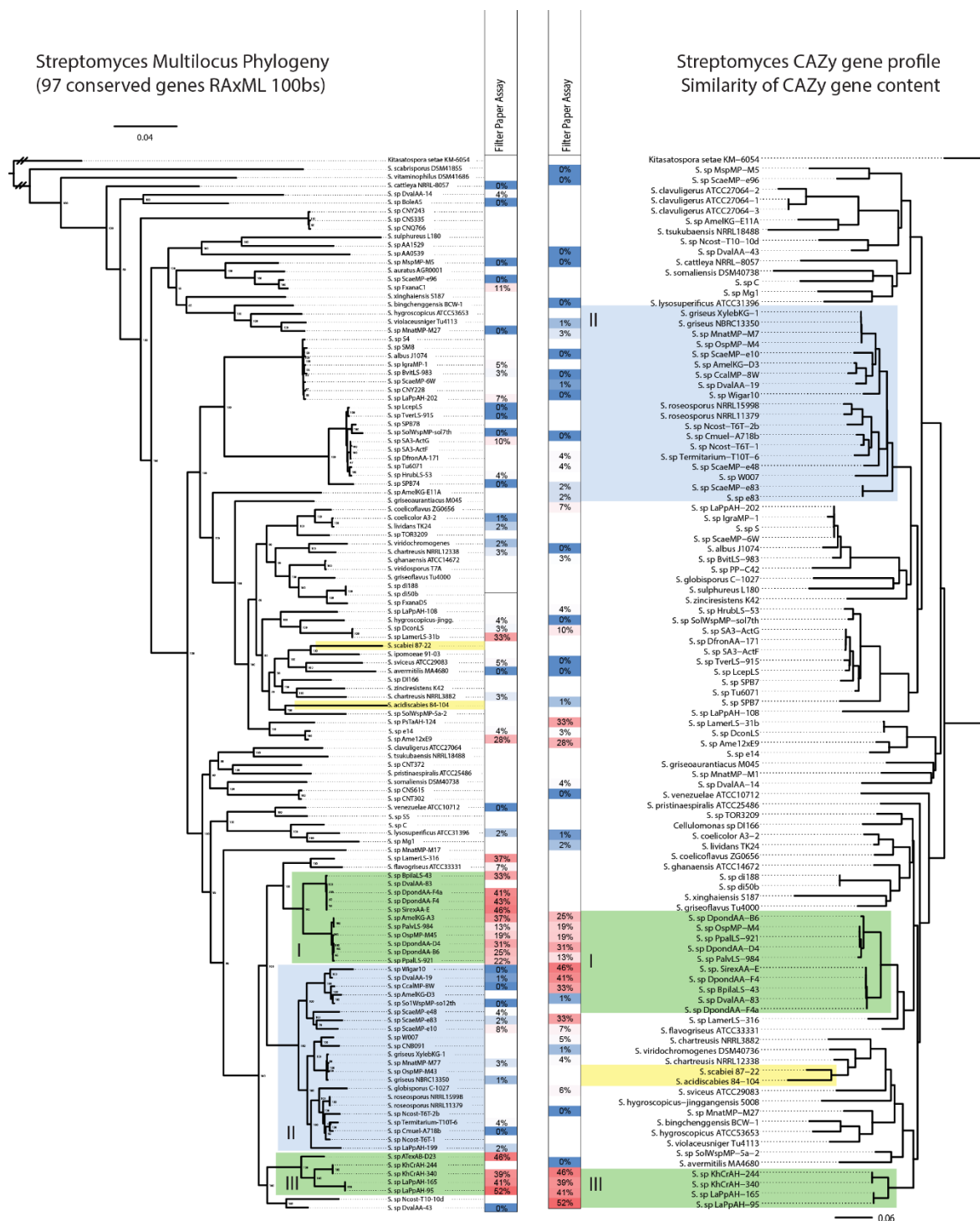


**Figure A3.4. Comparative analysis of *Streptomyces* phylogeny and the CAZy gene profile.** *Streptomyces*

multilocus phylogeny is shown on the left side and is annotated with the filter-paper degrading activity. The tree on the right represents the similarity of CAZy gene content in each genome. Total CAZy profiles were analyzed by Spearman rank correlation and tree was generated by phylip. Highly cellulolytic clades I and III are highlighted in green, low activity clade II in blue, and phytopathogens in yellow.



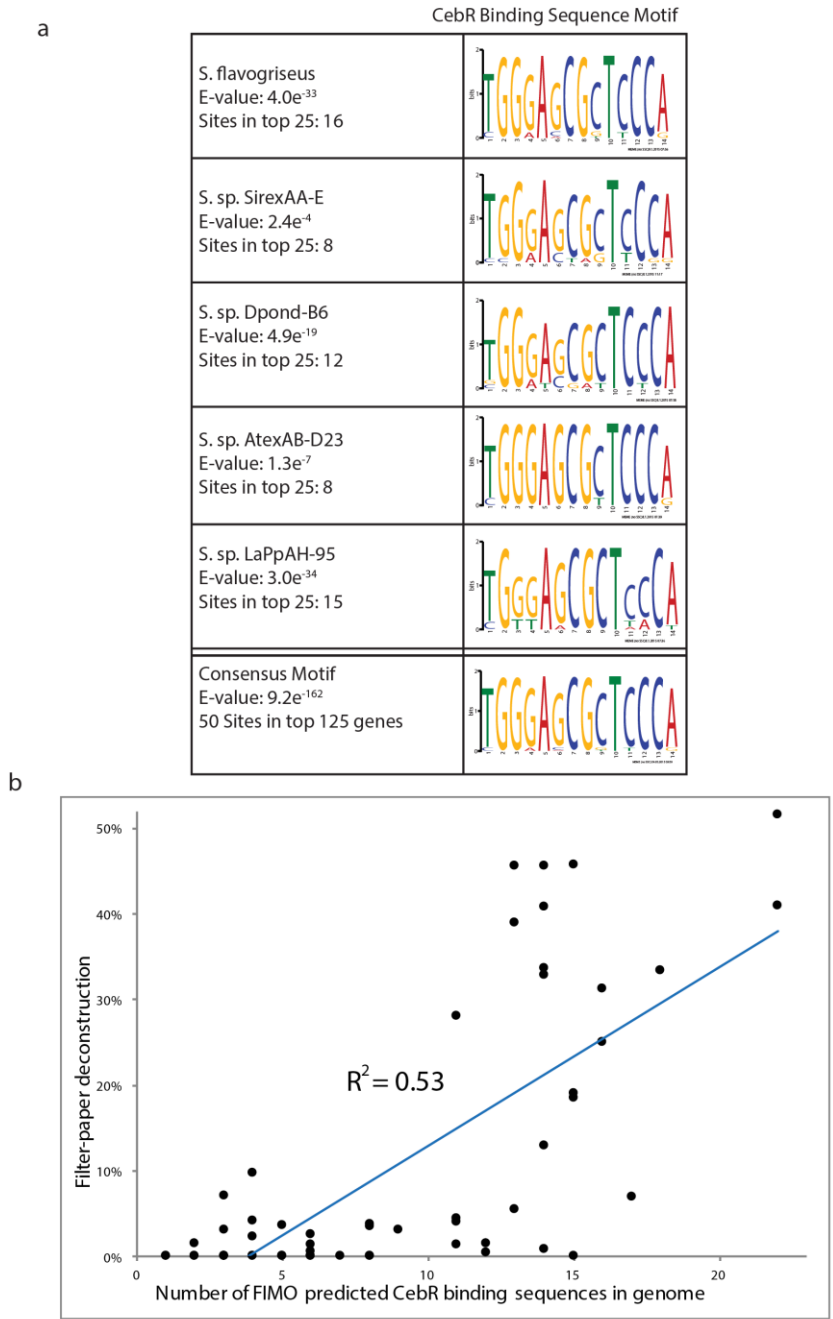


**Figure A3.6. Promoter motif analysis of cellulytic and non-cellulytic *Streptomyces*.** (a) Promoter sequences (500 bp upstream of start codon) for the top 25 most highly induced genes identified in the RNA-seq dataset of each strain were analyzed for conserved motifs. The consensus motif was generated by combining the top 25 genes from each strain and running MEME analysis (b) Correlation between the number of predicted CebR binding sites in each genome and the filter-paper degrading activity of each strain. Blue line indicates the linear regression line for the scatter plot and the  $R^2$  coefficient of determination is shown.

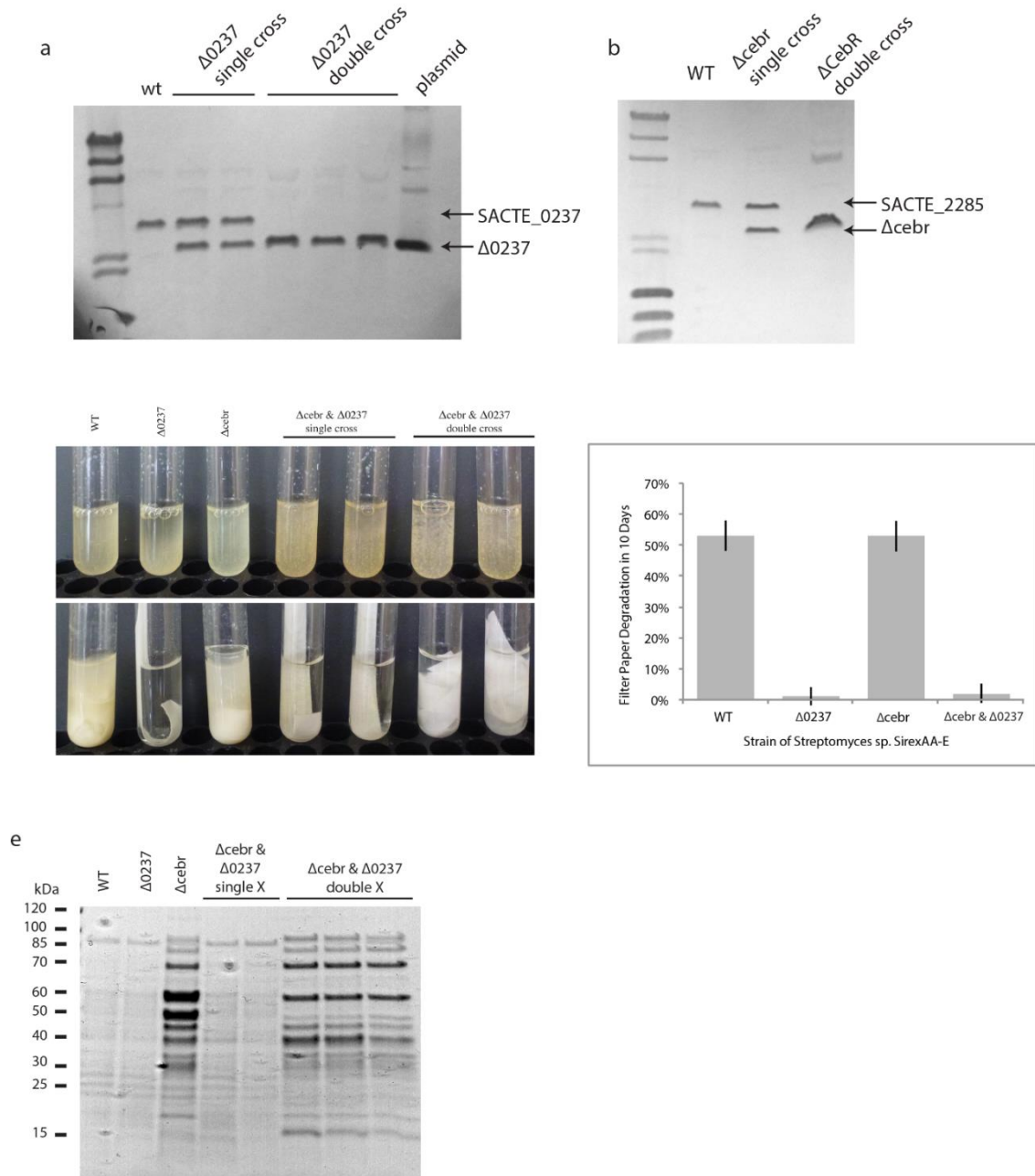


**Figure A3.7. Genetic engineering of *Streptomyces* sp. SirexAA-E.** (a) Southern blot analysis of  $\Delta 0237$  mutants showing deletion of full length SACTE\_0237 (GH6-CBM2 enzyme). (b) Southern blot analysis of  $\Delta cebr$  mutants showing deletion of full length SACTE\_2285 (CebR transcription regulator). (c) Phenotypic analysis of  $\Delta 0237$ ,  $\Delta cebr$ , and  $\Delta 0237\Delta cebr$  double mutants grown with glucose or filter-paper as the sole carbon source. (d) Quantitative filter-paper degrading activity of  $\Delta 0237$ ,  $\Delta cebr$ , and  $\Delta 0237\Delta cebr$  mutants. (e) Secreted protein profile of  $\Delta 0237$ ,  $\Delta cebr$ , and  $\Delta 0237\Delta cebr$  mutants grown with glucose as the sole carbon source.



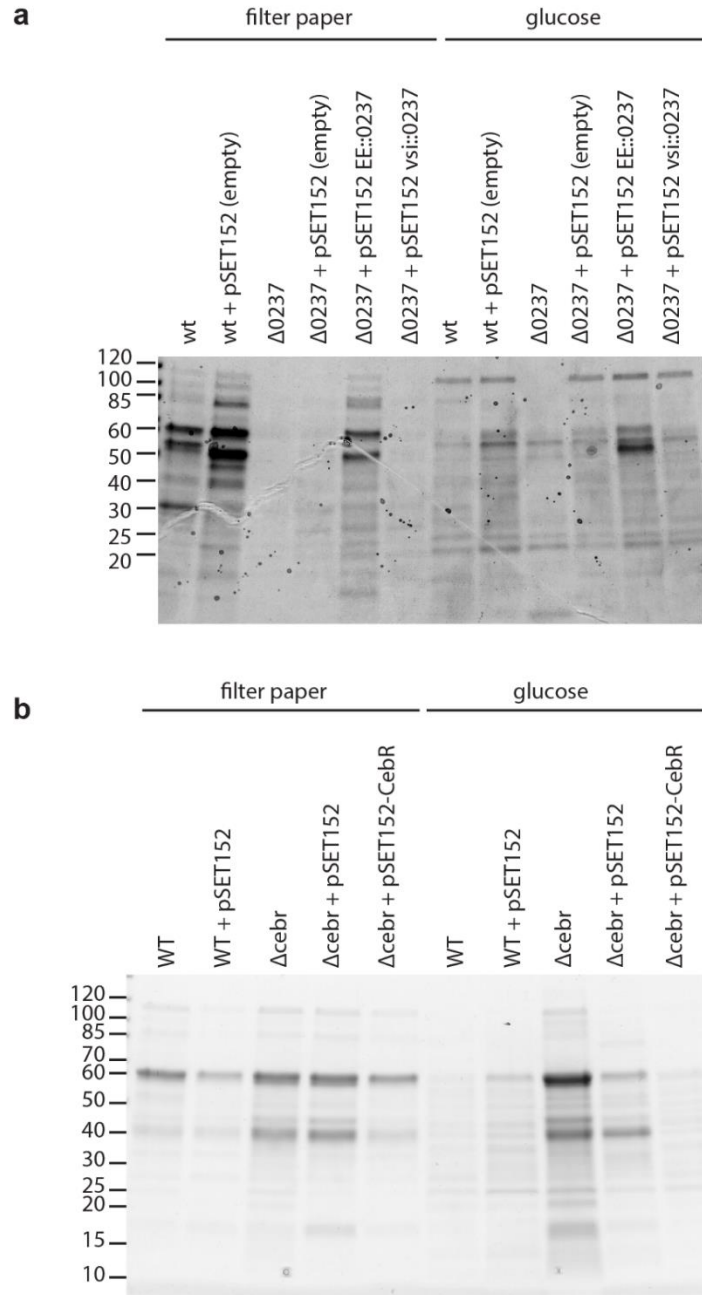


**Figure A3.8. Secreted protein profiles of rescued *Streptomyces* sp. SirexAA-E deletion strains. (a)** Rescue of WT protein profile of  $\Delta$ 0237 strain with pSET152EE::0237 transgene. This construction incorporates the WT SACTE\_0237 into the genome and is driven by the constitutive ErmE\* promoter. Rescued strain has regained ability to grow and secrete proteins in filter paper media. Glucose media shows constitutive expression of the protein (SACTE\_0237) **(b)** Rescue of WT protein profile of  $\Delta$ cebr strain with pSET152EE::CebR transgene in glucose media.

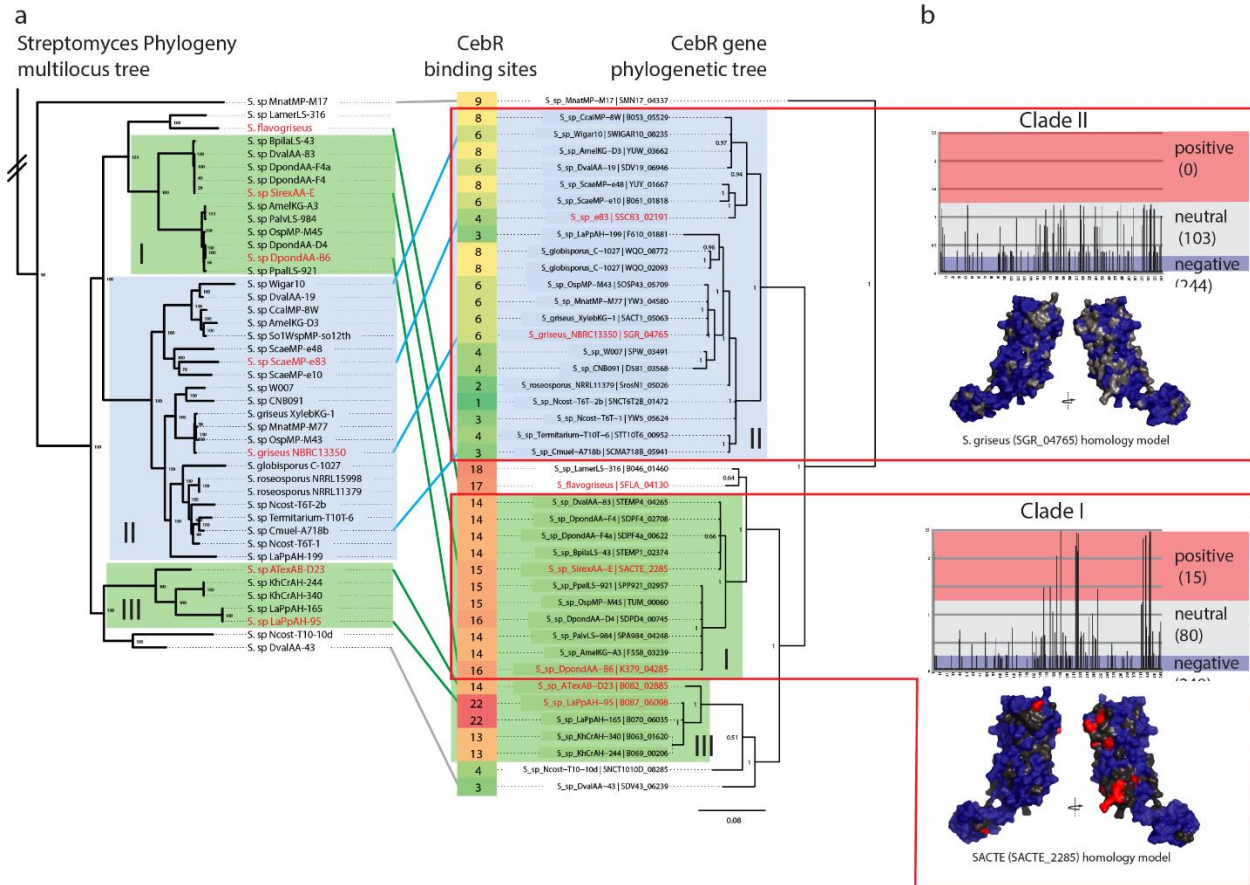


**Figure A3.9. Evolution of the CebR transcription regulator in cellulolytic and low activity strains. (a)**

Comparison of *Streptomyces* multilocus phylogenetic tree and the CebR gene tree. Tree is rooted with the outgroup *S. sp. MnaMP-M17*. Posterior probabilities for each node is indicated. CebR gene tree is annotated with the number of CebR binding sites (TGGGAGCGCTCCCA) present in each genome. **(b)** Site-specific estimation of dN/dS ratios for the cellulolytic clade I and clade II. Positively selected residues are colored in red, neutral in grey, and negative in blue. The x-axis corresponds to the protein sequence, and the y-axis corresponds to the posterior probability of the estimation. dN/dS ratios were also mapped onto the modeled structures for SACTE\_2285 and SGR\_04765. Colors correspond to selection rates described above.

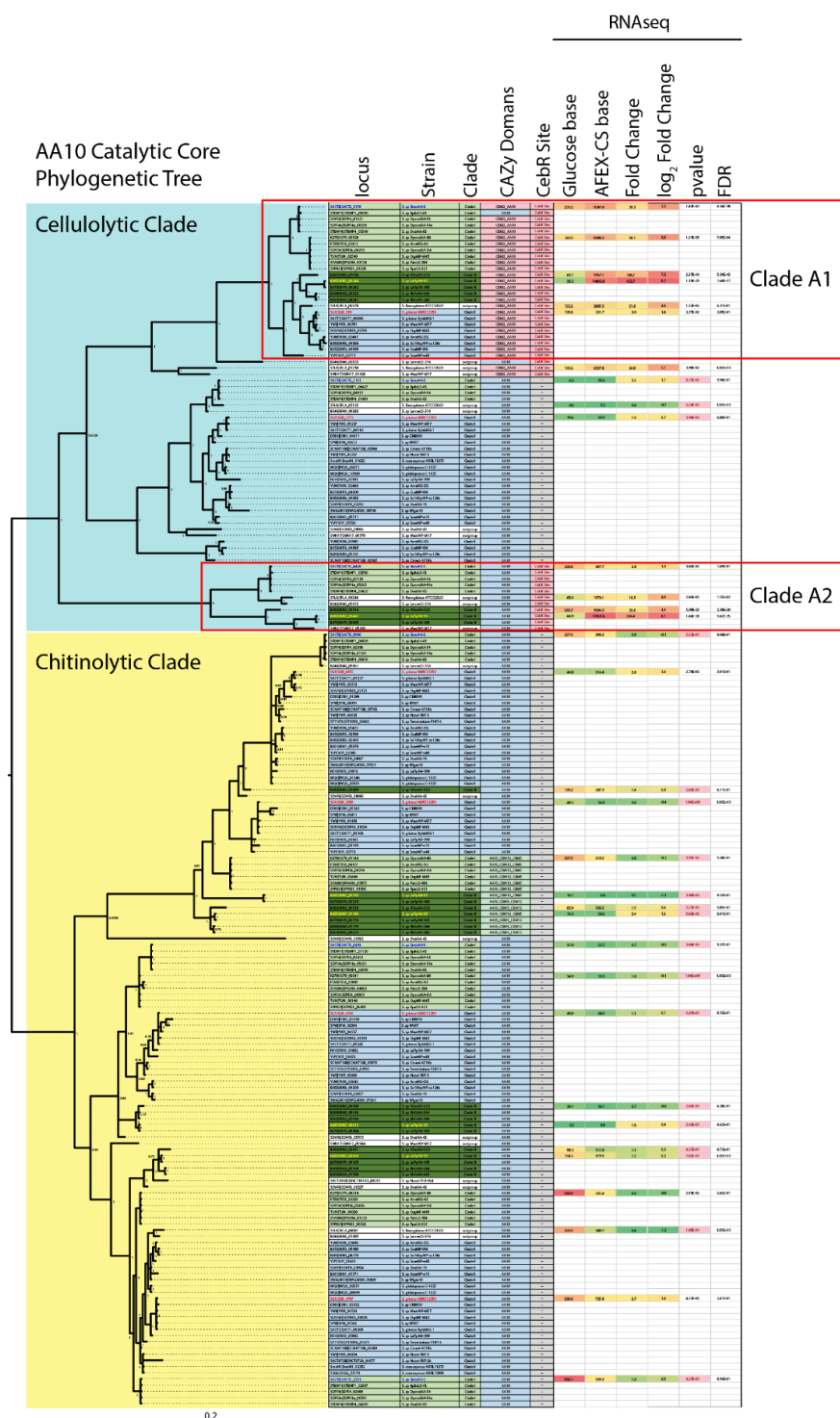
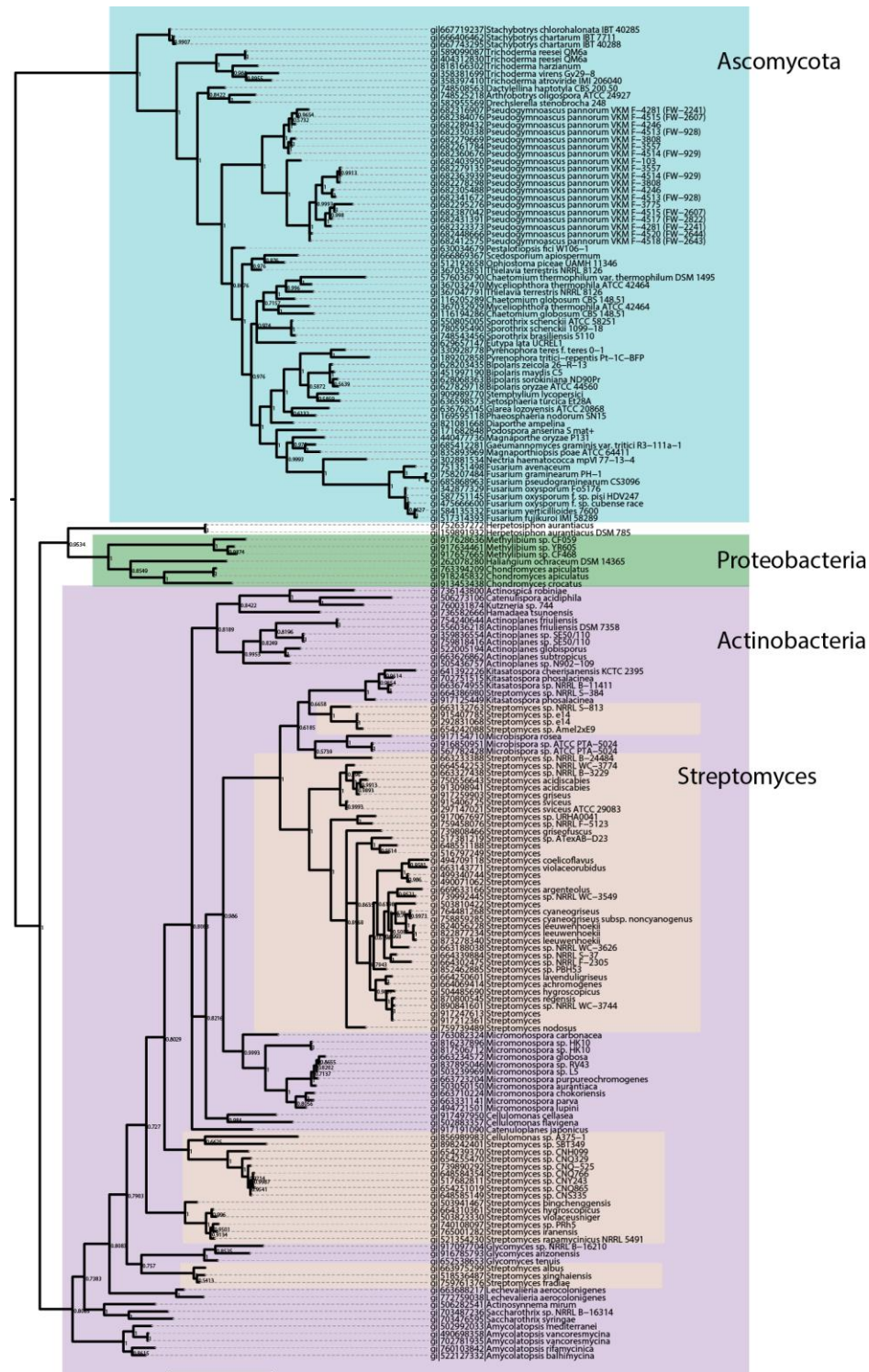
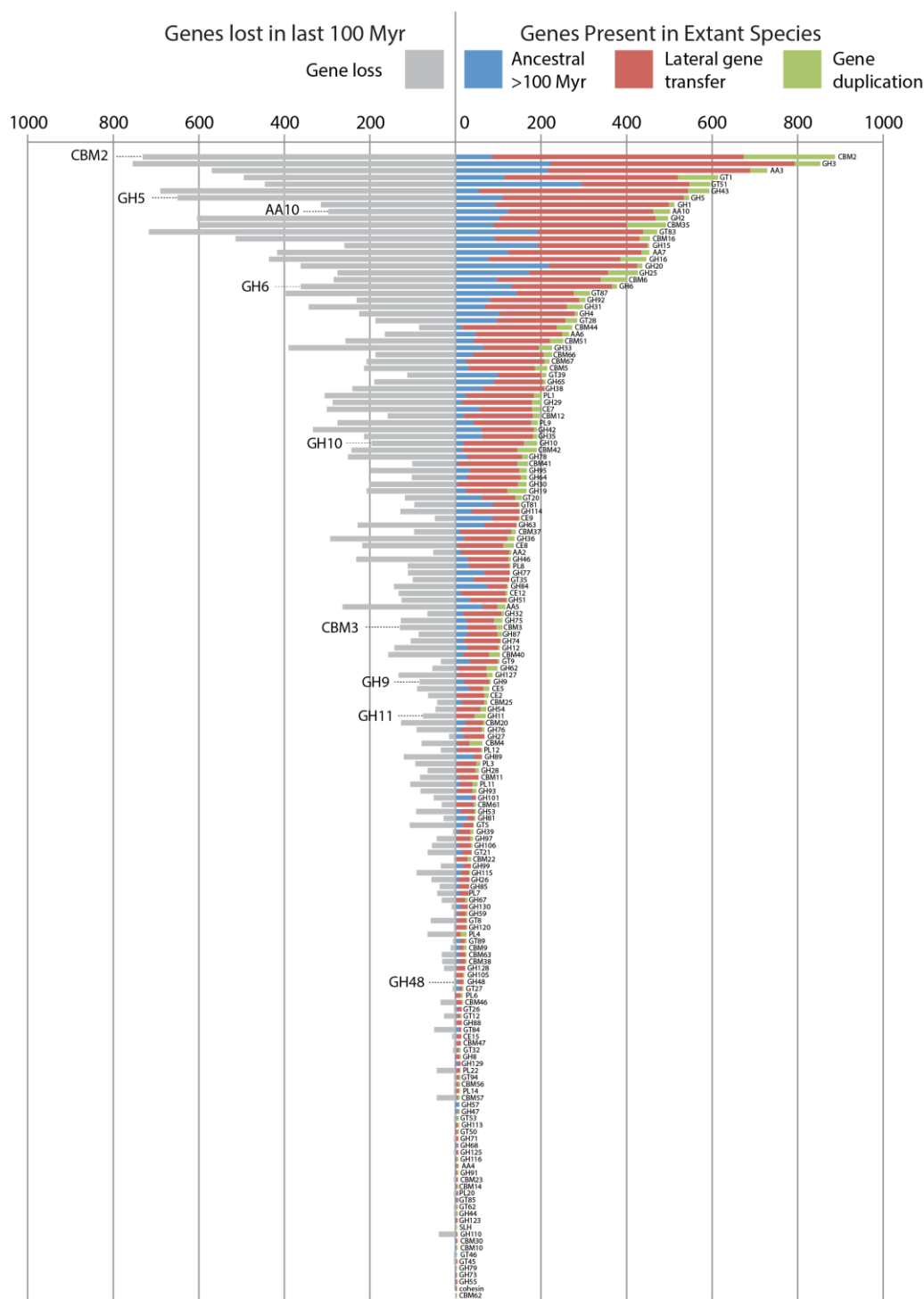


**Figure A3.10. Phylogenetic diversity of AA10 genes from strains in clades I-III.** Bayesian phylogenetic tree of AA10 protein homologs. The tree is midpoint rooted and nodes are labeled with posterior probability scores. Putative cellulolytic and chitinolytic clades are shaded. The tree is annotated with CAZy domains, CebR regulatory sequences, and RNA-seq expression data. The two most highly expressed and induced clades are indicated, A1 and A2 respectively.







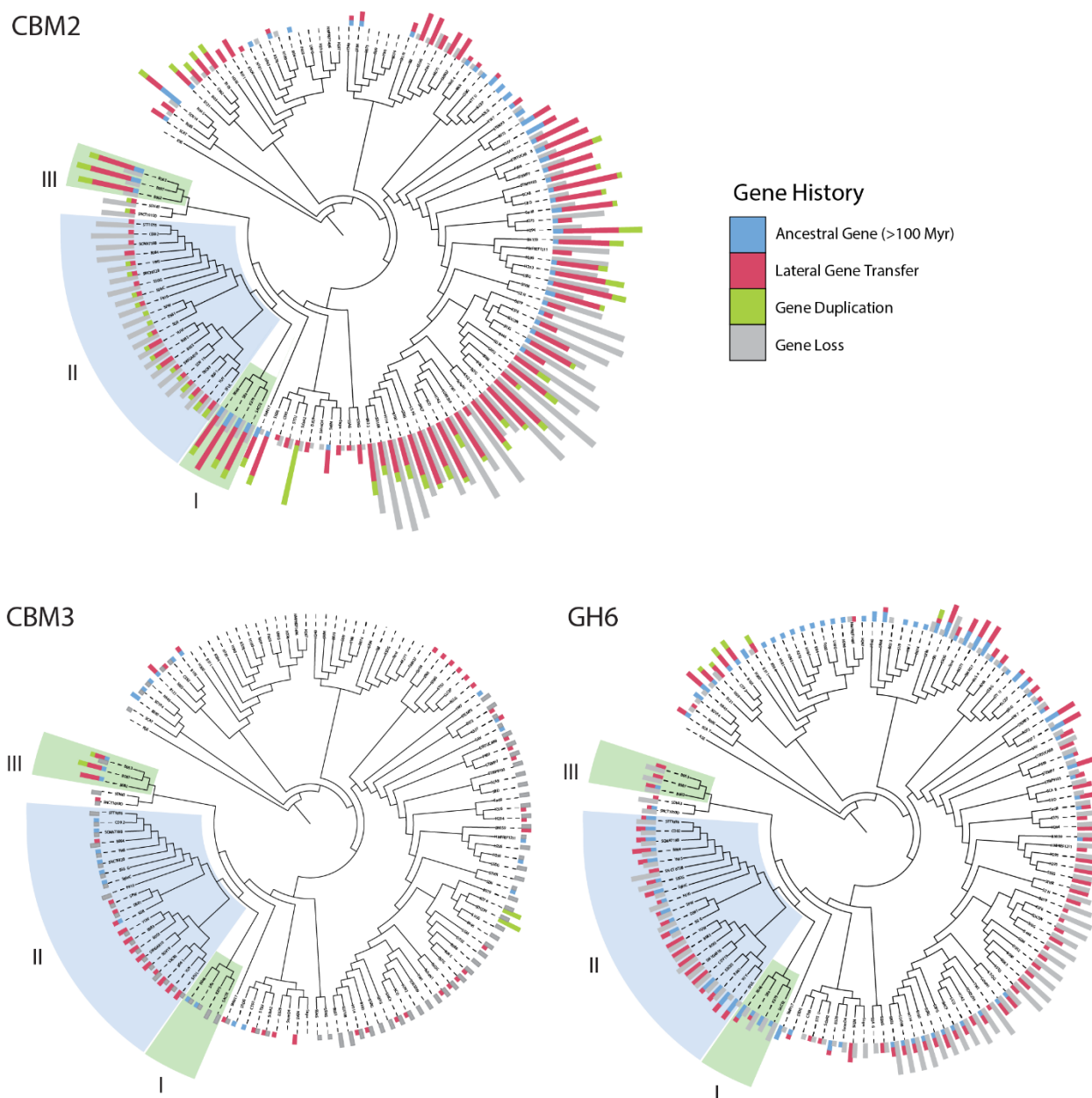


**Figure A3.13. Evolution of CBM2, CBM3 and GH6 gene families.** Gene gain and loss events for each gene family in our phylogenomic dataset. Gene loss is shown with grey bars. Gene gain is separated into three categories: ancestral genes older than 100my are shown in blue, laterally transferred genes within the last 100my are shown in red, and duplicated genes are shown in green. Cellulolytic clades (I and III) are highlighted in green and the clade II in blue.

**Table A3.1** Qualitative filter-paper degrading activity of *Streptomyces* [20,26,75-82]

| Strain                      | FP Deg Rank | Habitat         | Source     | Isolation Reference                                                                                                                                                                                                                                                      |
|-----------------------------|-------------|-----------------|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 14 Sirex 13bl               | 3           | Host-associated | Currie Lab | Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp <i>Sirex noctilio</i> . ISME J 5:1323-1331.                                                               |
| SA-ActH                     | 1           | Host-associated | Currie Lab | Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp <i>Sirex noctilio</i> . ISME J 5:1323-1331.                                                               |
| Strept SA3 ActG             | 1           | Host-associated | Currie Lab | Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp <i>Sirex noctilio</i> . ISME J 5:1323-1331.                                                               |
| Strept SirexAA E            | 5           | Host-associated | Currie Lab | Adams AS, Jordan MS, Adams SM, Suen G, Goodwin LA, Davenport KW, Currie CR, Raffa KF. 2011. Cellulose-degrading bacteria associated with the invasive woodwasp <i>Sirex noctilio</i> . ISME J 5:1323-1331.                                                               |
| 2-DV83                      | 3           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae)." <i>Environmental entomology</i> 39.2 (2010): 406-414.                                                                     |
| 21 Unknown Dv10 grey G21    | 0           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae)." <i>Environmental entomology</i> 39.2 (2010): 406-414.                                                                     |
| 8 Classic White Dv21 yellow | 1           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae)." <i>Environmental entomology</i> 39.2 (2010): 406-414.                                                                     |
| Classic White Dv26          | 1           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae)." <i>Environmental entomology</i> 39.2 (2010): 406-414.                                                                     |
| Grey Mix Dv19 consensus     | 3           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae)." <i>Environmental entomology</i> 39.2 (2010): 406-414.                                                                     |
| Grey mix Dv21 consensus GL  | 5           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Geographic variation in bacterial communities associated with the red turpentine beetle (Coleoptera: Curculionidae)." <i>Environmental entomology</i> 39.2 (2010): 406-414.                                                                     |
| MPBA-50                     | 5           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism." <i>Applied and environmental microbiology</i> 79.11 (2013): 3468-3475. |
| Strept DpondAA D4           | 5           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism." <i>Applied and environmental microbiology</i> 79.11 (2013): 3468-3475. |
| Strept DpondAA F4           | 3           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism." <i>Applied and environmental microbiology</i> 79.11 (2013): 3468-3475. |
| Strept DvalAA 19            | 1           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism." <i>Applied and environmental microbiology</i> 79.11 (2013): 3468-3475. |
| Strept DvalAA 43            | 2           | Host-associated | Currie Lab | Adams, Aaron S., et al. "Mountain pine beetles colonizing historical and naive host trees are associated with a bacterial community highly enriched in genes contributing to terpene metabolism." <i>Applied and environmental microbiology</i> 79.11 (2013): 3468-3475. |
| Strept DpondAA B6           | 4           | Host-associated | Currie Lab | Book, Adam J., et al. "Cellulolytic Streptomyces Strains Associated with Herbivorous Insects Share a Phylogenetically Linked Capacity To Degrade Lignocellulose." <i>Applied and environmental microbiology</i> 80.15 (2014): 4692-4701.                                 |
| M5                          | 1           | Host-associated | Currie Lab | Carr, Gavin, et al. "Microtermolides A and B from termite-associated Streptomyces sp. and structural revision of vinylamycin." <i>Organic letters</i> 14.11 (2012): 2822-2825.                                                                                           |
| MD8w                        | 2           | Host-associated | Currie Lab | Carr, Gavin, et al. "Microtermolides A and B from termite-associated Streptomyces sp. and structural revision of vinylamycin." <i>Organic letters</i> 14.11 (2012): 2822-2825.                                                                                           |
| Strept OspMP M45            | 3           | Host-associated | Currie Lab | Carr, Gavin, et al. "Microtermolides A and B from termite-associated Streptomyces sp. and structural revision of vinylamycin." <i>Organic letters</i> 14.11 (2012): 2822-2825.                                                                                           |
| Flower Actino BR A10        | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino BR A2         | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino BR A5         | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino BR A7         | 4           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino BR A9         | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino DL A11        | 5           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino DL B12        | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino DL B9         | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino GS A10        | 3           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino GS A11        | 1           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino GS A6         | 2           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino GS B12        | 2           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino GS B5         | 3           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino GS B6         | 3           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |
| Flower Actino LM A11        | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print.                                                                                          |



**Table A3.1** Qualitative filter-paper degrading activity of *Streptomyces* [20,26,75-82] (continued)

| Strain                | FP Deg Rank | Habitat         | Source     | Isolation Reference                                                                                                                                                             |
|-----------------------|-------------|-----------------|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Honey Bee 2nd Ext F1  | 4           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| Honey Bee 2nd Ext F10 | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| Honey Bee 2nd Ext F11 | 1           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| Honey Bee 2nd Ext F2  | 2           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| Honey Bee 2nd Ext F6  | 4           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| Honey Bee C3          | 2           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| Honey Bee C9          | 0           | Host-associated | Currie Lab | Grubbs, Kirk Justin, dissertation. Microbial Communities Associated with Ambrosia Beetles and the European Honey Bee, Apis Mellifera. Ann Arbor, MI :ProQuest LLC, 2013. Print. |
| 124.PsTaS2 4 2        | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 130.PsTaT12 PBS 5 2   | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 137.PsTaWA5 4 W 2     | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 149.PsTaWA9 1 W 2num2 | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 155.PsTaWD3 1 4       | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 165.LaPpA3 2 6        | 5           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 173.LaPpA81 3 1       | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 180.LaPpL24 1 4       | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 193.LaPpP19 1 5       | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 201.LaPpPBS4 1 3D     | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 223.LaPpWD7 2 3       | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 233.LaPpWD31 1 1      | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 239.LaPpWD87 1 3      | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 246.KhCrP40 1 1       | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 256.KhCrWA5 1 5       | 4           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 257.KhCrWA5 3 2       | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 274.KhCrWA14 3 7      | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 290.KhCrWA22 2 2      | 3           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 298.KhCrWA23 3 4      | 2           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 30.KhCrWA7 3 5        | 4           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 307.KhCrWD31 2 4      | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 314.KhCrWD34 1 5      | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 33.KhCrWA13 1 7       | 4           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 337.KhCrWA17 1 6      | 4           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 40.KhCrWA23 3 3       | 5           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 43.KhCrWD37 1 1       | 3           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 47.KhCrWA12 3 3       | 2           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 63.KhCrWA19 3 2       | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 87.PsTaWA7 2 3        | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| 92.PsTaWD3 1 5        | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| LaPp-PBS-2-5-4        | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |
| Strept KhCrAH 340     | 5           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                      |

**Table A3.1** Qualitative filter-paper degrading activity of *Streptomyces* [20,26,75-82] (continued)

| Strain                   | FP Deg Rank | Habitat         | Source     | Isolation Reference                                                                                                                                                                          |
|--------------------------|-------------|-----------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Strept LaPpAH 108        | 1           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                                   |
| Strept LaPpAH 202        | 0           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                                   |
| Strept LaPpAH 95         | 5           | Host-associated | Currie Lab | Hanshew, Alissa S., et al. "Characterization of Actinobacteria Associated with Three Ant-Plant Mutualisms." <i>Microbial ecology</i> 69.1 (2015): 192-203.                                   |
| e48                      | 1           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept Cnuel A718b       | 1           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept IgraMP 1          | 0           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept Ncost T10 10d     | 0           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept ScaeMP e83        | 3           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept ScaeMP-e96        | 0           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept SolWspMP sol7th   | 2           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept Temitarium T10T 6 | 0           | Host-associated | Currie Lab | Poulsen, Michael, et al. "Chemical analyses of wasp-associated <i>Streptomyces</i> bacteria reveal a prolific potential for natural products discovery." <i>PLoS One</i> 6.2 (2011): e16763. |
| Strept BpilalS 43        | 2           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept BvitLS 983        | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept DconLS            | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept LamerLS 316       | 4           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept LceplLS           | 1           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept PalvLS 984        | 3           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept TverLS 915        | 1           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| Strept Wigar10           | 3           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| SS2h                     | 5           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 1015 | 3           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 1047 | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 317  | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 432  | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 53   | 2           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 884  | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 895  | 4           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 896  | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 904  | 4           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 909  | 3           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 969  | 3           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| <i>Streptomyces</i> 986  | 0           | Host-associated | Currie Lab | Schwalb, Laura. "Investigation of Endosymbiotic Actinobacteria Occurring in Higher Fungi and Medicinal Prairie Plants." University of Wisconsin - Madison, 2011. Print.                      |
| di162b                   | 0           | Host-associated | Currie Lab | Scott, Jarrod Jude. "Microbial Community Ecology of Leaf-Cutter Ant Refuse Dumps." N. p., 0ADAD. Print.                                                                                      |
| Strept TexD              | 3           | Host-associated | Currie Lab | This Study                                                                                                                                                                                   |
| M77                      | 0           | Host-associated | Currie Lab | Visser, Anna A., et al. "Exploring the potential for actinobacteria as defensive symbionts in fungus-growing termites." <i>Microbial ecology</i> 63.4 (2012): 975-985.                       |

**Table A3.1** Qualitative filter-paper degrading activity of *Streptomyces* [20,26,75-82] (continued)

| Strain                                       | FP Deg Rank | Habitat         | Source     | Isolation Reference |
|----------------------------------------------|-------------|-----------------|------------|---------------------|
| Strept flavogriseus ATCC33331                | 2           | Host-associated | Currie Lab |                     |
| Kitasatospora setae U93332                   | 0           | Free-living     | NRRL       |                     |
| Strept coelicolor A3.2                       | 2           | Free-living     | NRRL       |                     |
| Strept hygrosopicus jinggangensis 5008       | 2           | Free-living     | NRRL       |                     |
| Streptomyces albolongus AB184425             | 0           | Free-living     | NRRL       |                     |
| Streptomyces albulus AB024440                | 0           | Free-living     | NRRL       |                     |
| Streptomyces albus albus AJ621602            | 0           | Free-living     | NRRL       |                     |
| Streptomyces althoticus AY999808             | 3           | Free-living     | NRRL       |                     |
| Streptomyces anulatus DQ026637               | 0           | Free-living     | NRRL       |                     |
| Streptomyces atratus DQ026638                | 0           | Free-living     | NRRL       |                     |
| Streptomyces aurantiacus AJ781383            | 0           | Free-living     | NRRL       |                     |
| Streptomyces aureovorticellatus AY999774     | 0           | Free-living     | NRRL       |                     |
| Streptomyces bambergiensis AB184869          | 1           | Free-living     | NRRL       |                     |
| Streptomyces celluloflavus AB184476          | 0           | Free-living     | NRRL       |                     |
| Streptomyces cellulosae DQ442495             | 0           | Free-living     | NRRL       |                     |
| Streptomyces chattanoogensis AY295791        | 1           | Free-living     | NRRL       |                     |
| Streptomyces chryseus AY999787               | 0           | Free-living     | NRRL       |                     |
| Streptomyces cinerochromogenes AB184507      | 1           | Free-living     | NRRL       |                     |
| Streptomyces cirratus AY999794               | 0           | Free-living     | NRRL       |                     |
| Streptomyces coeruleofuscus DQ026668         | 0           | Free-living     | NRRL       |                     |
| Streptomyces coeruleorubidus NBRC 12761      | 0           | Free-living     | NRRL       |                     |
| Streptomyces daghestanicus DQ442497          | 0           | Free-living     | NRRL       |                     |
| Streptomyces diastaticus ardesiacus DQ026631 | 2           | Free-living     | NRRL       |                     |
| Streptomyces drozdowiczii AB249957           | 3           | Free-living     | NRRL       |                     |
| Streptomyces eurocidicus AY999790            | 0           | Free-living     | NRRL       |                     |
| Streptomyces eurythermus D63870              | 3           | Free-living     | NRRL       |                     |
| Streptomyces filamentosus AB184130           | 0           | Free-living     | NRRL       |                     |
| Streptomyces flaveus DQ026643                | 0           | Free-living     | NRRL       |                     |
| Streptomyces flavotricini AY999915           | 0           | Free-living     | NRRL       |                     |
| Streptomyces flavovirens DQ026635            | 1           | Free-living     | NRRL       |                     |
| Streptomyces geysiriensis DQ442501           | 3           | Free-living     | NRRL       |                     |
| Streptomyces ghanaensis AY999851             | 0           | Free-living     | NRRL       |                     |
| Streptomyces globisporus EF178686            | 0           | Free-living     | NRRL       |                     |
| Streptomyces griseiniger AJ391818            | 0           | Free-living     | NRRL       |                     |
| Streptomyces griseoflavus AJ781322           | 0           | Free-living     | NRRL       |                     |



**Table A3.1** Qualitative filter-paper degrading activity of *Streptomyces* [20,26,75-82] (continued)

| Strain                                          | FP Deg |  | Habitat     | Source    | Isolation Reference |
|-------------------------------------------------|--------|--|-------------|-----------|---------------------|
|                                                 | Rank   |  |             |           |                     |
| <i>Streptomyces griseolus</i> AB184768          | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces halstedii</i> EF178695          | 4      |  | Free-living | NRRL      |                     |
| <i>Streptomyces hawaiiensis</i> AB184143        | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces hiroshimensis</i> AB184789      | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces lavenduligriseus</i> DQ442515   | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces lavendulocolor</i> DQ442516     | 1      |  | Free-living | NRRL      |                     |
| <i>Streptomyces libani rufus</i> AJ781351       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces longisporoflavus</i> DQ442520   | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces lucensis</i> DQ442522           | 1      |  | Free-living | NRRL      |                     |
| <i>Streptomyces luridus</i> DQ442523            | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces massasporeus</i> AB184152       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces melanosporefaciens</i> AJ271887 | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces microflavus</i> DQ445795        | 2      |  | Free-living | NRRL      |                     |
| <i>Streptomyces misakiensis</i> AB217605        | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces misionensis</i> EF178678        | 3      |  | Free-living | NRRL      |                     |
| <i>Streptomyces niger</i> AJ621607              | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces nitrosporeus</i> EF178680       | 4      |  | Free-living | NRRL      |                     |
| <i>Streptomyces omiyaensis</i> EF178697         | 2      |  | Free-living | NRRL      |                     |
| <i>Streptomyces phaeoluteigriseus</i> AJ391815  | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces phaeopurpureus</i> DQ026666     | 3      |  | Free-living | NRRL      |                     |
| <i>Streptomyces pluricologrescens</i> DQ442540  | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces poonensis</i> DQ445792          | 3      |  | Free-living | NRRL      |                     |
| <i>Streptomyces prasinus</i> DQ026658           | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces pseudovenezuelae</i> AB184233   | 2      |  | Free-living | NRRL      |                     |
| <i>Streptomyces purpeofuscus</i> AJ781364       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces rectiviolaceus</i> DQ026660     | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces regensis</i> DQ026649           | 1      |  | Free-living | NRRL      |                     |
| <i>Streptomyces rishiriensis</i> EF178682       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces sampsonii</i> ATCC 25495        | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces sioyaensis</i> DQ026654         | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces sparsogenes</i> AB184301        | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces tauricus</i> AB045879           | 2      |  | Free-living | NRRL      |                     |
| <i>Streptomyces tendae</i> D63873               | 2      |  | Free-living | NRRL      |                     |
| <i>Streptomyces thermocarboxydus</i> U94490     | 2      |  | Free-living | NRRL      |                     |
| <i>Streptomyces tubercidicus</i> AJ621612       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces variabilis</i> DQ442551         | 3      |  | Free-living | NRRL      |                     |
| <i>Streptomyces variegatus</i> AJ781371         | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces varsoviensis</i> DQ026653       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces violaceusniger</i> AB184420     | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces virens</i> DQ442554             | 4      |  | Free-living | NRRL      |                     |
| <i>Streptomyces werraensis</i> DQ442558         | 1      |  | Free-living | NRRL      |                     |
| <i>Streptomyces xanthochromogenes</i> DQ442559  | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces xinghaiensis</i> EF577247       | 0      |  | Free-living | NRRL      |                     |
| <i>Streptomyces zaomyceticus</i> EF178685       | 0      |  | Free-living | NRRL      |                     |
| WMMB 322                                        | 0      |  | Free-living | Bugni Lab |                     |
| WMMB 712                                        | 0      |  | Free-living | Bugni Lab |                     |

**Table A3.2.** Genome Accession Numbers

| Strain Name                                      | Genome Abbreviation <sup>a</sup> | Source <sup>b</sup> | RefSeq Accession | Reference  |
|--------------------------------------------------|----------------------------------|---------------------|------------------|------------|
| Streptomyces_sp_DI166                            | CDI166                           | CL                  | PRJNA319242      | This Study |
| Streptomyces_sp_AmelKG-A3                        | F558                             | CL                  | PRJNA302567      | This Study |
| Streptomyces_sp_BpilaLS-43                       | STEMP1                           | CL                  | PRJNA319235      | This Study |
| Streptomyces_sp_BvitLS-983                       | SBV983                           | CL                  | PRJNA319064      | This Study |
| Streptomyces_sp_DfronAA-171                      | SDF171                           | CL                  | PRJNA319234      | This Study |
| Streptomyces_sp_di50b                            | SDI50B                           | CL                  | PRJNA319212      | This Study |
| Streptomyces_sp_DpondAA-D4                       | SDPD4                            | CL                  | PRJNA319225      | This Study |
| Streptomyces_sp_DpondAA-F4                       | SDPF4                            | CL                  | PRJNA319226      | This Study |
| Streptomyces_sp_DpondAA-F4a                      | SDPF4a                           | CL                  | PRJNA319227      | This Study |
| Streptomyces_sp_IgraMP-1                         | SIGACT1                          | CL                  | PRJNA319223      | This Study |
| Streptomyces_sp_LamerLS-31b                      | STEMP5                           | CL                  | PRJNA319221      | This Study |
| Streptomyces_sp_MnatMP-M77                       | YW3                              | CL                  | PRJNA163021      | This Study |
| Streptomyces_sp_OspMP-M43                        | SOSP43                           | CL                  | PRJNA319239      | This Study |
| Streptomyces_sp_OspMP-M45                        | TUM                              | CL                  | PRJNA163005      | This Study |
| Streptomyces_sp_PalvLS-984                       | SPA984                           | CL                  | PRJNA319240      | This Study |
| Streptomyces_sp_PpalLS-921                       | SPP921                           | CL                  | PRJNA319228      | This Study |
| Streptomyces_sp_ScaeMP-6W                        | SSC6W                            | CL                  | PRJNA319236      | This Study |
| Streptomyces_sp_SolWspMP-5a-2                    | STEMP6                           | CL                  | PRJNA319033      | This Study |
| Streptomyces_sp_SolWspMP-sol7th                  | SSOL7TH                          | CL                  | PRJNA319233      | This Study |
| Streptomyces_sp_TverLS-915                       | STV915                           | CL                  | PRJNA319217      | This Study |
| Streptomyces_sp_DvalAA-83                        | STEMP4                           | CL                  | PRJNA201129      | This Study |
| Streptomyces_sp_FxanaD5                          | B066                             | CL                  | PRJNA199244      | This Study |
| Streptomyces_sp_HrubLS-53                        | SHR53                            | CL                  | PRJNA64421       | This Study |
| Streptomyces_sp_KhCrAH-244                       | B069                             | CL                  | PRJNA199246      | This Study |
| Streptomyces_sp_LaPpAH-165                       | B070                             | CL                  | PRJNA202776      | This Study |
| Streptomyces_sp_LaPpAH-202                       | B048                             | CL                  | PRJNA199232      | This Study |
| Streptomyces_sp_SA3-ActF                         | STREPF                           | CL                  | PRJNA61901       | This Study |
| Streptomyces_sp_AmelKG-D3                        | YUW                              | CL                  | PRJNA163015      | This Study |
| Streptomyces_sp_AmelKG-E11A                      | YW7                              | CL                  | PRJNA163025      | This Study |
| Streptomyces_sp_Cmuel-A718b                      | SCMA718B                         | CL                  | PRJNA319231      | This Study |
| Streptomyces_sp_DconLS                           | SDCON                            | CL                  | PRJNA319222      | This Study |
| Streptomyces_sp_di188                            | SDI188                           | CL                  | PRJNA319209      | This Study |
| Streptomyces_sp_DvalAA-14                        | SDV14                            | CL                  | PRJNA319210      | This Study |
| Streptomyces_sp_DvalAA-19                        | SDV19                            | CL                  | PRJNA319214      | This Study |
| Streptomyces_sp_DvalAA-43                        | SDV43                            | CL                  | PRJNA319215      | This Study |
| Streptomyces_sp_e83 (Streptomyces_sp_ScaeMP-e83) | SSC83                            | CL                  | PRJNA319216      | This Study |
| Streptomyces_sp_LamerLS-316                      | B046                             | CL                  | PRJNA169718      | This Study |
| Streptomyces_sp_LaPpAH-199                       | F610                             | CL                  | PRJNA302568      | This Study |
| Streptomyces_sp_LcepLS                           | SLCEP                            | CL                  | PRJNA319218      | This Study |
| Streptomyces_sp_MnatMP-M17                       | SMN17                            | CL                  | PRJNA319237      | This Study |
| Streptomyces_sp_MnatMP-M27                       | SMN27                            | CL                  | PRJNA319224      | This Study |
| Streptomyces_sp_Ncost-T10-10d                    | SNCT1010D                        | CL                  | PRJNA319230      | This Study |
| Streptomyces_sp_Ncost-T6T-1                      | YW5                              | CL                  | PRJNA163023      | This Study |
| Streptomyces_sp_Ncost-T6T-2b                     | SNCT6T2B                         | CL                  | PRJNA319229      | This Study |
| Streptomyces_sp_ScaeMP-e48                       | YUY                              | CL                  | PRJNA163017      | This Study |
| Streptomyces_sp_ScaeMP-e96                       | YWI                              | CL                  | PRJNA163037      | This Study |
| Streptomyces_sp_Termitarium-T10T-6               | STT10T6                          | CL                  | PRJNA319232      | This Study |
| Kitasatospora_setae_KM-6054                      | KSE                              | NCBI                | PRJNA77027       |            |
| Strept_chartreusis_NRRL3882                      | KIK                              | NCBI                | PRJNA200829      |            |
| Strept_vitaminophilus_DSM41686                   | A3IG                             | NCBI                | PRJNA199207      |            |

<sup>a</sup>Genome ID used in this study<sup>b</sup>Source of genome: Currie Lab (CL), Public Genome (NCBI)

**Table A3.2.** Genome Accession Numbers (continued)

| Strain Name                                                 | Genome Abbreviation <sup>a</sup> | Source <sup>b</sup> | RefSeq Accession | Reference |
|-------------------------------------------------------------|----------------------------------|---------------------|------------------|-----------|
| <i>Streptomyces_acidiscabies</i> _84-104                    | B243                             | NCBI                | PRJNA170242      |           |
| <i>Streptomyces_albus</i> _J1074                            | XNR                              | NCBI                | PRJNA196849      |           |
| <i>Streptomyces_auratus</i> _AGR0001                        | SU9                              | NCBI                | PRJNA171646      |           |
| <i>Streptomyces_avermitilis</i> _MA4680                     | SAV                              | NCBI                | PRJNA57739       |           |
| <i>Streptomyces_bingchenggensis</i> _BCW-1                  | SBI                              | NCBI                | PRJNA82931       |           |
| <i>Streptomyces_cattlea</i> _NRRL-8057                      | YKY                              | NCBI                | PRJNA162187      |           |
| <i>Streptomyces_chartreusis</i> _NRRL12338                  | KIM                              | NCBI                | PRJNA156769      |           |
| <i>Streptomyces_clavuligerus</i> _ATCC27064                 | SSCG                             | NCBI                | PRJNA64411       |           |
| <i>Streptomyces_coelicoflavus</i> _ZG0656                   | SMCF                             | NCBI                | PRJNA180030      |           |
| <i>Streptomyces_coelicolor</i> _A3-2                        | SCO                              | NCBI                | PRJNA57801       |           |
| <i>Streptomyces_flavogriseus</i> _ATCC33331                 | SFLA                             | NCBI                | PRJNA40839       |           |
| <i>Streptomyces_ghanaensis</i> _ATCC14672                   | SSFG                             | NCBI                | PRJNA55543       |           |
| <i>Streptomyces_globisporus</i> _C-1027                     | WQO                              | NCBI                | PRJNA178734      |           |
| <i>Streptomyces_griseoaurantiacus</i> _M045                 | SGM                              | NCBI                | PRJNA66149       |           |
| <i>Streptomyces_griseoflavus</i> _Tu4000                    | SSRG                             | NCBI                | PRJNA55831       |           |
| <i>Streptomyces_griseus</i> _NBRC13350                      | SCR                              | NCBI                | PRJNA58983       |           |
| <i>Streptomyces_hygroscopicus</i> _ATCC53653                | SSOG                             | NCBI                | PRJNA33605       |           |
| <i>Streptomyces_hygroscopicus</i> -subsp-jinggangensis_5008 | SHJG                             | NCBI                | PRJNA89409       |           |
| <i>Streptomyces_ipomoeae</i> _91-03                         | STRIP9103                        | NCBI                | PRJNA183480      |           |
| <i>Streptomyces_lividans</i> _TK24                          | SSPG                             | NCBI                | PRJNA55825       |           |
| <i>Streptomyces_lyso superficialis</i> _ATCC31396           | KII                              | NCBI                | PRJNA200828      |           |
| <i>Streptomyces_pristinae spiralis</i> _ATCC25486           | SSDG                             | NCBI                | PRJNA59511       |           |
| <i>Streptomyces_roseosporus</i> _NRRL11379                  | SrosN1                           | NCBI                | PRJNA55565       |           |
| <i>Streptomyces_roseosporus</i> _NRRL15998                  | SSGG                             | NCBI                | PRJNA55545       |           |
| <i>Streptomyces_scabiei</i> _87-22                          | SCAB                             | NCBI                | PRJNA46531       |           |
| <i>Streptomyces_scabrisporus</i> _DSM41855                  | A3IC                             | NCBI                | PRJNA199206      |           |
| <i>Streptomyces_somaliensis</i> _DSM40738                   | O3Q                              | NCBI                | PRJNA176939      |           |
| <i>Streptomyces_sp</i> _AA0539                              | B311                             | NCBI                | PRJNA199548      |           |
| <i>Streptomyces_sp</i> _AA1529                              | A949                             | NCBI                | PRJNA199497      |           |
| <i>Streptomyces_sp</i> _C                                   | SSNG                             | NCBI                | PRJNA55823       |           |
| <i>Streptomyces_sp</i> _CNB091                              | D581                             | NCBI                | PRJNA199379      |           |
| <i>Streptomyces_sp</i> _CNQ766                              | C591                             | NCBI                | PRJNA199337      |           |
| <i>Streptomyces_sp</i> _CNS335                              | C592                             | NCBI                | PRJNA199338      |           |
| <i>Streptomyces_sp</i> _CNS615                              | D583                             | NCBI                | PRJNA199988      |           |
| <i>Streptomyces_sp</i> _CNT302                              | D329                             | NCBI                | PRJNA199358      |           |
| <i>Streptomyces_sp</i> _CNT372                              | C593                             | NCBI                | PRJNA199339      |           |
| <i>Streptomyces_sp</i> _CNY228                              | D330                             | NCBI                | PRJNA199359      |           |
| <i>Streptomyces_sp</i> _CNY243                              | C594                             | NCBI                | PRJNA199340      |           |
| <i>Streptomyces_sp</i> _Mg1                                 | SSAG                             | NCBI                | PRJNA207881      |           |
| <i>Streptomyces_sp</i> _S4                                  | StrS4                            | NCBI                | PRJNA78151       |           |
| <i>Streptomyces_sp</i> _SM8                                 | SM8                              | NCBI                | PRJNA180938      |           |
| <i>Streptomyces_sp</i> _SS                                  | Y19                              | NCBI                | PRJNA199004      |           |
| <i>Streptomyces_sp</i> _TOR3209                             | TOR3209                          | NCBI                | PRJNA198964      |           |
| <i>Streptomyces_sp</i> _Tu6071                              | STTU                             | NCBI                | PRJNA66919       |           |
| <i>Streptomyces_sp</i> _W007                                | SPW                              | NCBI                | PRJNA80699       |           |
| <i>Streptomyces_sulphureus</i> _L180                        | UWG                              | NCBI                | PRJNA200371      |           |
| <i>Streptomyces_sviceus</i> _ATCC29083                      | SSEG                             | NCBI                | PRJNA59513       |           |
| <i>Streptomyces_tsukubaensis</i> _NRRL18488                 | STSU                             | NCBI                | PRJNA162933      |           |
| <i>Streptomyces_venezuelae</i> _ATCC10712                   | SVEN                             | NCBI                | PRJNA177080      |           |
| <i>Streptomyces_violaceus niger</i> _Tu4113                 | Strvi                            | NCBI                | PRJNA52609       |           |

<sup>a</sup>Genome ID used in this study<sup>b</sup>Source of genome: Currie Lab (CL), Public Genome (NCBI)

**Table A3.2.** Genome Accession Numbers (continued)

| Strain Name                                     | Genome Abbreviation <sup>a</sup> | Source <sup>b</sup> | RefSeq Accession | Reference |
|-------------------------------------------------|----------------------------------|---------------------|------------------|-----------|
| <i>Streptomyces_viridochromogenes</i> _DSM40736 | SSQG                             | NCBI                | PRJNA55829       |           |
| <i>Streptomyces_viridosporus</i> _T7A           | STRVIR                           | NCBI                | PRJNA200053      |           |
| <i>Streptomyces_xinghaiensis</i> _S187          | SXIN                             | NCBI                | PRJNA67543       |           |
| <i>Streptomyces_zinciresistens</i> _K42         | SZN                              | NCBI                | PRJNA72955       |           |
| <i>Streptomyces_griseus</i> _XylebKG-1          | SACT1                            | CL                  | PRJNA42965       |           |
| <i>Streptomyces_sp_Ame12xE9</i>                 | B065                             | CL                  | PRJNA201126      |           |
| <i>Streptomyces_sp_ATexAB-D23</i>               | B082                             | CL                  | PRJNA199252      |           |
| <i>Streptomyces_sp_BoleA5</i>                   | B085                             | CL                  | PRJNA169757      |           |
| <i>Streptomyces_sp_CcalMP-8W</i>                | B053                             | CL                  | PRJNA199236      |           |
| <i>Streptomyces_sp_DpondAA-B6</i>               | K379                             | CL                  | PRJNA195846      |           |
| <i>Streptomyces_sp_e14</i>                      | SSTG                             | CL                  | PRJNA47353       |           |
| <i>Streptomyces_sp_FxanaC1</i>                  | B074                             | CL                  | PRJNA199074      |           |
| <i>Streptomyces_sp_KhCrAH-340</i>               | B063                             | CL                  | PRJNA199243      |           |
| <i>Streptomyces_sp_LaPpAH-108</i>               | B079                             | CL                  | PRJNA199251      |           |
| <i>Streptomyces_sp_LaPpAH-95</i>                | B087                             | CL                  | PRJNA199076      |           |
| <i>Streptomyces_sp_MspMP-M5</i>                 | B073                             | CL                  | PRJNA199249      |           |
| <i>Streptomyces_sp_PsTaAH-124</i>               | B086                             | CL                  | PRJNA199254      |           |
| <i>Streptomyces_sp_SA3-ActG</i>                 | STREPG                           | CL                  | PRJNA61899       |           |
| <i>Streptomyces_sp_ScaeMP-e10</i>               | B061                             | CL                  | PRJNA199241      |           |
| <i>Streptomyces_sp_SirexAA-E</i>                | SACTE                            | CL                  | PRJNA72627       |           |
| <i>Streptomyces_sp_So1WspMP-so12th</i>          | B083                             | CL                  | PRJNA169755      |           |
| <i>Streptomyces_sp_SPB74</i>                    | SSBG                             | CL                  | PRJNA48415       |           |
| <i>Streptomyces_sp_SPB78</i>                    | SSLG                             | CL                  | PRJNA55819       |           |
| <i>Streptomyces_sp_Wigar10</i>                  | SWIGAR10                         | CL                  | PRJNA200830      |           |

<sup>a</sup>Genome ID used in this study<sup>b</sup>Source of genome: Currie Lab (CL), Public Genome (NCBI)

**Table A3.3.** Summary of genes upregulated  $\geq 2$ -fold

|                        | Total | CAZy<br>genes | GH's | Cellulases <sup>a</sup> | Xylanases <sup>b</sup> | ABC<br>Transporters | Transcriptional<br>Regulators |
|------------------------|-------|---------------|------|-------------------------|------------------------|---------------------|-------------------------------|
| <i>S. flavogriseus</i> | 54    | 21            | 17   | 11                      | 3                      | 7                   | 4                             |
| SirexAA-E              | 90    | 17            | 12   | 6                       | 2                      | 6                   | 6                             |
| DpondAA-B6             | 91    | 11            | 9    | 7                       | 2                      | 10                  | 0                             |
| e83                    | 146   | 15            | 12   | 1                       | 0                      | 8                   | 4                             |
| <i>S. griseus</i>      | 122   | 4             | 4    | 1                       | 0                      | 4                   | 8                             |
| ATexAB-D23             | 205   | 19            | 15   | 10                      | 2                      | 6                   | 7                             |
| LaPpAH-95              | 186   | 39            | 32   | 15                      | 4                      | 23                  | 4                             |

<sup>a</sup>gene containing a GH5, GH6, GH8, GH9, GH12, GH48, GH74, or AA10

<sup>b</sup>gene containing a GH10, GH11, or GH30

**Table A3.4.** Expression and annotation of genes predicted to be directly controlled by the CebR repressor

|            | Binding Site(s)                 | LocusID    | Position <sup>a</sup> | log <sub>2</sub> fold<br>change <sup>b</sup> | pvalue   | Rank <sup>b</sup> | Annotation                                                          | CAZY           |
|------------|---------------------------------|------------|-----------------------|----------------------------------------------|----------|-------------------|---------------------------------------------------------------------|----------------|
| SirexAA-E  | TGGGAGCGCTCCCA                  | SACTE_0236 | -167                  | 4.64                                         | < 0.0001 | 15                | glycoside hydrolase family protein                                  | GH48_GH74_CBM2 |
|            |                                 | SACTE_0237 | -876                  | 7.87                                         | < 0.0001 | 1                 | 1, 4-beta cellobiohydrolase                                         | GH6_CBM2       |
|            | TGGGAGCGCTCCCA                  | SACTE_0352 | -33                   | 0.77                                         | NS       | 809               | GCN5-like N-acetyltransferase                                       |                |
|            |                                 | SACTE_0353 | -761                  | 0.49                                         | NS       | 1362              | hypothetical protein                                                |                |
|            | TGGGAGCGCTCCCA                  | SACTE_0482 | -51                   | 5.14                                         | < 0.0001 | 10                | glycoside hydrolase family protein                                  | CBM2_GH5_CBM35 |
|            | GGGAGCGCTCCCA*                  | SACTE_0483 | -119                  | 3.09                                         | < 0.0001 | 41                | cellulose-binding family protein                                    | CBM2           |
|            | TGGGAGCGCTCCCA                  | SACTE_0562 | -15                   | 1.45                                         | 0.0002   | 236               | cellulose-binding family protein                                    | GH74_CBM2      |
|            | TGGGAGCGCTCCCA                  | SACTE_2285 | -15                   | 0.33                                         | NS       | 1842              | LacI family transcriptional regulator                               |                |
|            | GGGAGCGCTCCCA*                  | SACTE_2286 | -15                   | 2.25                                         | < 0.0001 | 89                | beta-galactosidase                                                  | GH1            |
|            | TGGGAGCGCTCCCA                  | SACTE_2289 | -148                  | 6.01                                         | < 0.0001 | 3                 | cebE; cellobiose transport system substrate-binding protein         |                |
|            | TGGGAGCGCTCCCA                  | SACTE_2347 | -65                   | 4.98                                         | < 0.0001 | 14                | cellulose-binding family protein                                    | CBM2_GH5_CBM37 |
|            |                                 | SACTE_2345 | -1168                 | -0.64                                        | 0.0471   | 5583              | hydrophobic                                                         |                |
|            | TGGGAGCGCTCCCA                  | SACTE_3160 | -667                  | -1.82                                        | 0.0003   | 6291              | cold-shock DNA-binding domain-containing protein                    |                |
|            |                                 | SACTE_3159 | -161                  | 5.30                                         | < 0.0001 | 7                 | chitin-binding domain 3 protein                                     | CBM2_AA10      |
|            | TGGGAGCGCTCCCA                  | SACTE_3717 | -13                   | 3.96                                         | < 0.0001 | 24                | glycoside hydrolase family protein                                  | GH9_CBM4_CBM2  |
|            | TGGGAACGCTCCCA*                 | SACTE_5229 | -50                   | 1.68                                         | < 0.0001 | 170               | ABC.MS.S; multiple sugar transport system substrate-binding protein |                |
| DpondAA-B6 | GGGAGCGCTCCCA*                  | SACTE_6428 | -53                   | 1.55                                         | 0.0096   | 202               | chitin-binding domain 3 protein                                     | AA10           |
|            | TGGGAGCGCTCCCA                  | SACTE_6490 | 55                    | 0.88                                         | NS       | 672               | aldo/keto reductase                                                 |                |
|            | TGGGAGCGCTCCCA                  | K379_00590 | -200                  | 3.92                                         | 0.0420   | 25                | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | CE15_CBM2      |
|            | TGGGAGCGCTCCCA ; TGGGAGCGCTCCCA | K379_01939 | -68; -112             | 4.36                                         | 0.0009   | 16                | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH9_CBM4_CBM2  |
|            | TGGGAGCGCTCCCA                  | K379_02726 | -72                   | 3.02                                         | NS       | 51                | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH74_CBM2      |
|            | TGGGAGCGCTCCCA                  | K379_03111 | -262                  | 1.93                                         | < 0.0001 | 135               | Endopolygalacturonase                                               | GH87_CBM32     |
|            |                                 | K379_03112 | -54                   | 5.60                                         | < 0.0001 | 6                 | Mannanase                                                           | GH5_CBM37      |
|            | TGGGAGCGCTCCCA                  | K379_03258 | -66                   | 4.65                                         | < 0.0001 | 14                | Uncharacterized protein contain chitin-binding domain type 3        | CBM2_GH5_CBM37 |
|            | TGGGAGCGCTCCCA                  | K379_03538 | -141                  | 5.86                                         | < 0.0001 | 5                 | Uncharacterized protein conserved in bacteria                       | CBM2_AA10      |
|            | GGGAGCGCTCCCA*                  | K379_03942 | -110                  | 1.37                                         | 0.0006   | 267               | ABC-type dipeptide transport system, periplasmic component          |                |
|            |                                 | K379_03943 | -134                  | -0.25                                        | NS       | 4007              | transcriptional regulator, LacI family                              |                |
|            | TGGGAGCGCTCCCA                  | K379_04285 | -127                  | -0.37                                        | NS       | 4437              | transcriptional regulator, LacI family                              |                |
|            | GGGAGCGCTCCCA*                  | K379_04286 | -15                   | 3.52                                         | 0.0740   | 32                | beta-galactosidase                                                  | GH1            |
|            | TGGGAGCGCTCCCA                  | K379_04289 | -147                  | 5.43                                         | < 0.0001 | 9                 | cebE; cellobiose transport system substrate-binding protein         |                |
|            | TGGGAGCGCTCCCA                  | K379_04837 | -32                   | 1.38                                         | 0.0281   | 262               | Acetyltransferases                                                  |                |
|            | TGGGAGCGCTCCCA                  | K379_04855 | -40                   | 5.50                                         | 0.0164   | 7                 | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH12_CBM2      |
|            |                                 | K379_04854 | -162                  | -0.32                                        | NS       | 4251              | Predicted NTP pyrophosphohydrolase                                  |                |
|            | TGGGAGCGCTCCCA                  | K379_04935 | -877                  | 6.30                                         | < 0.0001 | 4                 | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH6_CBM2       |
|            |                                 | K379_04936 | -229                  | 7.79                                         | < 0.0001 | 2                 | cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)                     | GH48_GH74_CBM2 |
|            | TGGAAACGCTCCCA*                 | K379_05573 | -131                  | -0.08                                        | NS       | 3312              | Beta-lactamase class C and other penicillin binding proteins        |                |
|            | TGGGAGCGCTCCCA*                 | K379_05757 | -51                   | 1.99                                         | < 0.0001 | 125               | ABC.MS.S; multiple sugar transport system substrate-binding protein |                |

<sup>a</sup>Position relative to start of gene

<sup>b</sup>log<sub>2</sub> fold change and ranked fold change for growth on AFEX corn stover relative to glucose

\*noncanonical CebR binding sequence

**Table A3.4.** Expression and annotation of genes predicted to be directly controlled by the CebR repressor (continued)

|                 | Binding Site(s) | LocusID         | Position <sup>a</sup> | log <sub>2</sub> fold |          |          | Rank <sup>b</sup>                                                   | Annotation                          | CAZY          |
|-----------------|-----------------|-----------------|-----------------------|-----------------------|----------|----------|---------------------------------------------------------------------|-------------------------------------|---------------|
|                 |                 |                 |                       | change <sup>b</sup>   | pvalue   |          |                                                                     |                                     |               |
| ATexAB-D23      | TGGGAGCGCTCCCA  | B082DRAFT_01056 | -105                  | 7.22                  | < 0.0001 | 3        | Uncharacterized protein conserved in bacteria                       | CBM2_AA10                           |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_01057 | -220                  | 0.02                  | NS       | 3217     | Transcriptional regulators                                          |                                     |               |
|                 |                 | B082DRAFT_01058 | -161                  | 6.82                  | < 0.0001 | 7        | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | CBM2_GH5_CBM35                      |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_02427 | -140                  | 3.72                  | < 0.0001 | 58       | Beta-1,4-xylanase                                                   | GH10_CBM3                           |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_02566 | -349                  | 3.59                  | < 0.0001 | 62       | Predicted xylanase/chitin deacetylase                               | CE4_CBM36                           |               |
|                 |                 | B082DRAFT_02567 | -59                   | 7.11                  | < 0.0001 | 4        | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | CBM2                                |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_02898 | -127                  | 1.30                  | 0.0026   | 547      | Transcriptional regulators                                          |                                     |               |
|                 | GGGAGCGCTCCCA*  | B082DRAFT_02899 | -14                   | 4.61                  | < 0.0001 | 24       | beta-galactosidase                                                  | GH1                                 |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_02902 | -143                  | 5.89                  | < 0.0001 | 11       | cebE; cellobiose transport system substrate-binding protein         |                                     |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_02996 | -222                  | 1.04                  | NS       | 781      | Uncharacterized protein conserved in bacteria                       |                                     |               |
|                 |                 | B082DRAFT_02997 | -56                   | 3.24                  | < 0.0001 | 77       | Uncharacterized protein contain chitin-binding domain type 3        | CBM2_GH5_CBM37                      |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_04414 | -45                   | 4.38                  | < 0.0001 | 31       | Uncharacterized protein conserved in bacteria                       | AA10                                |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_05432 | -356                  | 2.00                  | NS       | 822      | glutamate N-acetyltransferase/amino-acid acetyltransferase          |                                     |               |
|                 |                 | B082DRAFT_05433 | -53                   | 4.05                  | < 0.0001 | 38       | Endoglucanase                                                       | GH5_CBM2                            |               |
|                 | TGGGAGCGCTCCCA  | B082DRAFT_05492 | -72                   | 3.79                  | < 0.0001 | 52       | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH9_CBM3_CBM2                       |               |
|                 | GGGAGCGCTCCCA*  | B082DRAFT_05579 | -109                  | 2.02                  | < 0.0001 | 229      | ABC-type dipeptide transport system, periplasmic component          |                                     |               |
|                 |                 | B082DRAFT_05580 | -125                  | -0.92                 | NS       | 6006     | Transcriptional regulators                                          |                                     |               |
|                 |                 | B082DRAFT_05958 | 1                     | 2.53                  | < 0.0001 | 143      | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH74_CBM2                           |               |
|                 |                 | B082DRAFT_06747 | -63                   | 4.89                  | < 0.0001 | 17       | Cellobiohydrolase A (1,4-beta-cellobiosidase A)                     | GH9_CBM4_CBM2                       |               |
|                 | S. flavigriseus | TGGGAGCGCTCCCA  | SFLA_00178            | -141                  | 4.24     | < 0.0001 | 23                                                                  | putative actNOG04893 family protein | GH9_CBM4_CBM2 |
| GGGAGCGCTCCCA*  |                 | SFLA_00246      | -59                   | 4.04                  | < 0.0001 | 27       | chitin-binding protein                                              | AA10                                |               |
| TGGGAGCGCTCCCA  |                 | SFLA_00603      | -46                   | 5.96                  | 0.0004   | 5        | putative actNOG01555 family protein                                 | GH6                                 |               |
|                 |                 | SFLA_00604      | -177                  | 0.93                  | NS       | 989      | LacI family transcriptional regulator                               |                                     |               |
| TGGGAGCGCTCCCA  |                 | SFLA_00912      | -53                   | 4.64                  | < 0.0001 | 16       | glucan 1,3-beta-glucosidase                                         | GH5_CBM37                           |               |
|                 |                 | SFLA_00913      | -242                  | -0.10                 | NS       | 3751     | putative actNOG14362 family protein                                 | GH87_CBM32                          |               |
| TGGGAGCGCTCCCA  |                 | SFLA_01241      | -67                   | 5.13                  | 0.0398   | 10       | chitin-binding protein                                              | CBM2_AA10                           |               |
|                 |                 | SFLA_01242      | -141                  | -1.18                 | NS       | 6005     | Kegg K07110 family protein                                          |                                     |               |
| TGGGAGCGCTCCCA* |                 | SFLA_01247      | -50                   | 3.59                  | 0.0378   | 38       | ABC.MS.S; multiple sugar transport system substrate-binding protein |                                     |               |
| TGGGAGCGCTCCCA  |                 | SFLA_01777      | -554                  | 1.96                  | NS       | 236      | hypothetical protein                                                |                                     |               |
|                 |                 | SFLA_01778      | -47                   | 3.64                  | 0.0369   | 36       | endoglucanase                                                       | CBM2_GH5_CBM35                      |               |
| GGGAGCGCTCCCA*  |                 | SFLA_03840      | -125                  | -0.31                 | NS       | 4360     | LacI family transcriptional regulator                               |                                     |               |
|                 |                 | SFLA_03841      | -111                  | 0.30                  | NS       | 2447     | peptide/nickel transport system substrate-binding protein           |                                     |               |
| TGGGAGCGCTCCCA  |                 | SFLA_04057      | -122                  | 2.67                  | NS       | 91       | mannan endo-1,4-beta-mannosidase                                    | CBM2_GH5_CBM37                      |               |
| TGGGAGCGCTCCCA  |                 | SFLA_04126      | -147                  | 6.64                  | < 0.0001 | 2        | cebE; cellobiose transport system substrate-binding protein         |                                     |               |
| GGGAGCGCTCCCA*  |                 | SFLA_04129      | -15                   | 3.80                  | 0.0001   | 33       | beta-galactosidase                                                  | GH1                                 |               |
| TGGGAGCGCTCCCA  |                 | SFLA_04130      | -127                  | 0.26                  | NS       | 2586     | LacI family transcriptional regulator                               |                                     |               |
| TGGGAGCGCTCCCA  |                 | SFLA_05761      | -718                  | -0.13                 | NS       | 3854     | hypothetical protein                                                |                                     |               |
|                 |                 | SFLA_05762      | -73                   | 3.45                  | 0.0364   | 42       | putative actNOG00503 family protein                                 | GH74_CBM2                           |               |
| TGGGAGCGCTCCCA  |                 | SFLA_05979      | -32                   | 0.97                  | NS       | 940      | ribosomal-protein-alanine N-acetyltransferase                       |                                     |               |
| TGGGAGCGCTCCCA  |                 | SFLA_06095      | -197                  | 7.06                  | 0.0008   | 1        | hypothetical protein                                                |                                     |               |
|                 |                 | SFLA_06096      | -156                  | 5.38                  | 0.0004   | 6        | putative actNOG01534 family protein                                 | GH48_GH74_CBM2                      |               |
| TGGGAGCGCTCCCA  |                 | SFLA_06342      | -94                   | 3.12                  | 0.0061   | 55       | alpha-N-arabinofuranosidase                                         | GH51_CBM2                           |               |
|                 |                 | SFLA_06343      | -782                  | 0.26                  | NS       | 2568     | Lrp/AsnC family transcriptional regulator, regulator for asnA, asnC |                                     |               |
| TGGGAGCGCTCCCA  |                 | SFLA_06402      | -84                   | 4.46                  | < 0.0001 | 21       | putative actNOG03075 family protein                                 | GH12_CBM2                           |               |

<sup>a</sup>Position relative to start of gene<sup>b</sup>log<sub>2</sub> fold change and ranked fold change for growth on AFEX corn stover relative to glucose

\*noncanonical CebR binding sequence

**Table A3.4.** Expression and annotation of genes predicted to be directly controlled by the CebR repressor (continued)

|                   | Binding Site(s)                  | LocusID     | Position <sup>a</sup> | log <sub>2</sub> fold |          | Rank <sup>b</sup> | Annotation                                                  | CAZY           |
|-------------------|----------------------------------|-------------|-----------------------|-----------------------|----------|-------------------|-------------------------------------------------------------|----------------|
|                   |                                  |             |                       | change <sup>b</sup>   | pvalue   |                   |                                                             |                |
| LaPpAH-95         | TGGGAGCGCTCCCA ; TGGGAGCGCTCCCA  | B087_00242  | -60; -105             | 7.11                  | < 0.0001 | 8                 | endoglucanase                                               | CBM2_GH5_CBM35 |
|                   |                                  | B087_00243  | -196; -151            | 0.49                  | NS       | 1391              | unknown function                                            |                |
|                   | TGGGAGCGCTCCCA                   | B087_00244  | -97                   | 8.72                  | < 0.0001 | 2                 | chitin-binding protein                                      | CBM2_AA10      |
|                   | TGGGAGCGCTCCCA                   | B087_00622  | -64                   | 2.57                  | 0.0039   | 128               | putative actNOG03075 family protein                         | GH12_CBM13     |
|                   | GGGAGCGCTCCCA*                   | B087_00624  | -59                   | 6.52                  | < 0.0001 | 13                | glucan 1,3-beta-glucosidase                                 | GH5_CBM37      |
|                   |                                  | B087_00625  | -362                  | 1.94                  | 0.0130   | 223               | putative isomerase                                          | CBM66_GH0      |
|                   | TGGGAGCGCTCCCA ; TGGGAGCGATCCCA* | B087_01830  | 32; -13               | 5.13                  | < 0.0001 | 24                | putative actNOG04893 family protein                         | GH9_CBM4_CBM2  |
|                   | TGGGAGCGCTCCCA                   | B087_03402  | -147                  | 1.12                  | NS       | 549               | hypothetical protein                                        |                |
|                   |                                  | B087_03403  | -65                   | 4.91                  | < 0.0001 | 30                | alpha-N-arabinofuranosidase                                 | GH51_CBM2      |
|                   | TGGGAGCGCTCCCA                   | B087_03467  | -48                   | 8.06                  | < 0.0001 | 5                 | chitin-binding protein                                      | AA10           |
|                   | GGGAGCGCTCCCA*                   | B087_04423  | -109                  | 2.43                  | < 0.0001 | 148               | peptide/nickel transport system substrate-binding protein   |                |
|                   |                                  | B087_04424  | -125                  | -0.43                 | NS       | 4442              | unknown function                                            |                |
|                   | TGGGAGCGCTCCCA                   | B087_05321  | -62                   | 5.89                  | < 0.0001 | 17                | glucan 1,3-beta-glucosidase                                 | GH5_CBM2       |
|                   | GGGAGCGCTCCCA*                   | B087_05341  | 1878                  | 1.70                  | 0.0049   | 288               | amino acid adenylation domain                               |                |
|                   | TGGGAGCGCTCCCA                   | B087_05389  | -85                   | 2.65                  | < 0.0001 | 119               | putative actNOG05475 family protein                         | GH9_CBM3_CBM2  |
|                   |                                  | B087_05390  | -201                  | -0.16                 | NS       | 3465              | unknown function                                            |                |
|                   | TGGGAGCGCTCCCA                   | B087_05593  | -141                  | 5.78                  | < 0.0001 | 19                | endo-1,4-beta-xylanase                                      | GH10_CBM3      |
|                   |                                  | B087_05594  | -633                  | -0.11                 | NS       | 3305              | putative actNOG12105 family protein                         |                |
|                   | TGGAAGCGCTCCCA*                  | B087_05778  | -505                  | 8.36                  | 0.0002   | 4                 | licheninase                                                 | GH16           |
|                   |                                  | B087_05779  | -65                   | 6.90                  | 0.0018   | 11                | putative actNOG01555 family protein                         | GH6            |
|                   | TGGAAGCGCTCCCA*                  | B087_05783  | -180                  | 4.01                  | 0.0319   | 49                | LacI family transcriptional regulator                       |                |
|                   |                                  | B087_05784  | -87                   | 5.08                  | 0.0127   | 27                | beta-galactosidase                                          | GH1_GH95       |
|                   | TGGGAGCGCTTCCA*                  | B087_05785  | -47                   | 2.39                  | NS       | 154               | cebE; cellobiose transport system substrate-binding protein |                |
|                   | TGGGAGCGCTCCCA                   | B087_06016  | -47                   | 5.03                  | < 0.0001 | 28                | mannan endo-1,4-beta-mannosidase                            | CBM2_GH5_CBM37 |
|                   |                                  | B087_06017  | -154                  | 2.43                  | 0.0002   | 149               | galactoside O-acetyltransferase                             |                |
|                   | TGGGAGCGCTCCCA                   | B087_06094  | -144                  | 7.63                  | < 0.0001 | 6                 | cebE; cellobiose transport system substrate-binding protein |                |
|                   | TGGGAGCGCTCCCA                   | B087_06098  | -126                  | 1.78                  | 0.0053   | 261               | LacI family transcriptional regulator                       |                |
|                   | TGGGAGCGCTCCCA                   | B087_06218  | -72                   | 4.20                  | < 0.0001 | 42                | putative actNOG00503 family protein                         | GH74_CBM2      |
|                   | TGGGAGCGCTCCCA                   | B087_06575  | -330                  | 6.51                  | < 0.0001 | 14                | chitin deacetylase                                          | CE4_CBM36      |
|                   |                                  | B087_06576  | -43                   | 8.66                  | 0.0021   | 3                 | putative actNOG05965 family protein                         | CBM2_Cip1      |
| <i>S. griseus</i> | TGGGAGCGCTCCCA                   | SGR_199     | -331                  | 1.60                  | 0.0377   | 388               | cellulose-binding protein                                   | CBM2_AA10      |
|                   | GGGAGCGCTCCCA*                   | SGR_217     | -109                  | -1.58                 | NS       | 5794              | pectate lyase                                               | PL1_CBM35      |
|                   | TGGGAGCGCTCCCA                   | SGR_2445    | -30                   | 1.58                  | 0.0025   | 396               | cellulase                                                   | GH6            |
|                   | TGGGAGCGCTCCCA                   | SGR_4734    | -255                  | -1.71                 | 0.0482   | 6012              | transport protein                                           |                |
|                   |                                  | SGR_4735    | -121                  | 3.29                  | < 0.0001 | 15                | cebE; cellobiose transport system substrate-binding protein |                |
|                   | TGGGAGCGCTCCCA                   | SGR_4739    | -106                  | 0.85                  | NS       | 1035              | LacI family transcriptional regulator                       |                |
|                   | TGGGAGCGCTCCCA                   | SGR_6927    | -303                  | -0.54                 | NS       | 3500              | hypothetical protein                                        |                |
|                   |                                  | SGR_6928    | -119                  | 2.59                  | < 0.0001 | 62                | cellulase                                                   | CBM3_GH6       |
| ScaeMP-e83        | TGGGAGCGCTCCCA                   | SSC83_00187 | -30                   | 2.29                  | 0.0013   | 166               | hypothetical protein                                        | GH6            |
|                   | TGGGAGCGCTCCCA                   | SSC83_00188 | -215                  | 1.41                  | NS       | 625               | hypothetical protein                                        |                |
|                   | TGGGAGCGCTCCCA                   | SSC83_02183 | -106                  | 1.03                  | NS       | 1149              | LacI family transcriptional regulator                       |                |
|                   | TGGGAGCGCTCCCA                   | SSC83_02184 | -15                   | 3.54                  | < 0.0001 | 21                | beta-galactosidase                                          | GH1            |
|                   | TGGGAGCGCTCCCA                   | SSC83_02187 | -78                   | 4.20                  | < 0.0001 | 5                 | cebE; cellobiose transport system substrate-binding protein |                |
|                   |                                  | SSC83_02188 | -255                  | 1.09                  | NS       | 1051              | aspartate-alanine antiporter                                |                |

<sup>a</sup>Position relative to start of gene<sup>b</sup>log<sub>2</sub> fold change and ranked fold change for growth on AFEX corn stover relative to glucose

\*noncanonical CebR binding sequence



**Dataset A3.1.** Comparative genomic analysis of all ABC sugar transporters in *Streptomyces*. Available online at <http://journals.plos.org/plosbiology/article/asset?unique&id=info:doi/10.1371/journal.pbio.1002475.s001>.

**Dataset A3.2.** RNAseq differential expression data table for SirexAA-E, DpondAA-B6, ATexAB-D23, LaPpAH-95, *S. griseus* and ScaeMP-e83. Available online at <http://journals.plos.org/plosbiology/article/asset?unique&id=info:doi/10.1371/journal.pbio.1002475.s002>.

**Appendix 4: Supplementary Material for Chapter 4**

**Previously published as Supplementary Material:** Gina R. Lewin, Amanda L. Johnson, Rolando D. Moreira-Soto, Kailene Perry, Adam J. Book, Heidi A. Horn, Adrián A. Pinto-Tomás, and Cameron R. Currie (2016). Cellulose-Enriched Microbial Communities from Leaf-Cutter Ant (*Atta colombica*) Refuse Dumps Vary in Taxonomic Composition and Degradation Ability. *PLoS ONE*, 11(3): e0151840.

**Author contributions are the same as in Chapter 4.**

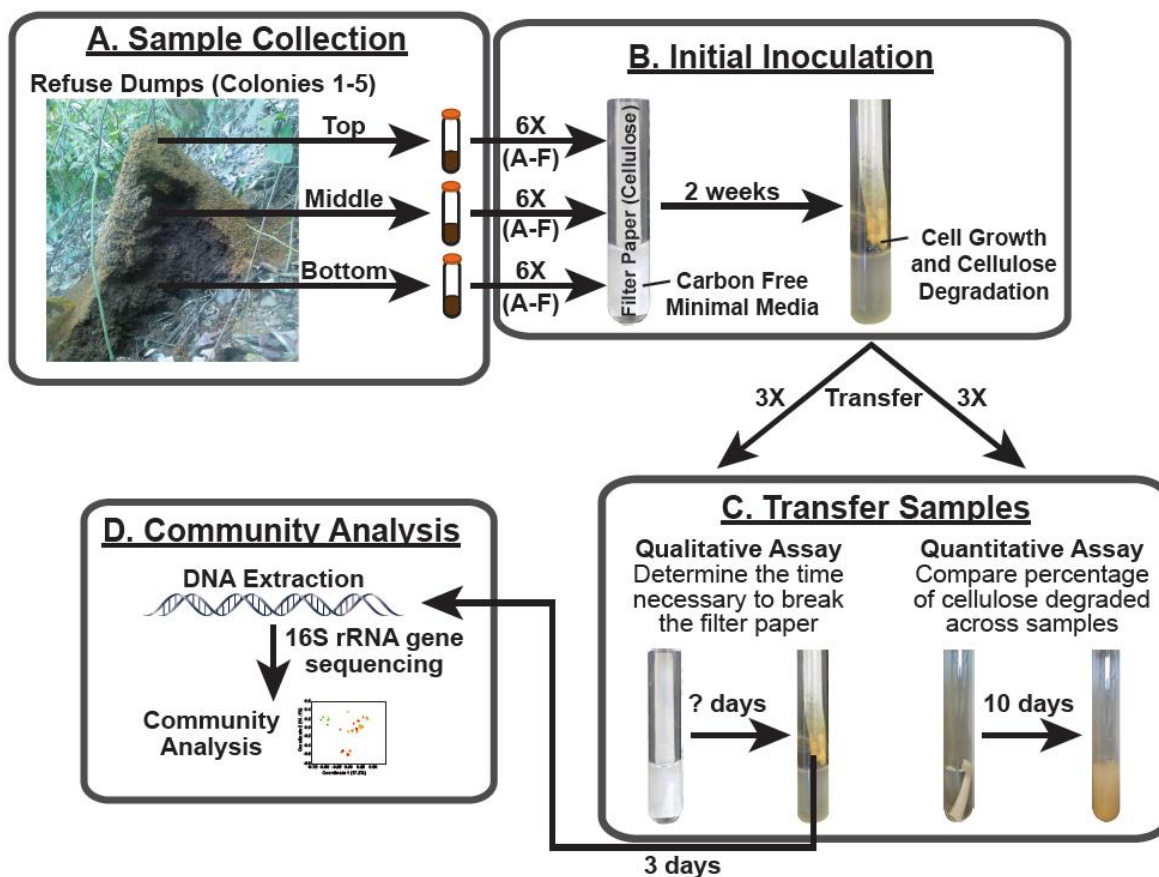


**Figure A4.1.** (A) The top (freshest), middle, and bottom (oldest) layers of five refuse dumps were collected. (B) Six unique samples from each collection were inoculated into test tubes with cellulose filter paper as the only carbon source in minimal media. These samples were allowed to grow for two weeks and then were transferred into new cultures. (C) Transfer cultures included qualitative and quantitative tubes. Qualitative tubes were visually observed daily to determine the number of days necessary for the microbial community to break apart the filter paper. Quantitative tubes were allowed to grow for ten days and then were processed to compare the percentage of cellulose degraded across samples. (D) Three days after the filter paper broke apart for each set of qualitative tubes, samples were collected for DNA extraction, 16S rRNA gene amplicon sequencing, and community analysis. A three day period between initial cellulose degradation and DNA extraction ensured that enough growth was present in the tube to analyze the microbial community.

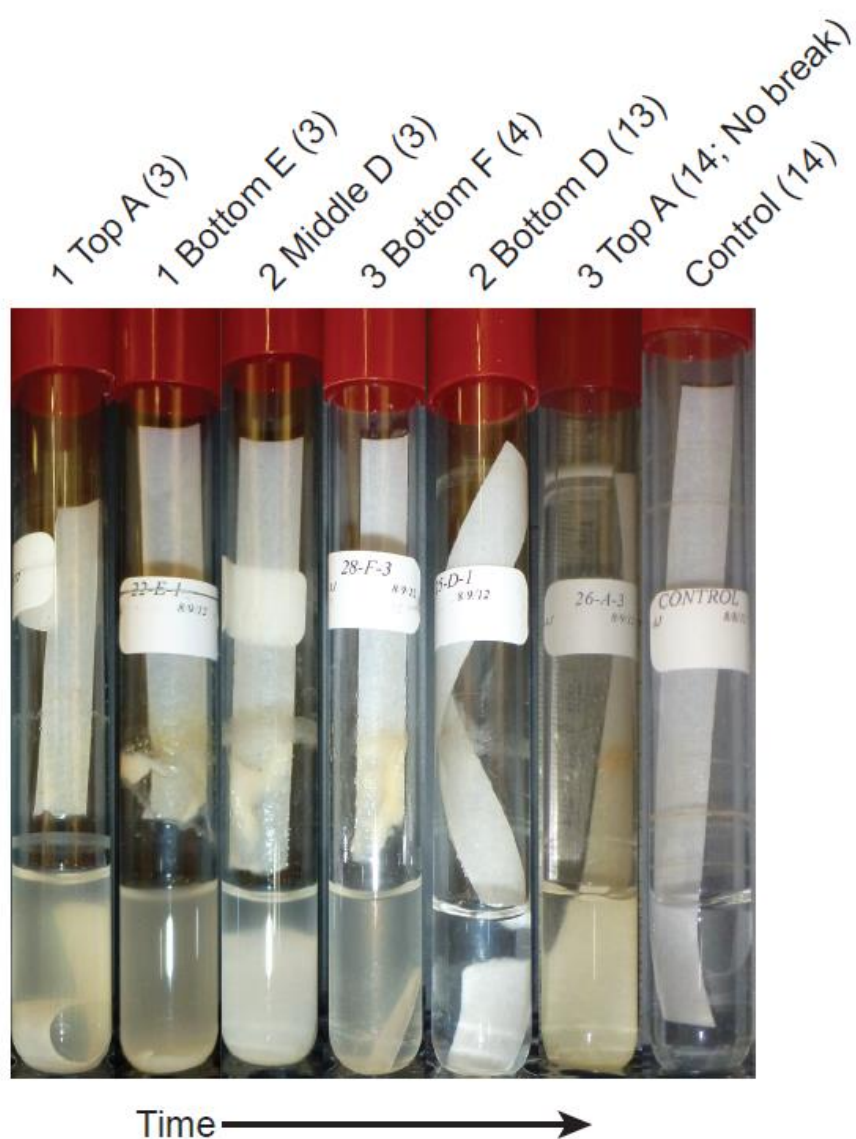


**Figure A4.2. Representative qualitative samples.** Images were taken on the day the filter paper broke in half (indicated in parentheses after the sample name) for all cultures except for 3 Top A and the control, which were imaged on day 14.

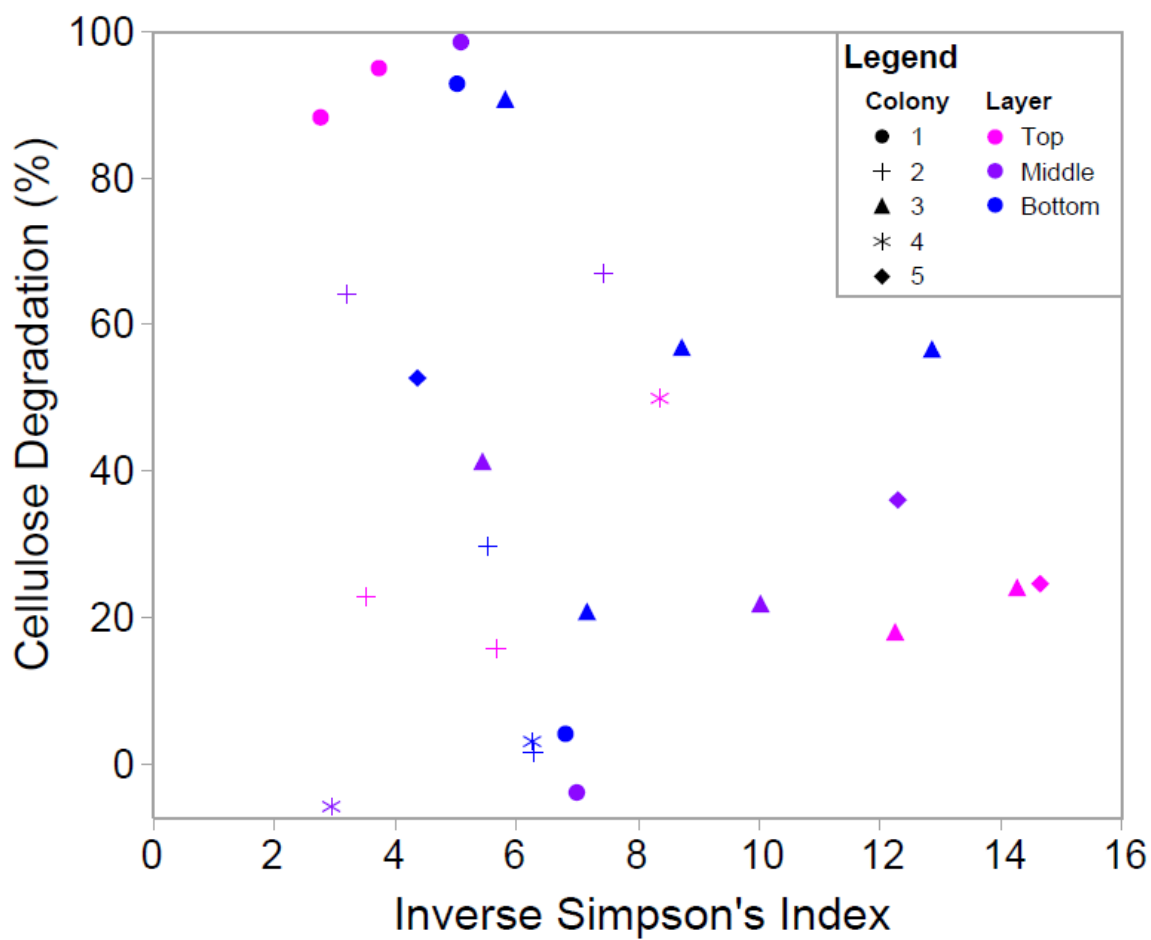


**Figure A4.3. Relationship between the percentage of cellulose degraded and diversity (Inverse Simpson's Index) across sequenced samples.** Sample shape indicates colony. Sample color indicates layer.

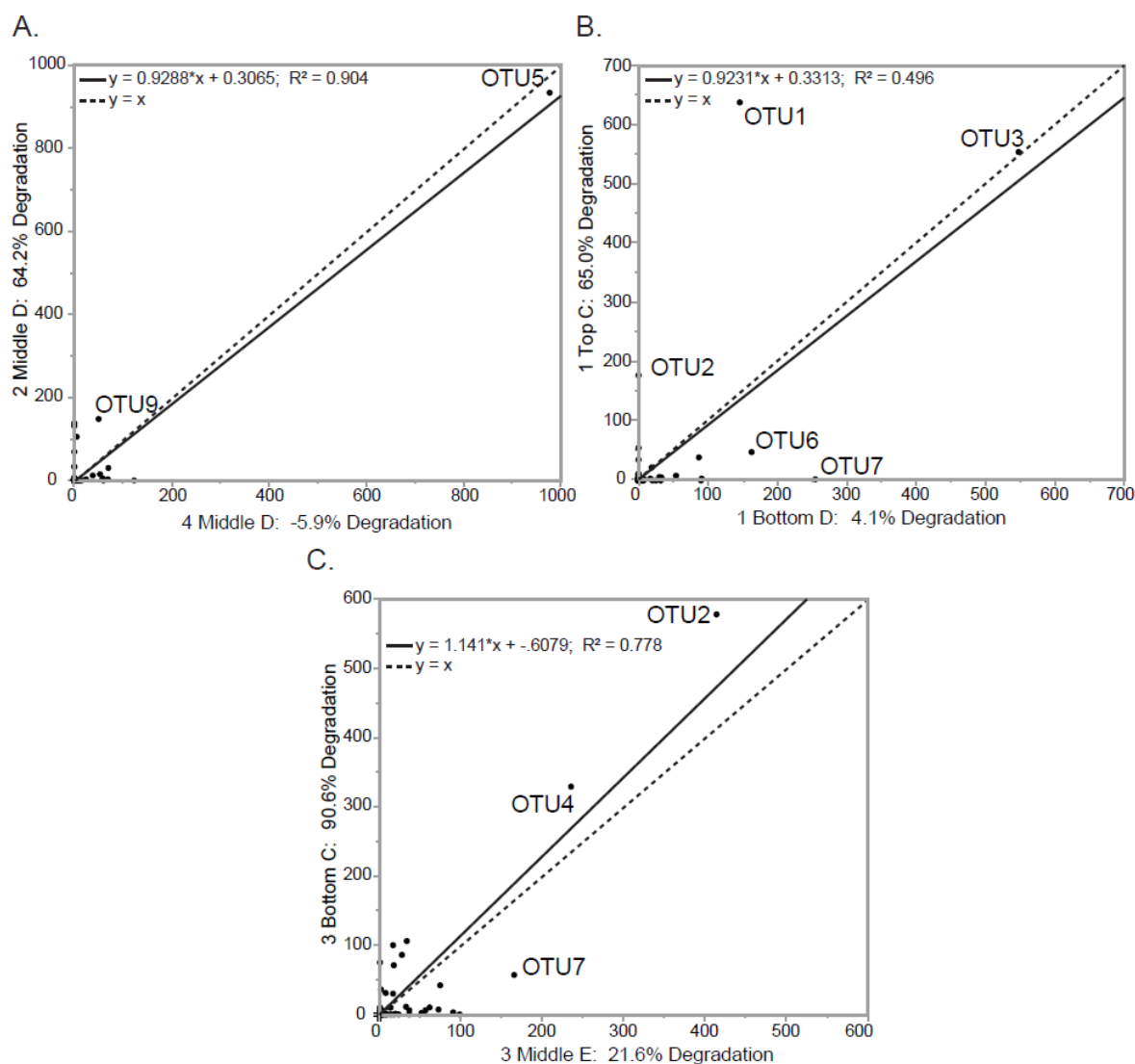


**Figure A4.4. Pairwise comparisons of OTU abundance between samples with similar diversity.** Samples were selected with high similarity in diversity but large differences in cellulolytic ability according to the cladogram in Figure 4.4. The abundance of reads for each OTU is plotted for the indicated samples. We show the best fit line for each plot to indicate the correlation between read abundances of each sample and the  $y = x$  line. Abundant OTUs are indicated. See Table 4.2 or Dataset A4.1 for taxonomic classification of each OTU.

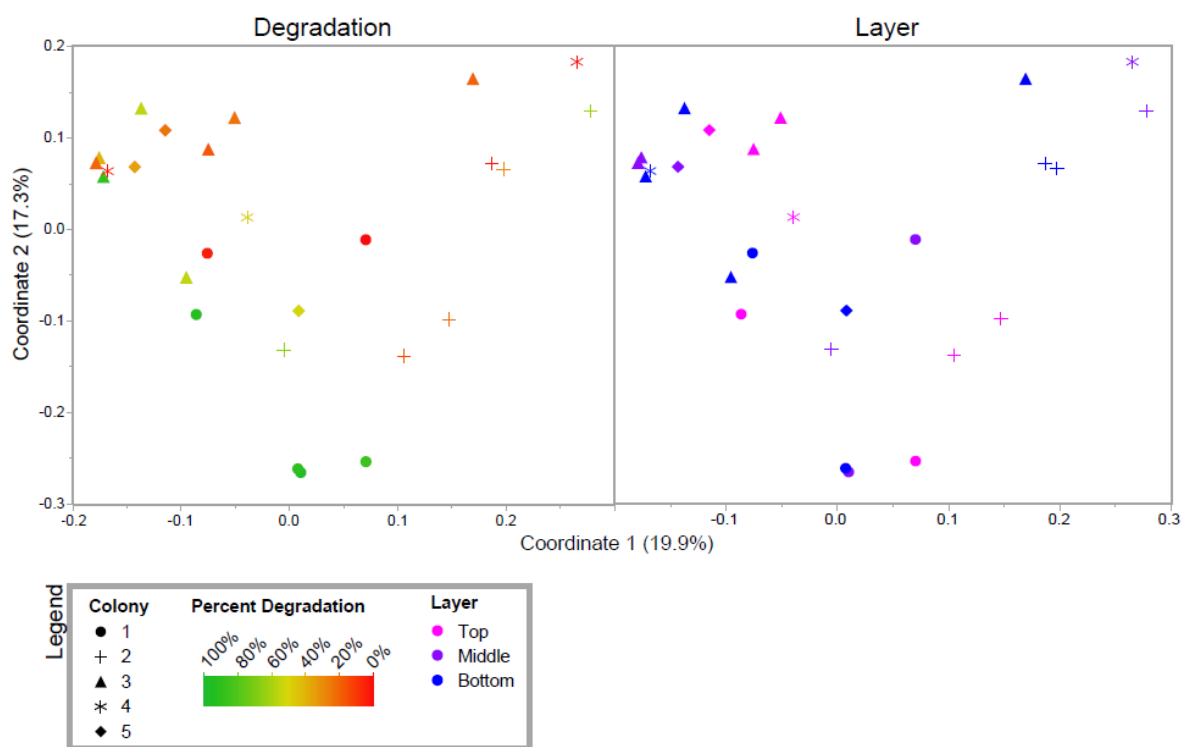


**Figure A4.5. PCoA clustering of the weighted Unifrac metric of similarity among samples.** Sample shape indicates colony. Sample color indicates degradation (A) or layer (B).

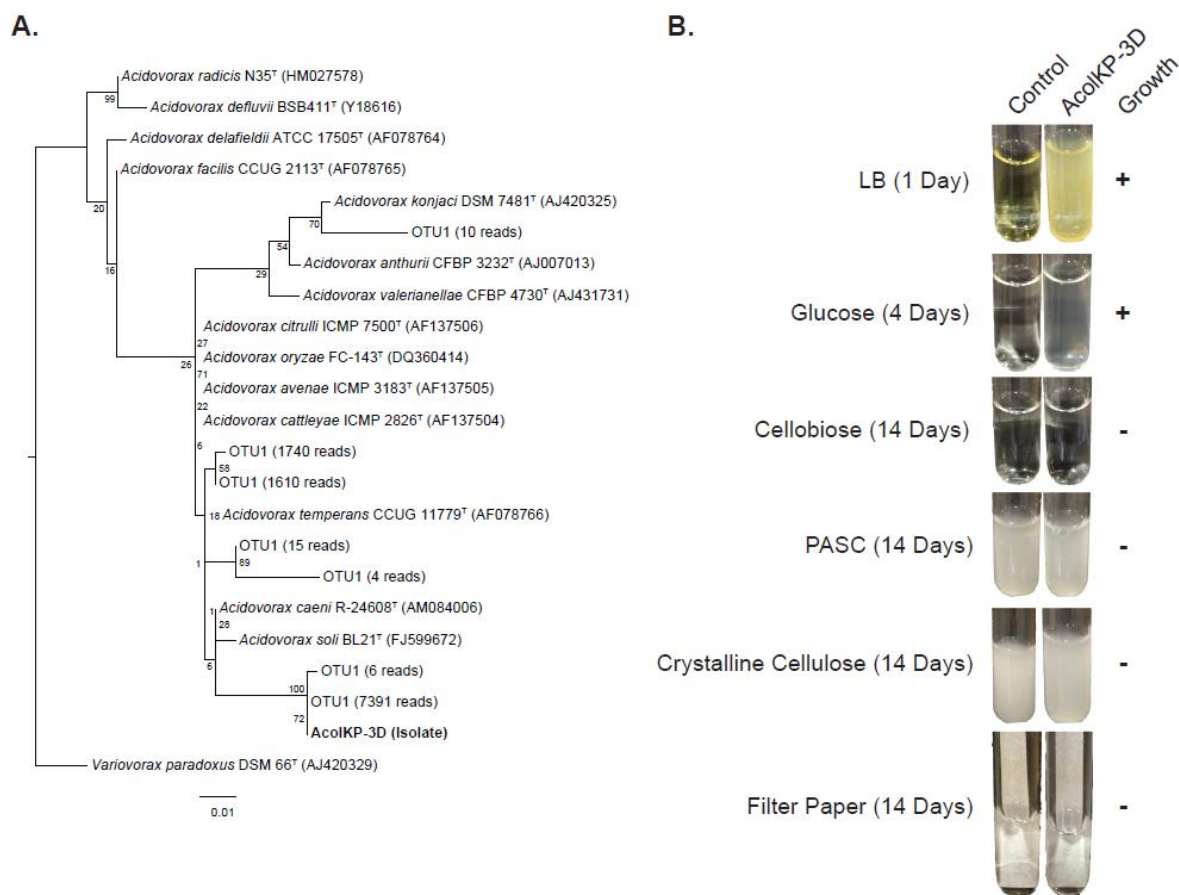


**Figure A4.6. Phylogeny and growth of an *Acidovorax* isolate.** (A) Maximum likelihood comparison of the 16S rRNA gene sequence of the AcolKP-3D isolate with all type strains of *Acidovorax*, the outgroup *Variovorax paradoxus*, and amplicon sequencing reads from OTU1. The number of reads matching each sequence of OTU1 is indicated in parentheses. (B) Representative images of control and inoculated cultures of AcolKP-3D on LB, glucose, cellobiose, phosphoric acid swollen cellulose (PASC), crystalline cellulose, and filter paper. Images were taken after the number of days indicated. Analysis of growth is indicated, based on visual assessment of liquid cultures and plating.



**Table A4.1.** Qualitative and quantitative data for all enrichment communities.

| Enriched Community | Qualitative Data |               |                     | Quantitative Data |                    |
|--------------------|------------------|---------------|---------------------|-------------------|--------------------|
|                    | Growth?          | Days to Break |                     | Average           | Standard Deviation |
|                    |                  | Filter Paper  | Censor <sup>1</sup> |                   |                    |
| 1 Top A            | yes              | 3.7           | 0                   | 88.3%             | 1.5%               |
| 1 Top B            | yes              | 5.0           | 0                   | 96.1%             | 1.5%               |
| 1 Top C            | yes              | 4.0           | 0                   | 95.0%             | 2.2%               |
| 1 Top D            | yes              | 3.0           | 0                   | 96.1%             | 0.4%               |
| 1 Top E            | yes              | 5.0           | 0                   | 75.3%             | 2.5%               |
| 1 Top F            | yes              | 3.3           | 0                   | 96.3%             | 4.3%               |
| 1 Middle A         | yes              | 4.3           | 0                   | 70.4%             | 4.3%               |
| 1 Middle B         | no               | 14.0          | 1                   | -1.6%             | 1.8%               |
| 1 Middle C         | no               | 14.0          | 1                   | -3.9%             | 1.1%               |
| 1 Middle D         | no               | 14.0          | 1                   | 4.6%              | 3.8%               |
| 1 Middle E         | yes              | 2.0           | 0                   | 98.6%             | 1.8%               |
| 1 Middle F         | yes              | 14.0          | 1                   | 6.1%              | 3.1%               |
| 1 Bottom A         | yes              | 6.0           | 0                   | 51.1%             | 9.5%               |
| 1 Bottom B         | yes              | 3.0           | 0                   | 94.0%             | 2.0%               |
| 1 Bottom C         | no               | 14.0          | 1                   | 2.3%              | 1.1%               |
| 1 Bottom D         | yes              | 14.0          | 1                   | 4.1%              | 2.0%               |
| 1 Bottom E         | yes              | 3.0           | 0                   | 92.9%             | 3.6%               |
| 1 Bottom F         | yes              | 10.7          | 0                   | 17.8%             | 6.5%               |
| 2 Top A            | yes              | 10.3          | 0                   | 22.8%             | 8.4%               |
| 2 Top B            | yes              | 7.0           | 0                   | 15.8%             | 1.8%               |
| 2 Top C            | yes              | 5.0           | 0                   | 26.6%             | 1.5%               |
| 2 Top D            | yes              | 8.5           | 0                   | 29.2%             | 2.9%               |
| 2 Top E            | no               | 14.0          | 1                   | 1.6%              | 1.8%               |
| 2 Top F            | yes              | 2.0           | 0                   | 72.2%             | 2.8%               |
| 2 Middle A         | yes              | 5.3           | 0                   | 45.0%             | 1.2%               |
| 2 Middle B         | yes              | 6.3           | 0                   | 67.1%             | 5.2%               |
| 2 Middle C         | yes              | 4.7           | 0                   | 28.3%             | 2.3%               |
| 2 Middle D         | yes              | 3.7           | 0                   | 64.2%             | 4.4%               |
| 2 Middle E         | yes              | 5.0           | 0                   | 55.1%             | 4.8%               |
| 2 Middle F         | yes              | 4.0           | 0                   | 72.2%             | 5.5%               |
| 2 Bottom A         | yes              | 7.0           | 0                   | 29.6%             | 2.6%               |
| 2 Bottom B         | yes              | 8.0           | 0                   | 17.7%             | 5.9%               |
| 2 Bottom C         | yes              | 9.3           | 0                   | 21.3%             | 2.1%               |
| 2 Bottom D         | yes              | 14.0          | 1                   | 1.6%              | 2.1%               |
| 2 Bottom E         | yes              | 11.0          | 0                   | 21.5%             | 0.8%               |
| 2 Bottom F         | yes              | 14.0          | 1                   | 13.0%             | 4.3%               |
| 3 Top A            | yes              | 14.0          | 1                   | 24.1%             | 2.6%               |
| 3 Top B            | yes              | 11.3          | 0                   | 37.0%             | 5.8%               |
| 3 Top C            | yes              | 14.0          | 1                   | 28.6%             | 5.5%               |
| 3 Top D            | yes              | 14.0          | 1                   | 17.8%             | 5.7%               |
| 3 Top E            | yes              | 14.0          | 1                   | 26.7%             | 6.6%               |
| 3 Top F            | yes              | 11.0          | 0                   | 17.8%             | 1.7%               |
| 3 Middle A         | yes              | 6.0           | 0                   | 41.0%             | 8.0%               |
| 3 Middle B         | yes              | 8.3           | 0                   | 54.0%             | 18.4%              |

<sup>1</sup>A censor of 1 indicates that the days to break the filter paper was not determined (the experiment ended before the filter paper broke in half).

**Table A4.1.** Qualitative and quantitative data for all enrichment communities (continued).

| Enriched Community | Growth? | Qualitative Data |                     | Quantitative Data |                    |
|--------------------|---------|------------------|---------------------|-------------------|--------------------|
|                    |         | Filter Paper     | Censor <sup>1</sup> | Average           | Standard Deviation |
|                    |         | Days to Break    |                     |                   |                    |
| 3 Middle C         | yes     | 10.7             | 0                   | 24.2%             | 6.0%               |
| 3 Middle D         | yes     | 9.7              | 0                   | 19.4%             | 4.0%               |
| 3 Middle E         | yes     | 14.0             | 1                   | 21.6%             | 5.0%               |
| 3 Middle F         | yes     | 8.3              | 0                   | 46.6%             | 4.2%               |
| 3 Bottom A         | yes     | 7.7              | 0                   | 73.3%             | 6.3%               |
| 3 Bottom B         | yes     | 7.7              | 0                   | 56.7%             | 0.0%               |
| 3 Bottom C         | yes     | 5.3              | 0                   | 90.6%             | 5.2%               |
| 3 Bottom D         | yes     | 9.0              | 0                   | 56.6%             | 3.6%               |
| 3 Bottom E         | yes     | 8.3              | 0                   | 68.4%             | 6.0%               |
| 3 Bottom F         | yes     | 4.0              | 0                   | 20.5%             | 6.3%               |
| 4 Top A            | yes     | 4.3              | 0                   | 15.9%             | 0.6%               |
| 4 Top B            | yes     | 6.0              | 0                   | 40.3%             | 2.5%               |
| 4 Top C            | yes     | 14.0             | 1                   | 49.8%             | 1.4%               |
| 4 Top D            | yes     | 8.3              | 0                   | 9.7%              | 2.0%               |
| 4 Top E            | yes     | 14.0             | 1                   | 1.3%              | 2.8%               |
| 4 Top F            | no      | 14.0             | 1                   | -2.3%             | 1.5%               |
| 4 Middle B         | yes     | 6.0              | 0                   | 6.0%              | 2.8%               |
| 4 Middle C         | yes     | 14.0             | 1                   | 2.3%              | 1.0%               |
| 4 Middle D         | yes     | 14.0             | 1                   | -5.9%             | 0.6%               |
| 4 Middle E         | yes     | 14.0             | 1                   | 2.3%              | 1.1%               |
| 4 Middle F         | yes     | 14.0             | 1                   | -5.2%             | 0.3%               |
| 4 Bottom A         | yes     | 14.0             | 1                   | 5.1%              | 1.6%               |
| 4 Bottom B         | yes     | 14.0             | 1                   | 5.7%              | 5.0%               |
| 4 Bottom D         | yes     | 14.0             | 1                   | 4.0%              | 3.8%               |
| 4 Bottom E         | yes     | 14.0             | 1                   | 3.1%              | 2.4%               |
| 4 Bottom F         | yes     | 14.0             | 1                   | 2.8%              | 2.7%               |
| 5 Top A            | yes     | 7.3              | 0                   | 24.6%             | 1.6%               |
| 5 Top B            | yes     | 4.3              | 0                   | 37.7%             | 0.9%               |
| 5 Top C            | yes     | 4.0              | 0                   | 21.2%             | 2.2%               |
| 5 Top D            | yes     | 6.0              | 0                   | 20.7%             | 2.9%               |
| 5 Top E            | yes     | 5.0              | 0                   | 48.0%             | 2.0%               |
| 5 Top F            | yes     | 4.0              | 0                   | 23.7%             | 4.3%               |
| 5 Middle A         | yes     | 5.3              | 0                   | 15.7%             | 1.1%               |
| 5 Middle B         | yes     | 5.0              | 0                   | 25.4%             | 3.3%               |
| 5 Middle C         | yes     | 4.3              | 0                   | 24.2%             | 2.1%               |
| 5 Middle D         | yes     | 4.7              | 0                   | 51.1%             | 1.8%               |
| 5 Middle E         | yes     | 7.0              | 0                   | 36.0%             | 2.9%               |
| 5 Middle F         | yes     | 4.0              | 0                   | 34.8%             | 7.5%               |
| 5 Bottom A         | yes     | 2.0              | 0                   | 63.6%             | 3.6%               |
| 5 Bottom B         | yes     | 5.0              | 0                   | 47.7%             | 5.1%               |
| 5 Bottom C         | yes     | 5.0              | 0                   | 52.7%             | 3.3%               |
| 5 Bottom D         | yes     | 3.3              | 0                   | 32.9%             | 9.2%               |
| 5 Bottom E         | yes     | 4.0              | 0                   | 30.6%             | 3.5%               |
| 5 Bottom F         | yes     | 2.0              | 0                   | 53.4%             | 2.5%               |

<sup>1</sup>A censor of 1 indicates that the days to break the filter paper was not determined (the experiment ended before the filter paper broke in half).

**Table A4.2.** Taxonomic identification of isolates.

| Enrichment Community | Taxonomic Classification                      | S_ab score |
|----------------------|-----------------------------------------------|------------|
| 1 Top A              | <i>Acidovorax caeni</i> R-24608               | 0.91       |
| 1 Top A              | <i>Asticcacaulis taihuensis</i> T3-B7         | 0.92       |
| 1 Top A              | <i>Klebsiella variicola</i> F2R9              | 0.99       |
| 1 Top A              | <i>Pseudomonas taiwanensis</i> BCRC 17751     | 0.97       |
| 1 Middle E           | <i>Acinetobacter calcoaceticus</i> NCCB 22016 | 0.97       |
| 1 Middle E           | <i>Enterobacter aerogenes</i> JCM1235         | 0.98       |
| 1 Middle E           | <i>Klebsiella oxytoca</i> ATCC 13182T         | 0.96       |
| 1 Middle E           | <i>Leclercia adecarboxylata</i> LMG 2803      | 0.93       |
| 1 Bottom E           | <i>Cedecea davisae</i> DSM 4568               | 0.96       |
| 1 Bottom E           | <i>Chryseobacterium indologenes</i> LMG 8337  | 0.87       |
| 1 Bottom E           | <i>Enterobacter amnigenus</i> JCM1237         | 0.92       |
| 1 Bottom E           | <i>Enterobacter asburiae</i> JCM6051          | 0.92       |
| 1 Bottom E           | <i>Klebsiella variicola</i> F2R9              | 0.98       |
| 1 Bottom E           | <i>Pseudomonas citronellolis</i> DSM 50332T   | 0.97       |
| 1 Bottom E           | <i>Pseudomonas delhiensis</i> RLD-1           | 0.96       |
| 1 Bottom E           | <i>Pseudomonas jinjuensis</i> Pss 26          | 0.98       |
| 3 Bottom C           | <i>Klebsiella variicola</i> F2R9              | 0.95       |
| 3 Bottom C           | <i>Pseudomonas nitroreducens</i> IAM 1439     | 0.90       |
| 3 Bottom C           | <i>Pseudomonas plecoglossicida</i> FPC951     | 0.93       |
| 3 Bottom C           | <i>Pseudomonas taiwanensis</i> BCRC 17751     | 0.93       |

**Dataset A4.1.** Taxonomic classification of reads and OTUs across samples. Available online at

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0151840#sec021>.

**Dataset A4.2.** Correlation analysis of degradation and OTUs with community diversity. Available online at

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0151840#sec021>.

**Appendix 5: Supplementary Material for Chapter 5.**

**In preparation as supplementary material for:** Gina R. Lewin, Nicole Davis, Bradon R.

McDonald, Adam J. Book, Marc G. Chevrette, Steven Suh, Ardina Boll, Cameron R. Currie.

Long-term cellulose enrichment selects for highly cellulolytic taxa and competition for public goods.

**Author contributions same as in Chapter 5.**

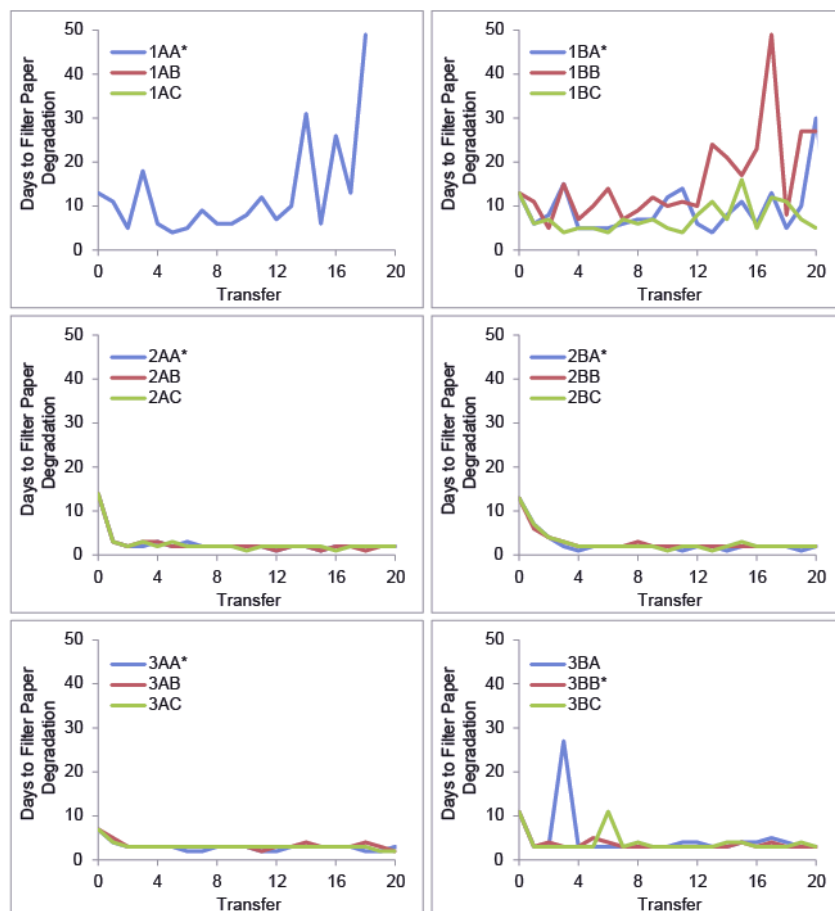


**Figure A5.1. Replicates of enrichment lines.** At the first transfer, each enrichment line was split into three replicate lines. Replicate lines were stopped after transfer 20 except for those marked with an asterisk, which were also chosen for sequencing analyses.

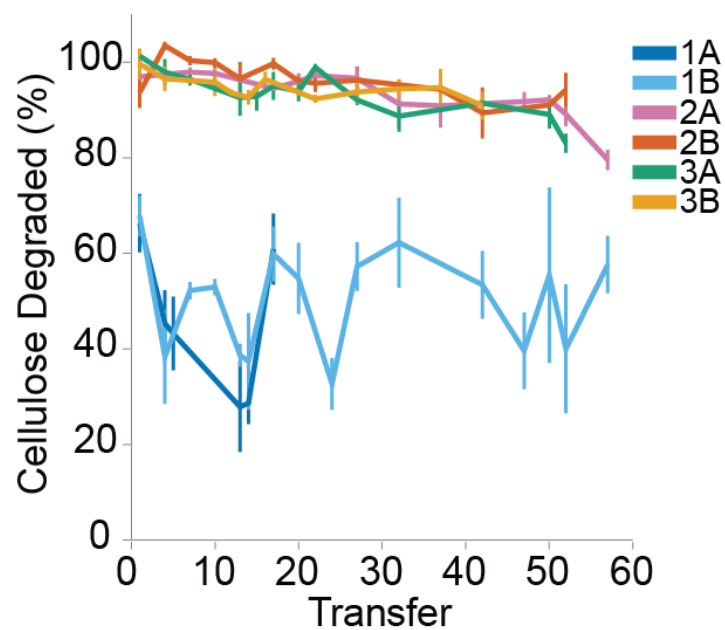


**Figure A5.2. Percentage of cellulose degradation in 10 days.** The amount of cellulose degraded in 10 days relative to controls using an acid detergent method with three biological replicates per sample.

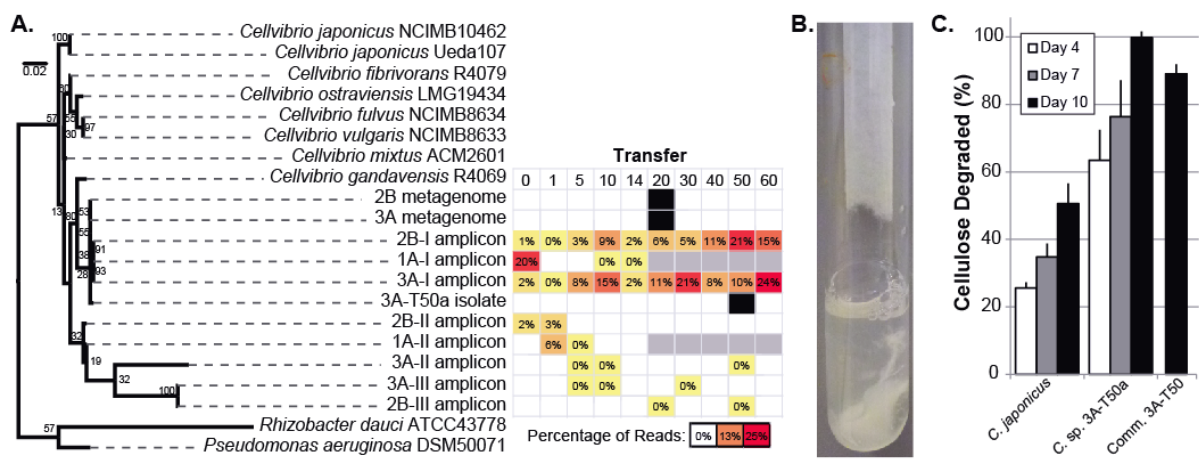


**Figure A5.3. Taxonomic and functional analysis of *Cellvibrio* isolates.** (A) Taxonomy of *Cellvibrio* type strains, sequences constructed from the 2B and 3A metagenome data using EMIRGE, amplicon sequences for OTU4 clustered at 100% identity, and the 16S rRNA gene sequence of the *Cellvibrio* isolate from community 3A at transfer 50. The abundance of each amplicon sequence in the corresponding sample is indicated in the heat map, and the corresponding transfer time point for the metagenome and isolate samples is indicated with black boxes. Taxonomy was built using RAxML with 100 bootstraps and *Rhizobacter* and *Pseudomonas* as outgroups. (B) Representative cellulose filter paper tube of *C. sp.* 3A-T50a after three days. (C) Quantitative cellulose filter paper degradation of *C. sp.* 3A-T50a isolate compared with *C. japonicas* and enriched community 3A at transfer 50.

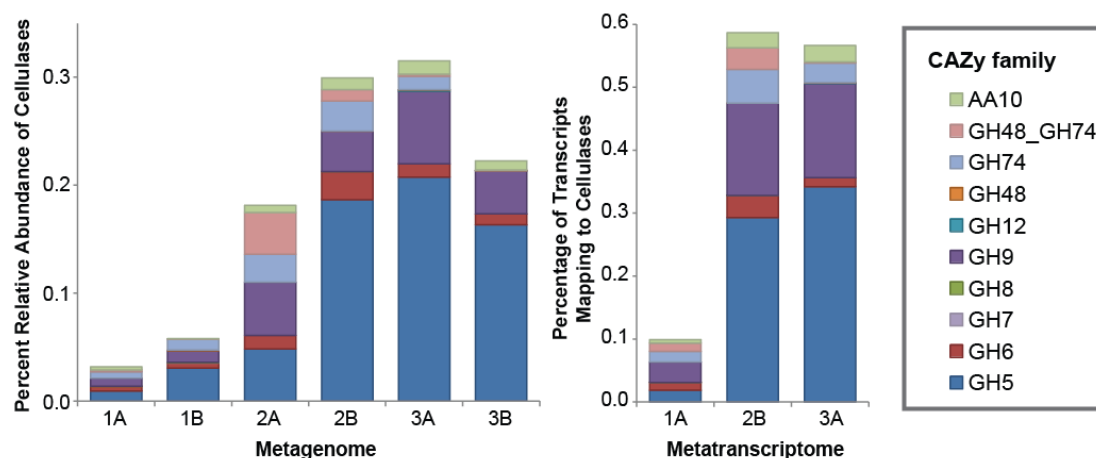


**Figure A5.4. Cellulase genes and transcript annotation and abundance.** The metagenome coverage or metatranscriptome read count of endocellulase, exocellulase, and lytic polysaccharide monooxygenase CAZy genes were summed and normalized to the total coverage of reads in each sample.



**Table A5.1.** Composite genome assembly quality and metatranscriptome mapping statistics

| <b>1A</b> |                    |                           |                            |                |           |          |                                 |                                  |
|-----------|--------------------|---------------------------|----------------------------|----------------|-----------|----------|---------------------------------|----------------------------------|
| Bin ID    | Lineage            | Completeness <sup>1</sup> | Contamination <sup>2</sup> | # of scaffolds | size (Mb) | n50 (bp) | % MT Reads Mapping <sup>3</sup> | Adjusted MT Mapping <sup>4</sup> |
| 1         | Sphingobacterium   | 100                       | 0.95                       | 67             | 5.4       | 157376   | 4%                              | 4%                               |
| 2         | Chitinophagaceae   | 99.01                     | 1.19                       | 148            | 4.9       | 59397    | 2%                              | 2%                               |
| 3         | Weeksella          | 98.39                     | 0                          | 32             | 3         | 131492   | 5%                              | 5%                               |
| 4         | Asticcacaulis      | 94.86                     | 1.71                       | 272            | 3.3       | 18879    | 4%                              | 4%                               |
| 5         | Actinobacteria     | 76.89                     | 3.18                       | 221            | 1.6       | 8580     | 5%                              | 7%                               |
| 6         | Pelagibacterium    | 65.19                     | 2.6                        | 440            | 2.6       | 6777     | 5%                              | 8%                               |
| 8         | Rhodanobacter      | 75.74                     | 1.83                       | 597            | 2.6       | 5251     | 17%                             | 22%                              |
| 9         | Cellulomonas       | 58.53                     | 3.26                       | 877            | 2         | 2491     | 7%                              | 12%                              |
|           |                    |                           |                            |                |           |          | 50%                             | 65%                              |
| <b>2B</b> |                    |                           |                            |                |           |          |                                 |                                  |
| Bin ID    | Lineage            | Completeness              | Contamination              | # of scaffolds | size (Mb) | n50 (bp) | % MT Reads Mapping              | Adjusted MT Mapping              |
| 1         | Flaobacterium      | 98.75                     | 0.35                       | 27             | 3.6       | 193554   | 25%                             | 25%                              |
| 2         | Cellvibrio         | 99.73                     | 0.24                       | 34             | 4.4       | 228391   | 37%                             | 37%                              |
| 3         | Chitinophagaceae   | 98.87                     | 0.02                       | 302            | 3.5       | 18749    | 3%                              | 3%                               |
| 4         | Cellulomonas       | 53.95                     | 0.87                       | 785            | 1.9       | 2637     | 20%                             | 37%                              |
| 7         | Sphingobacterium   | 90.1                      | 2.87                       | 954            | 6.2       | 9057     | 3%                              | 3%                               |
| 8         | Pseudomonas        | 68.73                     | 1.79                       | 1076           | 4.2       | 4365     | 3%                              | 4%                               |
| 10        | Unassigned         | 63.41                     | 2.02                       | 1086           | 3.1       | 3308     | 3%                              | 4%                               |
|           |                    |                           |                            |                |           |          | 93%                             | 113%                             |
| <b>3A</b> |                    |                           |                            |                |           |          |                                 |                                  |
| Bin ID    | Lineage            | Completeness              | Contamination              | # of scaffolds | size (Mb) | n50 (bp) | % MT Reads Mapping              | Adjusted MT Mapping              |
| 1         | Cellvibrio         | 94.69                     | 0.3                        | 205            | 4.03      | 38132    | 50%                             | 53%                              |
| 2         | Micavibrio         | 82.86                     | 1.74                       | 111            | 2.01      | 27116    | 2%                              | 3%                               |
| 3         | Flavobacterium     | 98.74                     | 0.44                       | 109            | 3.59      | 48700    | 6%                              | 6%                               |
| 4         | Chinophaga         | 97.85                     | 0.84                       | 124            | 3.23      | 41465    | 5%                              | 5%                               |
| 5         | Enterobacteriaceae | 60.52                     | 1.72                       | 1232           | 3.88      | 3722     | 2%                              | 3%                               |
| 6         | Azospirillum       | 57.39                     | 4.22                       | 1097           | 3.44      | 3496     | 3%                              | 6%                               |
| 7         | Opitutaceae        | 63.17                     | 3.79                       | 414            | 2.45      | 7308     | 6%                              | 9%                               |
| 8         | Pedobacter         | 63.41                     | 1.12                       | 695            | 2.08      | 3361     | 1%                              | 2%                               |
| 10        | Comamonadaceae     | 70.98                     | 1.8                        | 494            | 2.40      | 6016     | 2%                              | 3%                               |
| 11        | Betaproteobacteria | 60.07                     | 2.01                       | 769            | 1.73      | 2785     | 2%                              | 3%                               |
| 12        | Pseudomonas        | 63.57                     | 1.12                       | 450            | 3.20      | 7083     | 3%                              | 5%                               |
| 15        | TM7                | 61.68                     | 0.95                       | 22             | 0.76      | 68949    | 1%                              | 1%                               |
|           |                    |                           |                            |                |           |          | 83%                             | 98%                              |

<sup>1</sup>Completeness determined by CheckM based on taxonomy-specific conserved gene families<sup>2</sup>Contamination determined by CheckM based on multiple copies of conserved single-copy gene families<sup>3</sup>Percent of metatranscriptome reads mapping to composite genome relative to all trimmed reads<sup>4</sup>Adjusted for completeness of composite genome

## **Appendix 6: Aerobic deconstruction of cellulosic biomass by an insect-associated**

### ***Streptomyces*.**

**Previously published:** Taichi E. Takasuka, Adam J. Book, Gina R. Lewin, Cameron R. Currie, and Brian G. Fox (2013). Aerobic deconstruction of cellulosic biomass by an insect-associated *Streptomyces*. *Scientific Reports*, 3: 1030.

**Author contributions:** Conceived and designed the experiments: TET, AJB, BGF, CRC.

Performed the experiments: TET, AJB, GRL. Wrote the paper: TET, AJB, BGF, CRC.

### **A6.1 Abstract**

*Streptomyces* are best known for producing antimicrobial secondary metabolites, but they are also recognized for their contributions to biomass utilization. Despite their importance to carbon cycling in terrestrial ecosystems, our understanding of the cellulolytic ability of *Streptomyces* is currently limited to a few soil-isolates. Here, we demonstrate the biomass-deconstructing capability of *Streptomyces* sp. SirexAA-E (ActE), an aerobic bacterium associated with the invasive pine-boring woodwasp *Sirex noctilio*. When grown on plant biomass, ActE secretes a suite of enzymes including endo- and exo-cellulases, CBM33 polysaccharide-monooxygenases, and hemicellulases. Genome-wide transcriptomic and proteomic analyses, and biochemical assays have revealed the key enzymes used to deconstruct crystalline cellulose, other pure polysaccharides, and biomass. The mixture of enzymes obtained from growth on biomass has biomass-degrading activity comparable to a cellulolytic enzyme cocktail from the fungus *Trichoderma reesei*, and thus provides a compelling example of high cellulolytic capacity in an aerobic bacterium.

## A6.2 Introduction

Plant cell walls represent the largest reservoir of organic polymers in terrestrial ecosystems and are the most abundant source of renewable organic energy on Earth. Most of this energy is stored in the recalcitrant polysaccharide cellulose, which is difficult to hydrolyze because of the highly crystalline structure, and in hemicellulose, which presents challenges because of its structural diversity and complexity. In terrestrial ecosystems, cellulolytic microbes help drive carbon cycling through the deconstruction of biomass into simple sugars. The deconstruction is largely accomplished through the action of combinations of secreted glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and carbohydrate binding modules (CBMs) [1-4].

Actinobacteria in the genus *Streptomyces* are an ecologically important group, especially in soil environments, where they are considered to be vital players in the decomposition of cellulose and other biomass polymers [2,5-7]. *Streptomyces* are able to utilize a wide range of carbon sources, form spores when resources are depleted, and produce antimicrobial secondary metabolites to reduce competition [6,8]. Although a large number of *Streptomyces* species can grow on plant biomass, only a small percentage (14%) have been shown to efficiently degrade crystalline cellulose [9]. Furthermore, the secreted cellulolytic activities of only a few species have been biochemically characterized, and still fewer species have been examined to identify key biomass degrading enzymes [10,11]. *Streptomyces reticuli* is one of the best-studied cellulose- and chitin-degrading soil-dwelling *Streptomyces*; functional analyses of several important cellulases and other hydrolytic enzymes have been reported [9,12,13]. Furthermore, polysaccharide monooxygenase (PMO) activity with cellulose was identified using the CBM33

protein from *Streptomyces coelicolor* [14], which suggests *Streptomyces* may use both hydrolytic and oxidative enzymes to deconstruct biomass. With the tremendous amount of sequence data collected in the past few years, and despite the view that *Streptomyces* make important contributions to cellulose degradation in the soil, genome-wide analyses of cellulolytic *Streptomyces* have not been reported.

In addition to their putative roles in carbon cycling in the soil, *Streptomyces* may also potentiate biomass deconstruction in insects through symbiotic associations [15-18]. Recent work has identified cellulose degrading *Streptomyces* associated with the pine-boring woodwasp *Sirex noctilio*, including *Streptomyces* sp. SirexAA-E (ActE) [19]. *S. noctilio* is a highly destructive wood-feeding insect that is found throughout forests in Eurasia and North Africa and is spreading invasively in North America and elsewhere [20]. While the wasp itself does not produce cellulolytic enzymes, evidence supports the role of a symbiotic microbial community that secretes biomass-degrading enzymes to facilitate nutrient acquisition for developing larvae in the pine tree [21]. The white rot fungus, *Amylostereum areolatum*, is the best-described member of this community, and the success of *Sirex* infestations is thought to arise from the insect's association with this cellulolytic fungal mutualist. However, work with pure cultures has suggested that ActE and other *Sirex*-associated *Streptomyces* are more cellulolytic than *A. areolatum* [19].

Here, we report a systems-level analysis of the capacity of ActE for growth on pure polysaccharides and plant biomass, including genomic, transcriptomic and proteomic studies. We also isolated the secreted enzymes and characterized their reactions on a variety of biomass-

derived substrates, including a comparison with a commercial enzyme cocktail from *Trichoderma reesei*, an industrial standard for fungal biomass degradation<sup>4</sup>. This work provides insights into how highly cellulolytic *Streptomyces* may contribute to insect-associated symbiotic communities. Furthermore, the genome-wide analysis provides a basis for evaluating the general cellulolytic capabilities of other members of this important genus.

## A6.3 Results

### A6.3.1 Genomic characteristics of ActE

Table A6.1 compares the genomic characteristics of ActE with well-known soil-isolated *Streptomyces* that produce antibiotics and with two model cellulolytic bacteria, *Clostridium thermocellum* and *Cellvibrio japonicas* [3,22,23]. Putative biomass-degrading protein-coding sequences from ActE were identified by BLAST analysis of the finished genome to the Carbohydrate Active enZyme (CAZy) database. Among the 6357 predicted protein-coding genes, 167 have one or more domains assigned to CAZy families, including 119 GHs, 29 CEs, 6 PLs and 85 CBMs. ActE contains 45 different types of GH families, 4 PL families, 7 CE families, and 21 CBM families. The number of total CAZy domains and diversity of CAZy

**Table A6.1.** Comparison of genomic composition

|                                 | ActE    | <i>S. coelicolor</i> | <i>S. griseus</i> | <i>C. thermocellum</i> | <i>C. japonicus</i> |
|---------------------------------|---------|----------------------|-------------------|------------------------|---------------------|
| Genome size (nt)                | 7414440 | 8667507              | 8545929           | 3843301                | 4576573             |
| Proteome size                   | 6357    | 8153                 | 7136              | 3173                   | 3750                |
| Total CAZy Proteins             | 167     | 221                  | 132               | 103                    | 183                 |
| % CAZy Proteins <sup>a</sup>    | 2.6%    | 2.7%                 | 1.8%              | 3.2%                   | 4.9%                |
| Total GH <sup>b</sup>           | 119     | 154                  | 80                | 70                     | 124                 |
| Total PL <sup>c</sup>           | 6       | 11                   | 4                 | 6                      | 14                  |
| Total CE <sup>d</sup>           | 29      | 36                   | 23                | 20                     | 28                  |
| Total CBM <sup>e</sup>          | 85      | 98                   | 68                | 121                    | 134                 |
| antiSMASH clusters <sup>f</sup> | 22      | 24                   | 37                | 3                      | 4                   |
| Genes in clusters               | 620     | 718                  | 1139              | 89                     | 111                 |
| % antiSMASH                     | 9.8%    | 8.8%                 | 16.0%             | 2.8%                   | 3.0%                |

<sup>a</sup>Proteins classified as Carbohydrate Active Enzymes (CAZy).

<sup>b</sup>GH, glycoside hydrolase.

<sup>c</sup>PL, pectate lyase.

<sup>d</sup>CE, carbohydrate esterase.

<sup>e</sup>CBM, carbohydrate binding module.

<sup>f</sup>Putative antibiotic producing gene cluster.

families are comparable to other highly cellulolytic organisms.

Nearly all publically available *Streptomyces* genomes encode a relatively high percentage of genes for putative cellulolytic enzymes. Interestingly, ActE and the antibiotic producing *Streptomyces*, *S. griseus* and *S. coelicolor*, shown in Table A6.1 have similar numbers and compositions of CAZy families, but

substantially different genome sizes. However, these antibiotic-producing *Streptomyces* are not highly cellulolytic (Figure A6.1).

Relative to *S. griseus* and *S.*

*coelicolor*, the ActE genome

contains two unique CAZy families

but does not possess 16 CAZy families present in these species.

However, ActE contains more representatives in 13 CAZy families. Enrichment of certain

CAZy families is observed in other highly cellulolytic organisms. For example, *C. thermocellum* contains 16 genes in the GH9 family alone.

It is interesting to consider whether

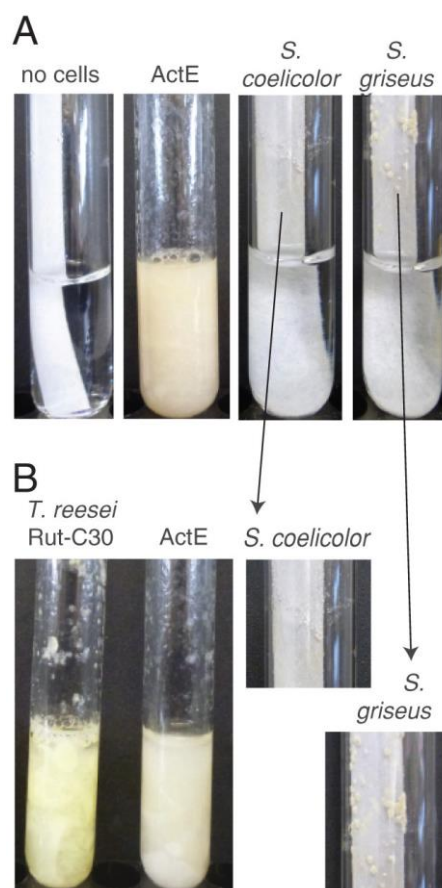


**Figure A6.1. Growth of ActE in minimal medium containing filter paper as the sole carbon source.** (A) Growth of ActE,

*Streptomyces coelicolor*, and *Streptomyces griseus* in minimal medium for 7 days at 30°C and pH 6.9. The expanded image shows small colonies of *S. coelicolor* and *S. griseus* forming on the surface of the paper. (B) Growth of ActE and *Trichoderma reesei* Rut-C30 for 7 days at 30°C and pH 6.0.

the reduction in total genome size and differences in CAZy composition between ActE and other closely related soil-dwelling *Streptomyces* might have arisen from evolutionary specialization of ActE, perhaps driven by association with the *Sirex*-fungal symbiosis.

### ***A6.3.2 Secreted proteins from ActE during growth on pure polysaccharides***

ActE grew well in minimal medium containing cellulose as the sole carbon source. The growth rate was similar to that of *Trichoderma reesei* Rut-C30, a model cellulolytic microbe (Figure A6.1), and both completely deconstructed filter paper strips in 5–7 days. By comparison, *S. coelicolor* and *S. griseus* grew only sparingly as small colonies on filter paper under the same conditions. This difference in growth capabilities among closely related *Streptomyces* prompted us to further examine ActE, which also grew with a filamentous morphology on polysaccharides including plant biomass (Supplementary Figure A6.1).

### ***A6.3.3 Reactions of the ActE secretomes***

The enzymatic activities of ActE secretomes were compared with a commercial secretome, Spezyme CP. This enzyme cocktail is prepared from *T. reesei* Rut-C30, and thus provides a useful, routinely available reference point for the capabilities of other cellulolytic organisms. HPLC analysis showed that the ActE cellulose secretome released cellobiose as the primary product during reaction with cellulose (Figure A6.2a, 95% of products), which is distinct from the higher proportion of glucose produced by the *T. reesei* secretome. Similarly, the primary products from xylan and mannan were xylobiose and mannobiose, respectively. Upon accounting for total glucose equivalents released, the ActE secretome obtained from growth on pure cellulose had specific activity that was about half of that provided by Spezyme CP (Figure

A6.2a, *inset*). Interestingly, the ActE secretome obtained from growth on pure cellulose had higher specific activity for deconstruction of pure mannan than Spezyme CP (Figure A6.2b). Additionally, the ActE secretome obtained from growth on pure xylan had higher specific activity for reaction with pure xylan than Spezyme CP. Cellulose, xylan, and mannan are all abundant in pinewood, thus accounting for the necessity of each of the major catalytic activities detected.

Anion exchange chromatography was performed to fractionate the ActE secretome obtained from cells grown on cellulose as the sole carbon source. We identified fractions that hydrolyzed pure polysaccharides by biochemical assays (Supplementary Figure A6.2), and confirmed the identity of the protein or proteins contained in these fractions by mass spectrometry (Supplementary Table A6.1). Where multiple polypeptides were present, the identity of each was confirmed by mass spectrometry to correspond to the indicated gene locus. In several cases, these most likely arise from proteolysis of a single protein found in the secretome. Fractions containing the maximum cellulase activity were highly enriched in SACTE\_0236 and SACTE\_0237, reducing and non-reducing end cellobiohydrolases from the GH6 and GH48 families, respectively. SACTE\_0265 and SACTE\_2347 were identified as the major proteins present in fractions associated with xylan and mannan hydrolysis, respectively. A CBM33 polysaccharide monooxygenase (SACTE\_3159) was also identified in the ion exchange profile. Moreover, beta-1,3 glucanase activity was identified in fractions that were enriched in SACTE\_4755.



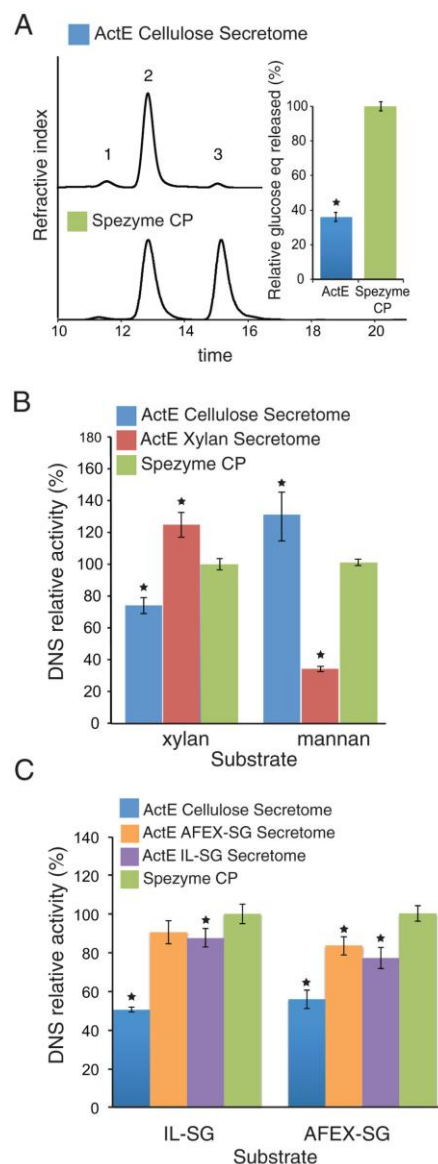


**Figure A6.2: Reactions of ActE secretomes and Spezyme CP.** (A) HPLC of sugars released from cellulose (1, cellotriose; 2, cellobiose; 3, glucose) and quantification of glucose equivalent (insert). (B) Reducing sugars released from xylan and mannan by the secretomes of ActE grown on cellulose and xylan. (C) Total reducing sugar released from ionic liquid-switchgrass (IL-SG) or AFEX-switchgrass (AFEX-SG) in reactions of the ActE cellulose, AFEX-SG, and IL-SG secretomes and Spezyme CP. Data represent the mean  $\pm$  s.d. from three experiments; \* indicates  $P < 0.01$  compared with Spezyme CP.

When ActE was grown on either ammonia fiber expansion-treated switchgrass (AFEX-SG) [24] or ionic liquid-treated switchgrass (IL-SG), the secretomes had ~2-fold increase in specific activity relative to the cellulose secretome and were equivalent to Spezyme CP for reaction with both the AFEX- and IL-treated biomass (Figure A6.2C) [24]. The ActE secretomes retained greater than 60% of maximal activity for the hydrolysis of AFEX- and IL-SG from 30 to 55°C and 35 to 47°C, respectively, which is comparable to recent reports on the temperature profile of

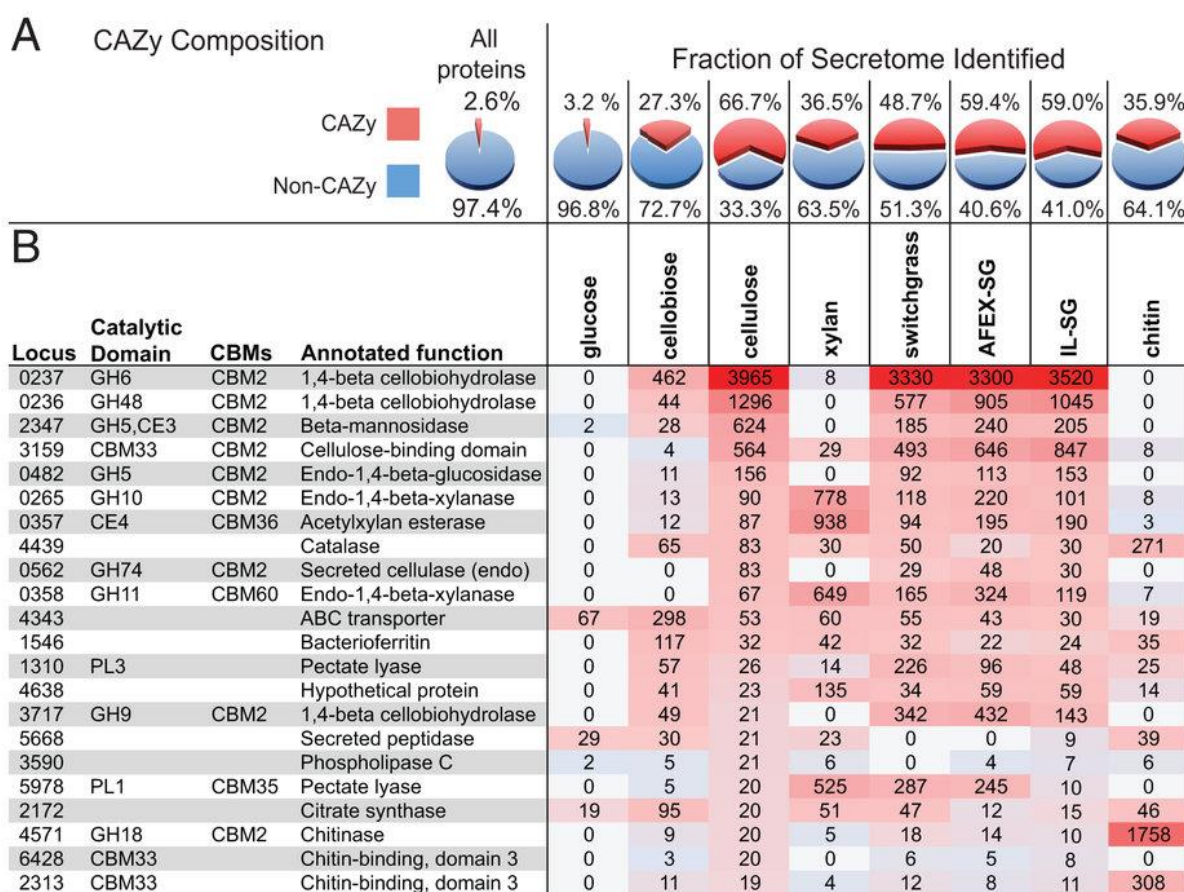


**Figure A6.3. Composition of ActE secretomes identified by LC-MS/MS.** (A) CAZy genes account for 2.6% of the 6357 predicted protein-coding sequences in the ActE genome. (B) Identity of most abundant proteins in the cellulose secretome proteins is sorted according to decreasing spectral counts (accounting for 95% of total spectral counts); corresponding spectral counts from other secretomes are also shown.

secretomes from thermophilic biomass-degrading fungi [25] (Supplementary Figure A6.3a). The secretomes showed a pH optimum of ~7 for reaction with AFEX-SG and a pH optimum of ~8 for reaction with IL-SG. Moreover, these secretomes retained greater than 60% of maximal activity over the ranges of pH 4.5 to 8.0, and pH 7.0 to 8.0, respectively (Supplementary Figure A6.3b). These optimal pH values are considerably higher than observed for Spezyme CP.

#### ***A6.3.4 Secretome analysis on pure polysaccharides and plant biomass***

To identify secreted proteins, supernatants from ActE cultures grown on glucose, cellobiose, cellulose, xylan, chitin, switchgrass, AFEX-SG, and IL-SG were analyzed by LC-MS/MS (Figure A6.3, Supplementary Table A6.2). The proteins were sorted into a descending rank according to spectral counts, and sets whose spectral counts summed to 95% of the total protein in each secretome are shown. Figure A6.3A summarizes the percentages of CAZy families in the detected proteins. The glucose secretome had a protein concentration of ~0.03 g/L of culture medium, and among the 136 proteins identified only 3% had a CAZy annotation. Indeed, the majority (>90%) likely originated from cell lysis. In contrast, the polysaccharide secretomes had a protein concentration of ~0.3 g/L of culture medium, a ~10-fold increase from the glucose secretome. Pectate lyase (SACTE\_1310), chondroitin/alginate lyase (SACTE\_4638), an extracellular solute binding protein (SACTE\_4343), bacterioferritin (SACTE\_1546), and catalase (SACTE\_4439) were observed in all polysaccharide secretomes. The first two proteins, SACTE\_1310 and SACTE\_4638, have signal peptides and are thus secreted as part of the response needed for growth on polysaccharides.

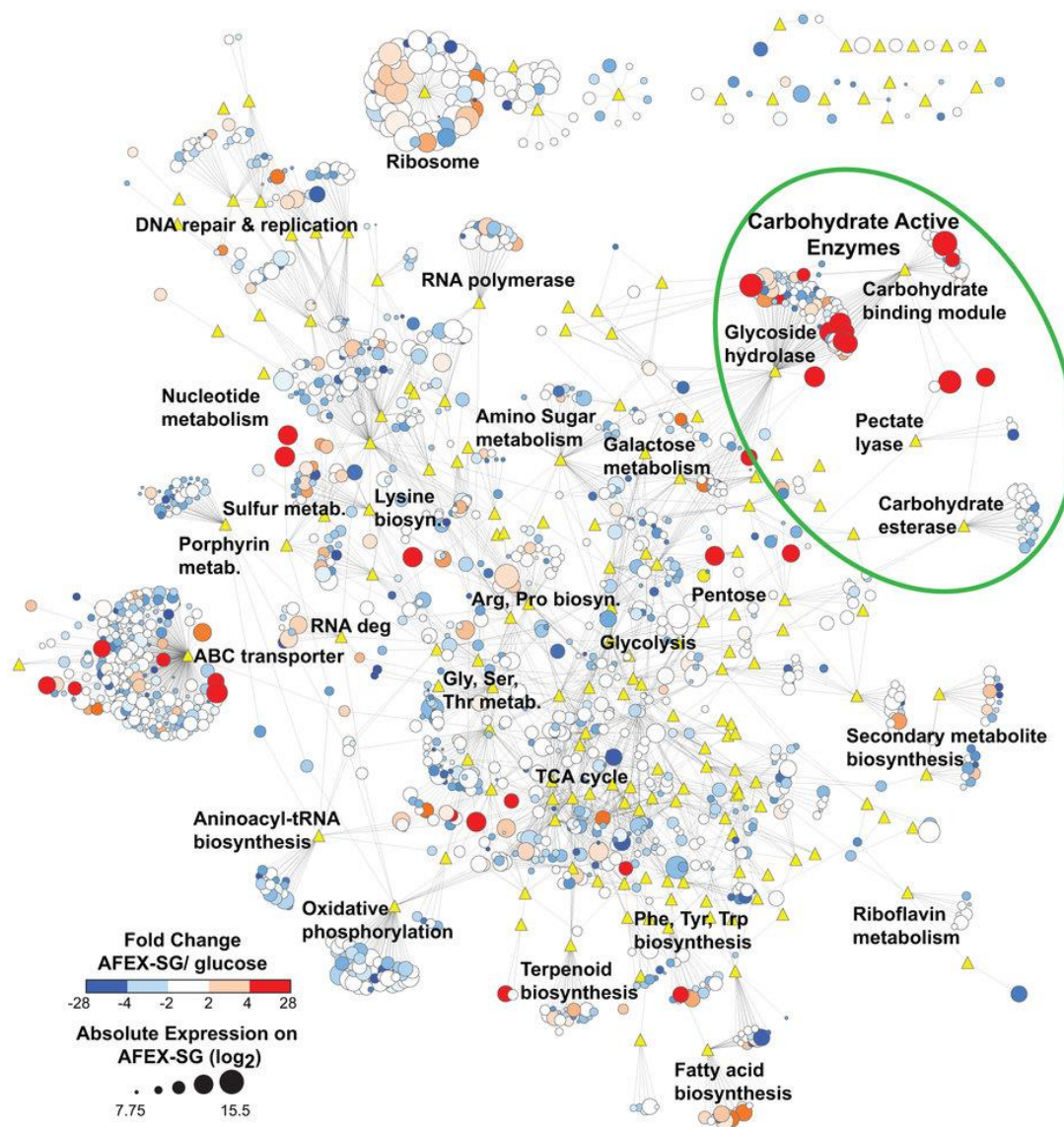


**Figure A6.4: Genome-wide changes in expression during growth of ActE on AFEX-treated switchgrass (AFEX-SG) versus glucose.** Nodes are genes (*circles*) or KEGG/CAZy functional categories (*yellow triangles*); edges indicate that the gene belongs to the indicated functional group as defined by either KEGG or CAZy analysis. Gene node sizes reflect expression intensity determined by microarray from growth on AFEX-SG as a  $\log_2$  ratio, where the genome-wide average transcriptional intensity was  $\sim 10.5$  for both substrates. Node colors represent expression changes as the  $\log_2$  ratio of AFEX-SG/glucose transcript intensities.



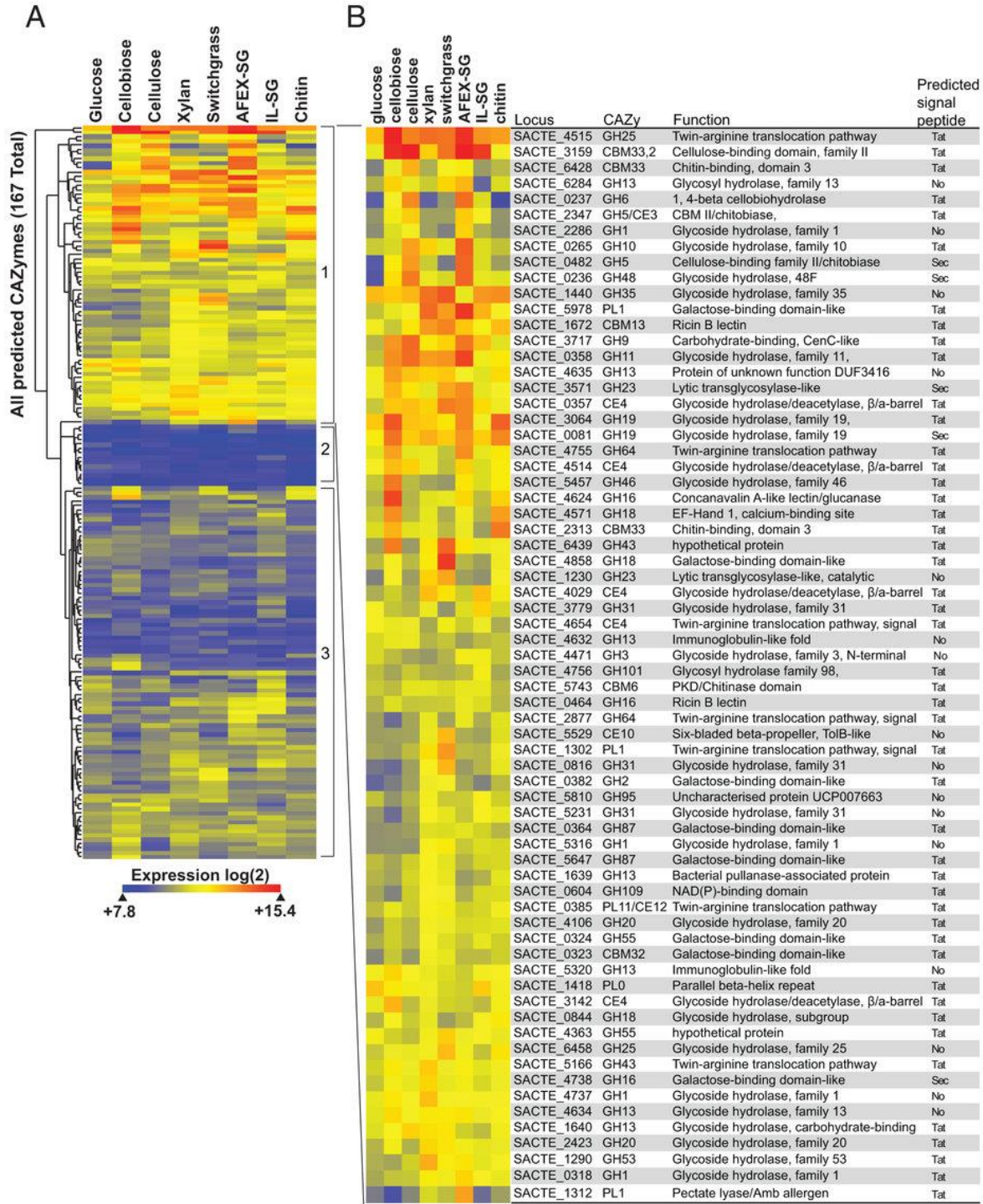


**Figure A6.5: Expression of ActE CAZy genes on various carbon sources.** (A) Hierarchical clustering of expression for 167 CAZy genes from the ActE genome during growth on the indicated substrates. (B) Identity of CAZy genes with distinct changes in expression observed in group 1 CAZy genes during growth in different carbon sources. Information for additional groups is provided in the Supplementary Information.

Figure A6.3 and Supplementary Table A6.2 demonstrate that 22 proteins accounted for 95% of the total spectral counts during growth on cellulose; two-thirds were from CAZy families. The five most abundant proteins, in order and representing ~85% of the total spectral counts, were reducing and non-reducing exoglucanases (SACTE\_0236 and SACTE\_0237), a CBM33 polysaccharide monooxygenase (SACTE\_3159), an endoglucanase (SACTE\_0482), and a  $\beta$ -mannosidase (SACTE\_2347). The first four proteins encode a non-redundant set of enzymes that likely provide the essential activities required for utilization of crystalline cellulose [22]. Among the 22 most abundant proteins, there were representatives from 9 different GH families, two CE families, two PL families, and two additional CBM33 proteins. Collectively, these secreted proteins represent ~20% of the CAZy composition in the ActE genome.

There were substantial differences in the composition of the xylan and chitin secretomes as compared to the cellulose secretome (Figure A6.3, Supplementary Table A6.2). In the xylan secretome, 92 proteins comprise 95% of the detected spectral counts. Twenty GHs from 18 different CAZy families were included, along with 1 CE4 and 2 PL family proteins. Thus, growth on xylan elicits secretion of representatives from half of the total CAZy families found in the ActE genome. The broad distribution of hemicellulytic enzymes in the xylan secretome contrasts with the considerably less diverse composition of the chitin secretome, which consists of 7 representatives from GH18 (e.g., chitinase, endo beta-N-acetylglucosaminidase), 2 from GH19 (e.g., chitinase, lysozyme), and 1 chitinolytic CBM33 (Supplementary Table A6.2). While chitinolytic CAZy families account for two-thirds of the proteins secreted during growth on chitin, they represent only ~6% of the diversity of CAZy families found in the genome. These

results document the substantially different substrate-specific responses of ActE during growth on different polysaccharides.

The secretomes isolated from cells grown on switchgrass, AFEX-SG, and IL-SG contained the highly abundant secreted proteins identified in the purified cellulose and xylan experiments and some additional proteins. These additional proteins likely reflect cellular response to the more complex composition of polysaccharides present in the biomass samples. The increased diversity of proteins present in the biomass secretome also increased the efficiency of reaction with plant biomass (Figure A6.2c). In total, the biomass secretomes contained 31 different CAZy families that contributed to the total spectral counts (~70% of the CAZy families present in the ActE genome), thus representing coordinated and extensive use of CAZyme families present in the ActE genome for biomass utilization.

#### ***A6.3.5 Gene expression analysis during growth on purified polysaccharides and plant biomass***

Gene expression profiles were determined for ActE grown on purified polysaccharides and plant biomass by whole genome microarrays (Figure A6.4 and Figure A6.5, Supplementary Figures A6.4 to A6.9). Genome-wide gene expression was analyzed as a functional annotation network composed of ActE genes (*circles*) connected to predicted functional groups (*triangles*; KEGG or CAZy). In Figure A6.4, the network was annotated with genome-wide microarray expression data to indicate genes that were differentially expressed when ActE was grown on either AFEX-SG or glucose, and further annotated to indicate normalized expression levels observed during growth on AFEX-SG. While many aspects of metabolism are modestly changed in response to these different carbon sources, the CAZy and ABC transporter categories were substantially

enriched in differentially expressed genes (Figure A6.4, *green circles*). Furthermore, pentose sugar metabolism, sulfur metabolism, and some amino acid biosynthesis pathways (e.g. aromatic amino acids) were also highly induced during growth on AFEX-SG relative to other carbon sources (Supplementary Figure A6.4 to A6.9). In contrast, ribosomal, secondary metabolite, and DNA repair genes showed little change in expression across the conditions examined. Within the CAZy functional group, there was a large induction of genes that contained both a GH domain and a CBM2 domain. Among the 11 genes in the ActE genome that contain a CBM2 domain, 6 were induced greater than 4-fold during growth on AFEX-SG. Furthermore, 9 of the 11 CBM2 containing proteins were identified in the secreted proteome (Figure A6.3).

Given the large number of differentially expressed CAZy genes identified in the network analysis, we analyzed the expression of this group of genes in cultures grown on different carbon sources (Figure A6.5). As with other cellulolytic organisms, there was strong correlation between the content of the secreted proteomes and the most highly expressed genes. Of the 167 ActE genes containing CAZy domains, 68 genes (Figure A6.5, group 1) showed distinct increases in expression when grown on different polymeric substrates, 14 genes (group 2, see Supplementary Figure A6.10) did not show any appreciable level of expression, and 85 genes (group 3, see Supplementary Figure A6.11) showed moderate changes in expression with the different substrates. A significant fraction of these genes contained translocation signals for either the Sec or twin-arginine translocation pathways, and genes encoding structural polypeptides for these translocation pathways were also highly expressed. Besides correlation with secreted proteins, the transcriptomic studies also gave insight into co-regulated gene clusters that potentially encode functional units for utilization of different polysaccharides by



ActE. In the following, the 130 genes with normalized expression intensities in the top 2% of all genes are described.

During growth on cellulose, four CAZy genes (SACTE\_0236, SACTE\_0237, SACTE\_3159, and SACTE\_0482) showed >15-fold increase in transcript abundance (Figure A6.5), and the corresponding proteins were highly enriched in the secreted proteome. None of these four were obviously placed in a gene cluster, and the two most highly expressed genes, SACTE\_0236 and SACTE\_0237, while adjacent on the chromosome, were transcribed in opposite directions. Nevertheless, these four most highly expressed genes and three others that showed >5-fold increase in transcript abundance (SACTE\_3717, SACTE\_6428, SACTE\_2347, Table A6.2) were associated with a conserved 14 bp palindromic promoter sequence, TGGGAGCGCTCCCA (the CebR binding element). CebR proteins are LacI/GalR-like transcriptional regulators shown to provide transcriptional control of gene expression in response to the presence of cellobiose or

**Table A6.2.** Analysis of upstream DNA sequence elements in ActE genes upregulated during growth on cellulose.

| Locus      | Catalytic domain | CBM  | Annotated function                            | Sequence <sup>a</sup> | Rank <sup>b</sup> | Fold change <sup>b</sup> |
|------------|------------------|------|-----------------------------------------------|-----------------------|-------------------|--------------------------|
| SACTE_0236 | GH48             | CBM2 | 1,4-beta cellobiohydrolase                    | TGGGAGCGCTCCCA        | 1                 | 21.7                     |
| SACTE_0237 | GH6              | CBM2 | 1,4-beta cellobiohydrolase                    | TGGGAGCGCTCCCA        | 2                 | 17.3                     |
| SACTE_3159 | CBM33            | CBM2 | Cellulose-binding domain                      | TGGGAGCGCTCCCA        | 3                 | 16.2                     |
| SACTE_0482 | GH5              | CBM2 | Endo-1,4-beta-glucosidase                     | TGGGAGCGCTCCCA        | 4                 | 15.4                     |
| SACTE_2288 |                  |      | Transport systems inner membrane component    | TGGGAGCGCTCCCA        | 5                 | 11.2                     |
| SACTE_3717 | GH9              | CBM2 | 1,4-beta cellobiohydrolase                    | TGGGAGCGCTCCCA        | 6                 | 9.7                      |
| SACTE_6428 | CBM33            |      | Chitin-binding, domain 3                      | GGGAGCGCTCCCA         | 9                 | 7.9                      |
| SACTE_2347 | GH5              | CBM2 | Beta-mannosidase                              | TGGGAGCGCTCCCA        | 11                | 5.0                      |
| SACTE_2287 |                  |      | Transport systems inner membrane component    | TGGGAGCGCTCCCA        | 15                | 4.3                      |
| SACTE_2289 |                  |      | Family 1 extracellular solute-binding protein | TGGGAGCGCTCCCA        | 19                | 3.9                      |
| SACTE_0352 |                  |      | GCN5-related N-acetyltransferase              | TGGGAGCGCTCCCA        | 22                | 3.6                      |
| SACTE_2286 | GH1              |      | Glycoside hydrolase 1                         | GGGAGCGCTCCCA         | 27                | 3.4                      |
| SACTE_0483 |                  | CBM2 | Cellulose-binding family protein              | GGGAGCGCTCCCA         | 503               | 1.6                      |
| SACTE_0562 | GH74             | CBM2 | Secreted cellulase (endo)                     | TGGGAGCGCTCCCA        | 5759              | 0.7                      |
| SACTE_2285 |                  |      | LacI family transcriptional regulator (CebR)  | TGGGAGCGCTCCCA        | 6229              | 0.6                      |

<sup>a</sup>Predicted binding sequence element found upstream from gene locus.

<sup>b</sup>Ranking and fold change in expression intensity detected by microarray for ActE genes when grown on cellulose relative to glucose.

other small oligosaccharides in *S. griseus*, *S. reticuli*, and *Thermobifida fusca* [26-28]. Likewise, the genes (SACTE\_2285 to SACTE\_2289) encoding a CebR regulator (SACTE\_2285), a GH1 protein ( $\beta$ -glucosidase), a two-protein cellobiose transporter system, and an extracellular solute binding protein were associated with a CebR binding element and were also among the most highly expressed genes during growth on cellulose. These latter five genes have 75% or greater sequence identity with the cellobiose utilization operon identified in *S. griseus* and *S. reticuli* [26,29]. There were only 15 genes annotated as hypothetical or domain of unknown function (12%) up-regulated during growth on cellulose, a considerably smaller percentage of these than in the entire genome (27%).

Several characteristics distinguished expression during growth on either xylan or chitin. First, unique sets of genes were induced, as there was only 14% and 10% overlap, respectively, when compared to cellulose. Second, ~33% of the top 2% of genes expressed during growth on either xylan or chitin were annotated as hypothetical or domain of unknown function, which greatly exceeds the unknown fraction in the cellulose secretome. During growth on xylan, two clusters of genes were up-regulated. One extended from SACTE\_0357 to SACTE\_0370, encoding proteins from the GH11, GH13, GH42, GH43, GH78, GH87, and CE4 families, a LacI-like transcriptional regulator, a secreted peptidase, and two sets of inner membrane transporters and associated solute binding proteins. Alternatively, during growth on chitin, three CBM33 proteins were up-regulated (SACTE\_0080, SACTE\_2313, SACTE\_6493), and two of these had an immediately adjacent gene encoding a GH18 (SACTE\_6494) or GH19 (SACTE\_0081) that was up-regulated.

When ActE was grown on biomass samples, 14 additional CAZy genes were uniquely up-regulated, and the corresponding proteins were identified in the proteomic analysis of biomass secretomes (Figures A6.3 and A6.4). A gene cluster extending from SACTE\_5858 to SACTE\_5864 was uniquely up regulated during growth on biomass. Among these genes, SACTE\_5860 and SACTE\_5862 are annotated as a twin-arginine translocation pathway protein and an ABC transporter, respectively, while the rest are annotated either as hypothetical protein or as domain of unknown function.

#### **A6.4 Discussion**

Symbiotic cellulolytic fungi and bacteria help a diverse array of eukaryotes access the energy stored in plant biomass and are fundamental to terrestrial energy flow and ecosystem function. However, our understanding of symbiotic cellulose degradation is limited to a few well-studied model organisms, most of which survive only in special anoxic digestive organs of the host. For example, ruminants carry communities of anaerobic microbes in their rumen that deconstruct biomass, facilitating nutrient acquisition from plant biomass for the host [30,31]. Insects also obtain nutrients from plant biomass by forming symbiotic relationships with cellulolytic microbes. Termites are a classic example, in which microbes deconstruct the plant biomass in the insect's hindgut [32]. Alternatively, leaf-cutter ants and other fungus-farming insects cultivate stable communities of externally associated aerobic microbes in specialized gardens [33,34]. The insects provide plant biomass to these fungal and bacterial communities, which in exchange help convert the plant material into usable nutrients for their host. Recently, aerobic *Streptomyces* were isolated from the microbial communities associated with the herbivorous woodwasp *Sirex noctilio* and they were shown to be efficient at degrading cellulose and plant

biomass [19]. While *Streptomyces* have long been considered important decomposers in the soil and may also be common cellulolytic symbionts associated with herbivorous insects, genome-wide studies characterizing the cellulolytic capacities of these ubiquitous bacterial species have been lacking. In this study, we have detailed the highly efficient biomass-degrading ability of an insect-associated *Streptomyces*, ActE, through a combination of biochemical, genomic, transcriptomic and proteomic analyses.

Biochemical assays of the ActE secretome showed comparable specific activity to a commercial enzyme cocktail prepared from the secretome of *T. reesei* Rut-C30, the major progenitor organism of biofuels enzymology. The secretomes of ActE grown on biomass contain enzymes with endocellulase, exocellulase, 1,3-beta-glucanase, mannanase, xylanase, esterase, pectate lyase, and other activities. The presence of endo- and exocellulases distinguishes ActE from *C. japonicus*, whose genome lacks one exocellulase that is usually associated with growth on cellulose [22]. The vigorous reactions of the ActE secretome with xylan and mannan are consistent with the abundance of arabinoxylan and glucomannan in pinewood. Furthermore, production of cellobiose, xylobiose, and mannobiose as the primary polysaccharide deconstruction products may preferentially channel energy to only select members of *Sirex/Amylostereum* community. Moreover, the high activity with 1,3-beta-glucan (callose) suggests a capability to attack the protective polymers produced by the pine tree during wound response and/or the cell wall of fungi present in the larval galleries.

During growth on biomass, ~75% of the CAZy families present in the ActE genome are expressed, and in many cases, these genes are part of gene clusters that may be suitable for

adaption to biotechnological processes. Interestingly, the dominant architecture for these enzymes is the fusion of a catalytic domain with a CBM2 domain. This is similar to the architecture used by *T. reesei*, whose enzymes also primarily consist of a single catalytic domain fused to a CBM1 domain.

GH61 and CBM33 proteins are now identified as copper-dependent polysaccharide monooxygenases<sup>35,36,37</sup> that function with extracellular oxidoreductases [38]. There are six CBM33 genes in the ActE genome, and this study provides insight into their function. Three (SACTE\_3159, SACTE\_6428, SACTE\_2313) were highly expressed and secreted during growth on cellulose, while three were observed only during growth on chitin (SACTE\_0080, SACTE\_2313, SACTE\_6493). SACTE\_3159 is unique among the six ActE CBM33 proteins, as it is fused to a CBM2 domain. It is also the most highly expressed CBM33 protein during growth on cellulose. Since genes encoding cellobiose oxidase and cellobiose dehydrogenase are not present in ActE, it is possible that an extracellular oxidoreductase is provided by other members of the *Sirex* community, in analogy to the combination of a *Thermoascus aurantiacus* GH61 and a *Humicola insolens* cellobiose dehydrogenase [38]. *A. areolatum*, a white rot basidiomycete from the *Sirex* community, is one reasonable source for these enzymes. The detection of a high level of catalase in the secreted proteome further suggests the overall environment in the insect-microbe community is oxidizing, with additional implications for how cellulose and lignin may be attacked.

The ActE genome contains 1738 genes (~30%) annotated as hypothetical or domain of unknown function. It is interesting that up to one-third of genes expressed during growth on xylan and

chitin were in these classes. This result implies a relative lack of understanding regarding proteins involved in utilization of these abundant polysaccharides by *Streptomyces*. Furthermore, several genes with unknown function were uniquely expressed during growth on biomass and provide a basis for future studies to improve understanding of natural biomass utilization.

Cellulolytic microorganisms have developed different enzymatic approaches to overcome the complexity of plant cell walls. Cellulolytic anaerobic bacteria such as *C. thermocellum* use a multi-protein complex called the cellulosome for biomass deconstruction [39,40], while other anaerobes use non-cellulosomic free enzymes [25] or cell-surface bound enzymes<sup>31</sup>. Aerobic cellulolytic fungi such as *Phanaerochaete* and *Postia* use a rich repertoire of both oxidative and hydrolytic enzymes to deconstruct biomass, and genome-enabled studies of their cellulolytic capacities are available [4,35-37,41]. This research has identified that an insect-associated *Streptomyces* isolated from an ecological niche defined by invasive attack on plant biomass has vigorous capabilities for attack of plant polysaccharides. Gene clusters that encode solute binding proteins, transporters, and extracellular enzymes were present and highly expressed during growth on the respective pure polysaccharide substrates, providing insights into the building blocks needed to install enhanced cellulolytic capacity into less reactive industrial strains of *Streptomyces*. Interestingly, the genes involved in extracellular aspects of cellulose utilization were not organized into extensive gene clusters but instead appear to be controlled by a CebR-like regulatory system distributed throughout the genome. In contrast xylan utilization appears to have a more organized structure also under control of a CebR-like regulatory system. Thus, genetic manipulation of CebR-like regulation in ActE may increase expression of hydrolytic and oxidative enzymes for biomass deconstruction.

Bacteria from the genus *Streptomyces* are best known for industrial production of antibiotics, and two well-known examples, *S. coelicolor* and *S. griseus*, have been intensively studied to better understand their capacity for producing antibiotics. Similar to these, ActE contains 22 clusters of genes potentially involved in secondary metabolite production (Table A6.1). However, none of the secondary metabolite genes found in ActE were expressed under the growth conditions investigated here (Figure A6.4, Supplementary Figures A6.4 to A6.9). While many *Streptomyces*, including *S. coelicolor*, *S. griseus* and others grow on plant polysaccharides<sup>6</sup>, they have considerable phenotypic differences observed during growth on pure crystalline cellulose compared to ActE (Figure A6.1).

In conclusion, we have shown that an aerobic *Streptomyces* bacterium associated with *Sirex* woodwasps secretes a mixture of enzymes with high specific activity for biomass deconstruction. Thus, ActE potentially makes profound contributions to nutrient acquisition in the *Sirex*-*Amyelosterum* mutualism. Given the associations of *Streptomyces* with other wood feeding insects [34], the present results suggest the contributions of externally-associated *Streptomyces* to insect-mediated biomass deconstruction are a widespread, important, and a heretofore underappreciated factor shaping carbon flow in terrestrial ecosystems. In contrast, to symbiotic cellulolytic bacteria that reside in the guts of animals (including insects such as termites), ActE is an externally propagating, aerobic bacterium that can secrete a potential mixture of hydrolytic and oxidative enzymes into the pinewood galleries created by its host. ActE thus provides a new example for how aerobic bacteria contribute to the deconstruction of plant biomass in Nature.

## A6.5 Methods

### A6.5.1 *Genome analysis*

The complete genome sequence of *Streptomyces* sp. SirexAA-E (ActE, taxonomy ID 862751) was determined by the Joint Genome Institute, project ID 4086644. Gene annotation models were predicted using Prodigal<sup>42</sup>, examined using Artemis<sup>43</sup>, and are available at NCBI with the following accession numbers, GenBank: CP002993.1; RefSeq: NC\_015953.1. Carbohydrate-active enzymes were annotated by comparison of all translated open-reading frames to the CAZy database<sup>2</sup>. We collected CAZy annotated genes from the CAZy database ([www.cazy.org](http://www.cazy.org)). We then used BLASTP to compare all ActE protein-coding sequences to the CAZy database and to the pfam database ([ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/little\\_endian/Pfam\\_LE.tar.gz](ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/little_endian/Pfam_LE.tar.gz)). These two annotations were then crosschecked, and proteins annotated by both databases were identified as our final CAZy annotation. Secreted proteins were identified by SignalP, TatP, and SecretomeP analyses. BLAST was used to identify sequence orthologs in other organisms. Secondary metabolite gene clusters were identified by AntiSmash analysis<sup>44</sup>. CebR boxes were identified by using BLAST comparison of the *S. griseus* CebR box sequence to the ActE genome<sup>26</sup>. Networks of expression and functional categories were visualized using Cytoscape<sup>45</sup>.

### A6.5.2 *Biomass substrates*

Switchgrass and ammonia fiber expansion (AFEX)-treated switchgrass were obtained from the Great Lakes Bioenergy Research Center. Extensively washed ionic liquid (IL)-treated switchgrass was the generous gift of Dr. Masood Hadi (Joint BioEnergy Institute).



### **A6.5.3 Growth of organisms**

ActE, *Streptomyces coelicolor*, *Streptomyces griseus* and *Trichoderma reesei* RUT-C30 were grown at pH 6.9 (pH 6.0 for *T. reesei*), where 1 L contains: 10.72 g K<sub>2</sub>HPO<sub>4</sub>; 5.24 g KH<sub>2</sub>PO<sub>4</sub>; 2 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 0.5 mL of an iron solution (1 mg of FeSO<sub>4</sub> dissolved in 1 mL in 0.01 M HCl); 1 mL of 1 M MgSO<sub>4</sub> solution; 1 mL of thiamine solution (1 mg/mL); and 5 mL of trace elements solution (SPV-4). The sole carbon source (0.5% w/v) in the medium was glucose, cellulose (either Whatman #1 filter paper or Sigmacell-20, Sigma/Aldrich, St. Louis, MO as indicated), xylan, chitin, switchgrass, AFEX-treated switchgrass or IL-treated switchgrass. Cultures were incubated for 7 d at 30°C with shaking.

### **A6.5.4 Preparation of secretomes**

Supernatants obtained from different culture media were prepared by centrifugation of the culture medium for 10 min at 3000 × *g*, which removed the remaining insoluble polysaccharide and adhered cells. The supernatant fraction was then passed through a 0.20 µm nylon membrane filter (Whatman) in order to remove any remaining cells. For enzymatic assays, the supernatants were concentrated using a 3 kDa cut off ultrafiltration membrane. The concentration of secretome protein was determined by Bradford assay.

### **A6.5.5 Enzyme activity measurements**

Reduced sugar assays were carried out by mixing secretome preparations with polysaccharide-containing substrates including cellulose (either Whatman #1 filter paper or Sigmacell-20 as indicated), xylan, chitin, mannan, switchgrass, AFEX pretreated switchgrass, or ionic-liquid

pretreated switchgrass<sup>24</sup>. After incubation in 0.1 M sodium phosphate, pH 6 at 40°C for 20 h, the reducing sugar content was detected by dinitrosalicylic acid assay<sup>46</sup> and calibrated by using glucose, xylose, or mannose as standards. Purified polysaccharide preparations had negligible background response in the absence of added enzymes. Cellobionic and gluconic acids were assayed by a coupled enzyme assay (K-GATE system, Megazyme, Bray Ireland). Spezyme CP was obtained from Genencor with batch number #4901522860. The distributions of soluble sugar oligomers obtained from secretome reactions were determined using a Shimadzu Liquid Chromatograph HPLC system (Shimadzu Scientific Instruments, Columbia, MD) equipped with a refractive index detector (RID-10A) and a Phenomenex Rezex RPM-monosaccharide column. The temperature was maintained at 85°C and Milli-Q water was used as the mobile phase at 0.6 mL min<sup>-1</sup> flow rate. Glucose, cellobiose, cellotriose, cellotetraose, cellopentaose, and cellohexaose (Sigma) were used as standards. The integrated areas of peaks were analyzed by EZ start 7.2 SP1 software (Shimadzu).

#### ***A6.5.6 Extracellular protein profiles***

Extracellular proteins from culture supernatants were precipitated with trichloroacetic acid (TCA), resuspended in denaturing sample buffer (SDS and 2-mercaptoethanol), and separated by SDS-PAGE in 4–20% gels. Protein bands of interest were excised from the gel, digested with trypsin, desalted with C18 pipette tips (Millipore, Billerica, MA) and identified by MALDI-TOF (MDS SCIEX 4800 MALDI TOF/TOF, Applied Biosystems, Foster City, CA). LC-MS/MS was performed at the University of Wisconsin Biotechnology Center. Samples were prepared by TCA precipitation of 100 ng of total secreted protein from 7-day old culture supernatants. Protein samples were digested with trypsin (sequencing grade trypsin, Promega, Madison, WI)

and were desalted using C18 pipette tips (Millipore, Billerica, MA). High-energy collision dissociation (HCD) MS analyses employing a capillary LC-MS/MS were performed on an electrospray ionization FT/ion-trap mass spectrometer (LTQ Orbitrap XL, Thermo Fisher Scientific, San Jose, CA). The MS and MS/MS spectra were searched against the spectra obtained from the ActE proteome using Scaffold (Scaffold\_3\_00\_06, Proteome Software, Portland, OR). Comparisons of secretome composition were carried out using Mathematica v.8.0.4.0 (Wolfram Research).

#### **A6.5.7 RNA microarrays**

ActE was grown in minimal medium plus the indicated substrate for 7 days. The cell pellet was separated from the culture medium by centrifugation for 10 min at  $3000 \times g$ . Microarray experiments were carried out as reported previously [23]. The total RNA was extracted from the cell pellet and purified. The University of Wisconsin Gene Expression Center carried out the syntheses of cDNA and array hybridizations. Four-plex arrays were constructed by Nimblegen and hybridized with 10  $\mu$ g of labeled cDNA. ArrayStar (v4.02, DNASTAR, Madison, WI) was used to quantify and visualize data. All analyses were based on three or more biological replicates per carbon source. Quantile normalization and robust multi-array averaging (RMA) were applied to the entire data set. Unless otherwise specified, expression levels are based on log<sub>2</sub> values and statistical analysis of the datasets were performed using the moderated *t*-test.

#### **A6.6 References**

1. Baldrian, P. & Valaskova, V. Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiol Rev* **32**, 501–521 (2008).

2. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**, D233–238 (2009).
3. Lynd, L. R., Weimer, P. J., van Zyl, W. H. & Pretorius, I. S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**, 506–577 (2002).
4. Schuster, A. & Schmoll, M. Biology and biotechnology of *Trichoderma*. *Appl Microbiol Biotechnol* **87**, 787–799 (2010).
5. Crawford, D. L. Lignocellulose decomposition by selected *Streptomyces* strains. *Appl Environ Microb* **35**, 1041–1045 (1978).
6. Goodfellow, M. & Williams, S. T. Ecology of actinomycetes. *Annu Rev Microbiol* **37**, 189–216 (1983).
7. McCarthy, A. J. & Williams, S. T. Actinomycetes as agents of biodegradation in the environment--a review. *Gene* **115**, 189–192 (1992).
8. Schlatter, D. *et al.* Resource amendments influence density and competitive phenotypes of *Streptomyces* in soil. *Microb Ecol* **57**, 413–420 (2009).
9. Wachinger, G., Bronnenmeier, K., Staudenbauer, W. L. & Schrempf, H. Identification of mycelium-associated cellulase from *Streptomyces reticuli*. *Appl Environ Microb* **55**, 2653–2657 (1989).
10. Ishaque, M. & Kluepfel, D. Cellulase complex of a mesophilic *Streptomyces* strain. *Can J Microbiol* **26**, 183–189 (1980).
11. Semedo, L. T. *et al.* *Streptomyces drozdowiczii* sp. nov., a novel cellulolytic streptomycete from soil in Brazil. *Int J Syst Evol Microbiol* **54**, 1323–1328 (2004).

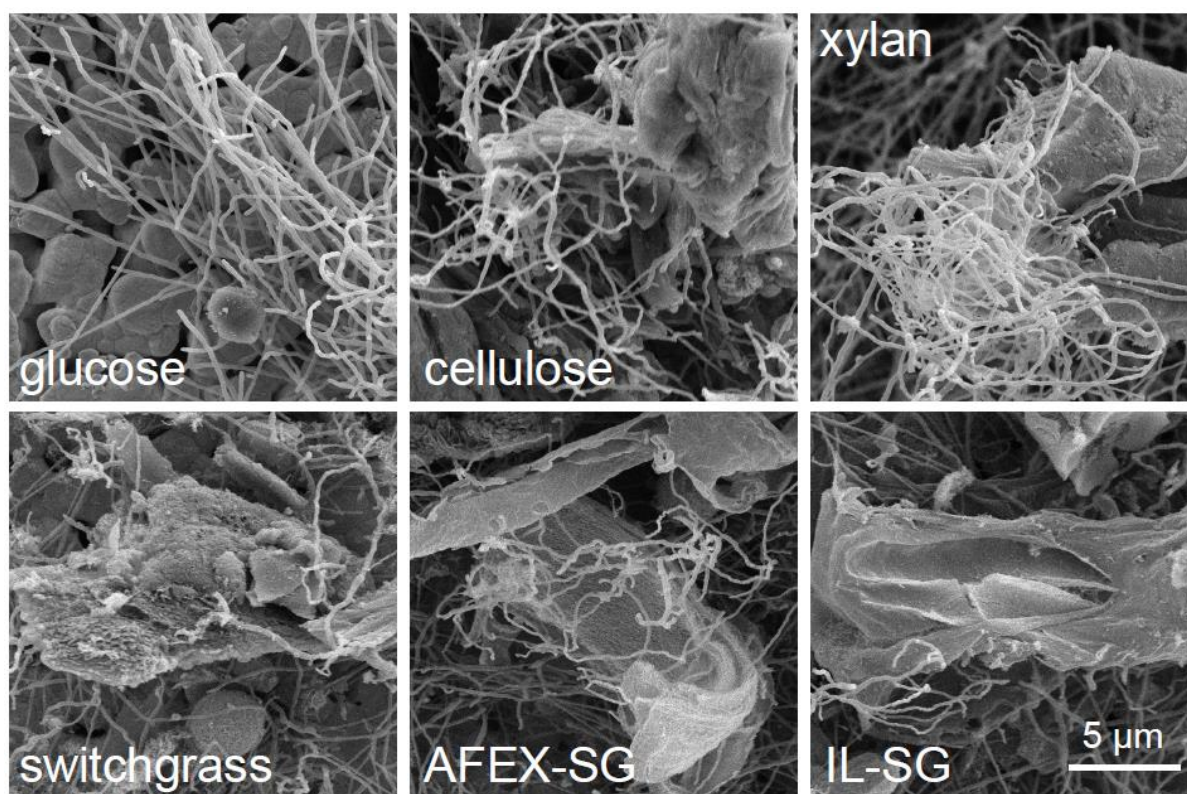
12. Schlochtermeyer, A., Walter, S., Schroder, J., Moorman, M. & Schrempf, H. The gene encoding the cellulase (Avicelase) Cel1 from *Streptomyces reticuli* and analysis of protein domains. *Mol Microbiol* **6**, 3611–3621 (1992).
13. Walter, S. & Schrempf, H. Physiological studies of cellulase (Avicelase) synthesis in *Streptomyces reticuli*. *Appl Environ Microb* **62**, 1065–1069 (1996).
14. Forsberg, Z. *et al.* Cleavage of cellulose by a CBM33 protein. *Protein Science* **20**, 1479–1483 (2011).
15. Bignell, D. E., Anderson, J. M. & Crosse, R. Isolation of facultatively aerobic Actinomycetes from the gut, parent soil and mound materials of the termites *Procubitermes-Aburiensis* and *Cubitermes-Severus*. *Fems Microbiol Ecol* **85**, 151–159 (1991).
16. Pasti, M. B. & Belli, M. L. Cellulolytic activity of actinomycetes isolated from termites (*Termitidae*) gut. *Fems Microbiol Lett* **26**, 107–112 (1985).
17. Pasti, M. B., Pometto, A. L., Nuti, M. P. & Crawford, D. L. Lignin-solubilizing ability of Actinomycetes isolated from termite (*Termitidae*) gut. *Appl Environ Microb* **56**, 2213–2218 (1990).
18. Schafer, A. *et al.* Hemicellulose-degrading bacteria and yeasts from the termite gut. *J Appl Bacteriol* **80**, 471–478 (1996).
19. Adams, A. S. *et al.* Cellulose-degrading bacteria associated with the invasive woodwasp *Sirex noctilio*. *ISME Journal* **5**, 1323–1331 (2011).
20. Bergeron, M. J. *et al.* Putative origin of clonal lineages of *Amylostereum areolatum*, the fungal symbiont associated with *Sirex noctilio*, retrieved from *Pinus sylvestris*, in eastern Canada. *Fungal Biol* **115**, 750–758 (2011).

21. Kukor, J. J. & Martin, M. M. Acquisition of digestive enzymes by siricid woodwasps from their fungal symbiont. *Science* **220**, 1161–1163 (1983).
22. Deboy, R. T. *et al.* Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J Bacteriol* **190**, 5455–5463 (2008).
23. Riederer, A. *et al.* Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl Environ Microb* **77**, 1243–1253 (2011).
24. Li, C. *et al.* Influence of physico-chemical changes on enzymatic digestibility of ionic liquid and AFEX pretreated corn stover. *Bioresour Technol* **102**, 6928–6936 (2011).
25. Tolonen, A. C. *et al.* Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol Syst Biol* **6**, 461 (2011).
26. Marushima, K., Ohnishi, Y. & Horinouchi, S. CebR as a master regulator for cellulose/celooligosaccharide catabolism affects morphological development in *Streptomyces griseus*. *J Bacteriol* **191**, 5930–5940 (2009).
27. Walter, S. & Schrempf, H. The synthesis of the *Streptomyces reticuli* cellulase (avicelase) is regulated by both activation and repression mechanisms. *Mol Gen Genet* **251**, 186–195 (1996).
28. Deng, Y. & Fong, S. S. Development and application of a PCR-targeted gene disruption method for studying CelR function in *Thermobifida fusca*. *Appl Environ Microb* **76**, 2098–2106 (2010).
29. Schlosser, A. & Schrempf, H. A lipid-anchored binding protein is a component of an ATP-dependent cellobiose/cellotriose-transport system from the cellulose degrader *Streptomyces reticuli*. *Eur J Biochem* **242**, 332–338 (1996).

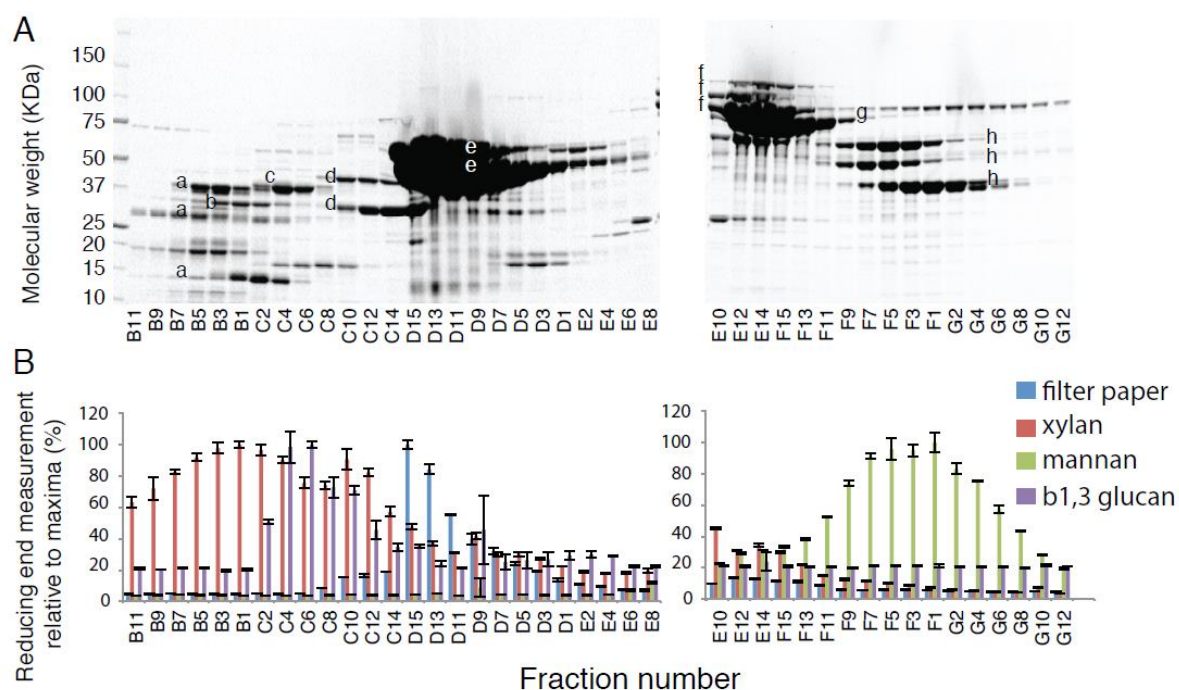
30. Hess, M. *et al.* Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**, 463–467 (2011).
31. Suen, G. *et al.* The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS One* **6**, e18814 (2011).
32. Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565 (2007).
33. Suen, G. *et al.* An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet* **6**, e1001129 (2010).
34. Scott, J. J. *et al.* Bacterial protection of beetle-fungus mutualism. *Science* **322**, 63 (2008).
35. Beeson, W. T., Phillips, C. M., Cate, J. H. D. & Marletta, M. A. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J Am Chem Soc* **134**, 890–892 (2012).
36. Quinlan, R. J. *et al.* Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *P Natl Acad Sci USA* **108**, 15079–15084 (2011).
37. Vaaje-Kolstad, G. *et al.* An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**, 219–222 (2010).
38. Langston, J. A. *et al.* Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl Environ Microb* **77**, 7007–7015 (2011).
39. Bayer, E. A., Lamed, R., White, B. A. & Flint, H. J. From cellulosomes to cellulosomes. *Chem Rec* **8**, 364–377 (2008).

40. Tamaru, Y., Miyake, H., Kuroda, K., Ueda, M. & Doi, R. H. Comparative genomics of the mesophilic cellulosome-producing *Clostridium cellulovorans* and its application to biofuel production via consolidated bioprocessing. *Environ Technol* **31**, 889–903 (2010).
41. Vanden Wymelenberg, A. *et al.* Comparative transcriptome and secretome analysis of wood decay fungi *Postia placenta* and *Phanerochaete chrysosporium*. *Appl Environ Microb* **76**, 3599–3610 (2010).
42. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
43. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
44. Medema, M. H. *et al.* antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* **39**, W339–346 (2011).
45. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).
46. Miller, G. L. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31**, 426–428 (1959).

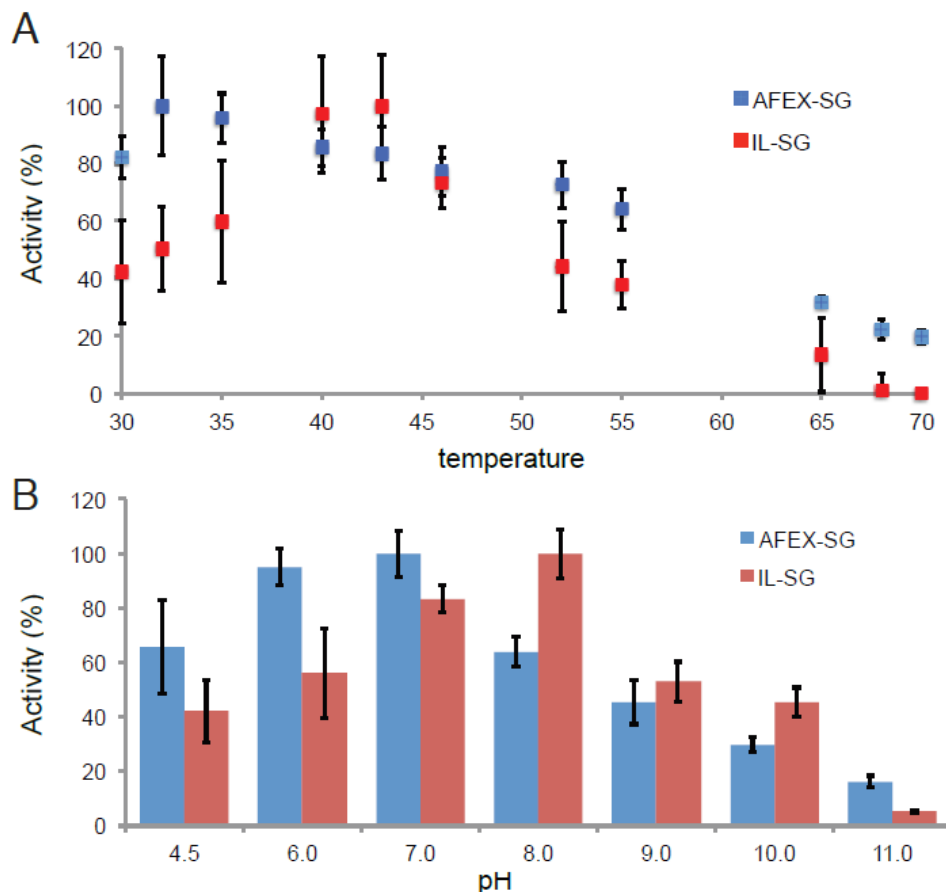


**A6.7 Supplementary Material**

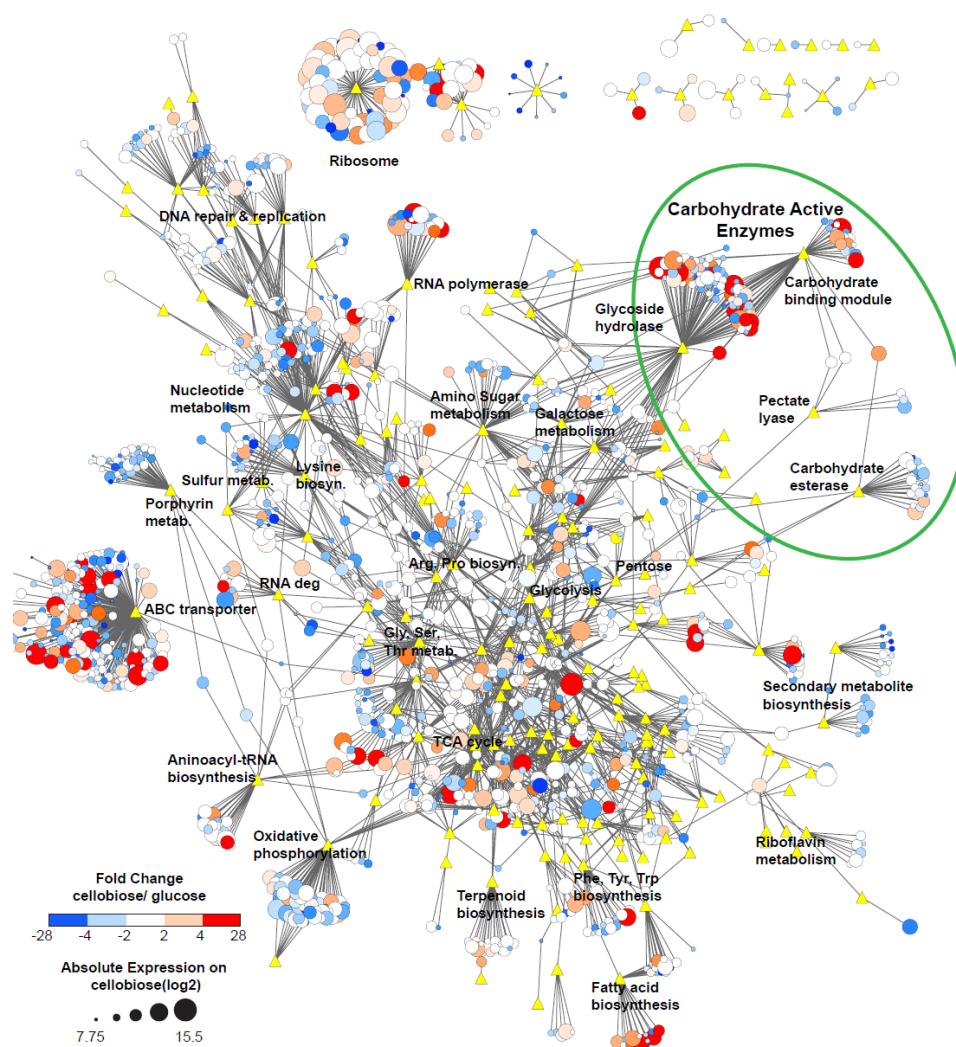
**Supplementary Figure A6.1.** Scanning electron microscopy images of ActE grown on different carbon sources including glucose, cellulose, xylan, switchgrass, ammonia ammonia fiber expansion-treated switchgrass (AFEX-SG) and ionic liquid-treated switchgrass (IL-SG). ActE cells were grown in minimum medium with the indicated substrate as a sole carbon source for 7 days at 30 °C. The scale bar indicates 5 μm.



**Supplementary Figure A6.2. Fractionation of the ActE cellulose secretome and assays of reactions with different polysaccharides.** (A) Anion exchange chromatography was performed using the ActE cellulose secretome, and fractions were collected and analyzed by SDS-PAGE. Lowercase letters indicate protein identified by MALDI-TOF MS shown in Supplementary Table S1. (B) Results from hydrolysis assays for reaction with filter paper (FP), xylan, mannan and beta-1,3 glucan as detected by DNS assay of each fraction. The percentage reactivity relative to the maximum activity observed for each substrate is shown. Error bars indicate the standard deviation, with  $n = 3$  for technical replicates.

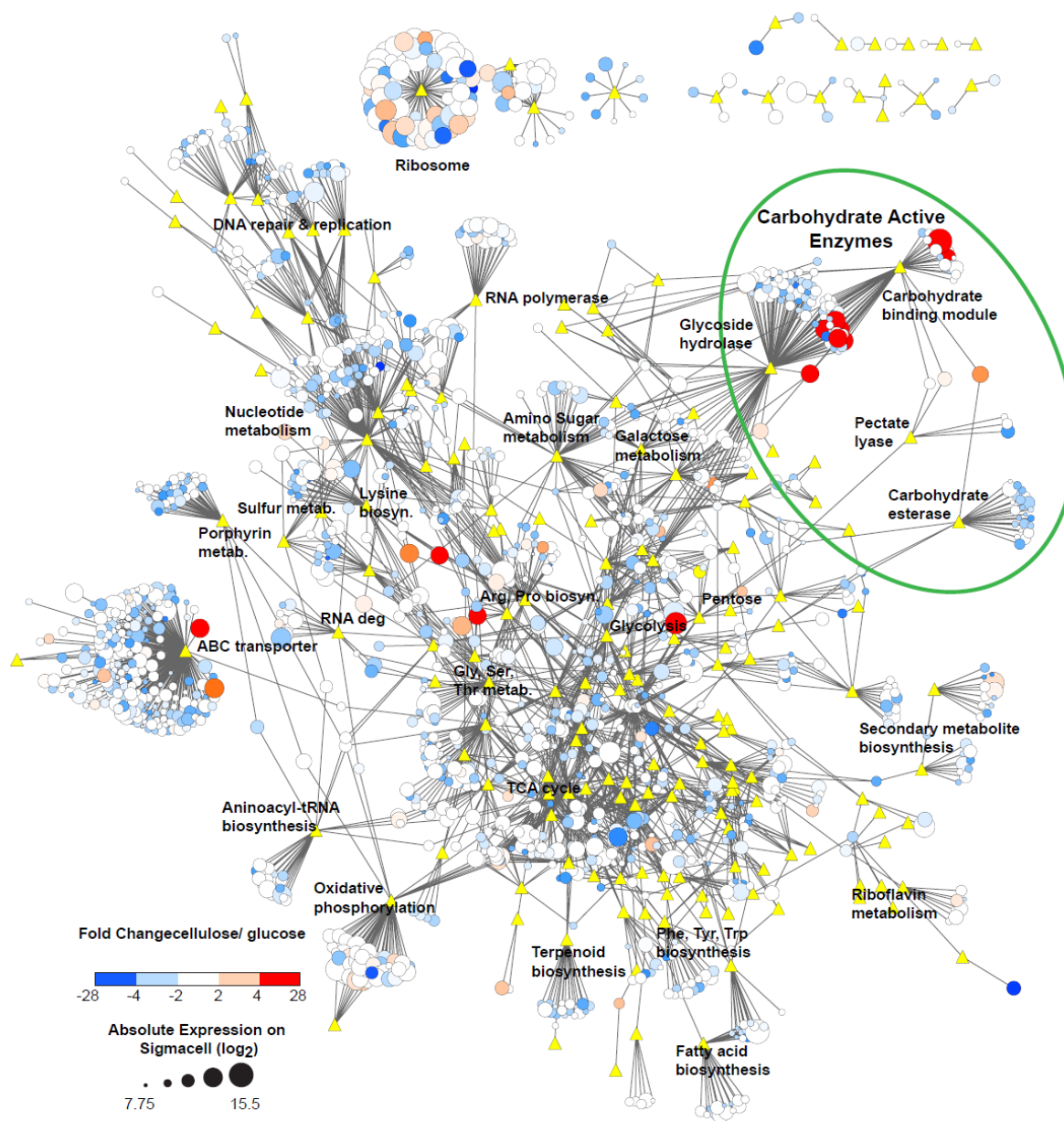


**Supplementary Figure A6.3. Temperature and pH profiles of the ActE secretome obtained from growth on AFEX-treated corn stover.** (A) The effect of temperature on the deconstruction of AFEX-treated switchgrass (AFEX-SG) and ionic liquid-treated switchgrass (IL-SG). The relative activity of the ActE secretome was compared to the maximal rates determined for reaction with AFEX-SG (blue star), and IL-SG (red star) at pH 6.0. (B) The effect of pH on the AFEX-SG and IL-SG deconstruction activities in the indicated ActE secretomes. The maximal rates observed for AFEX-SG and IL-SG were at pH 7.0 (blue star) and pH 8 (red star), respectively. Reactions were carried out at 40 °C and the 0.1 M buffers used were citrate (pH 4.5), phosphate (pH 6-8), CHES (pH 9-10), and CAPS (pH 11). The reaction was performed for 20 h and the reducing sugar content was measured by DNS assay.

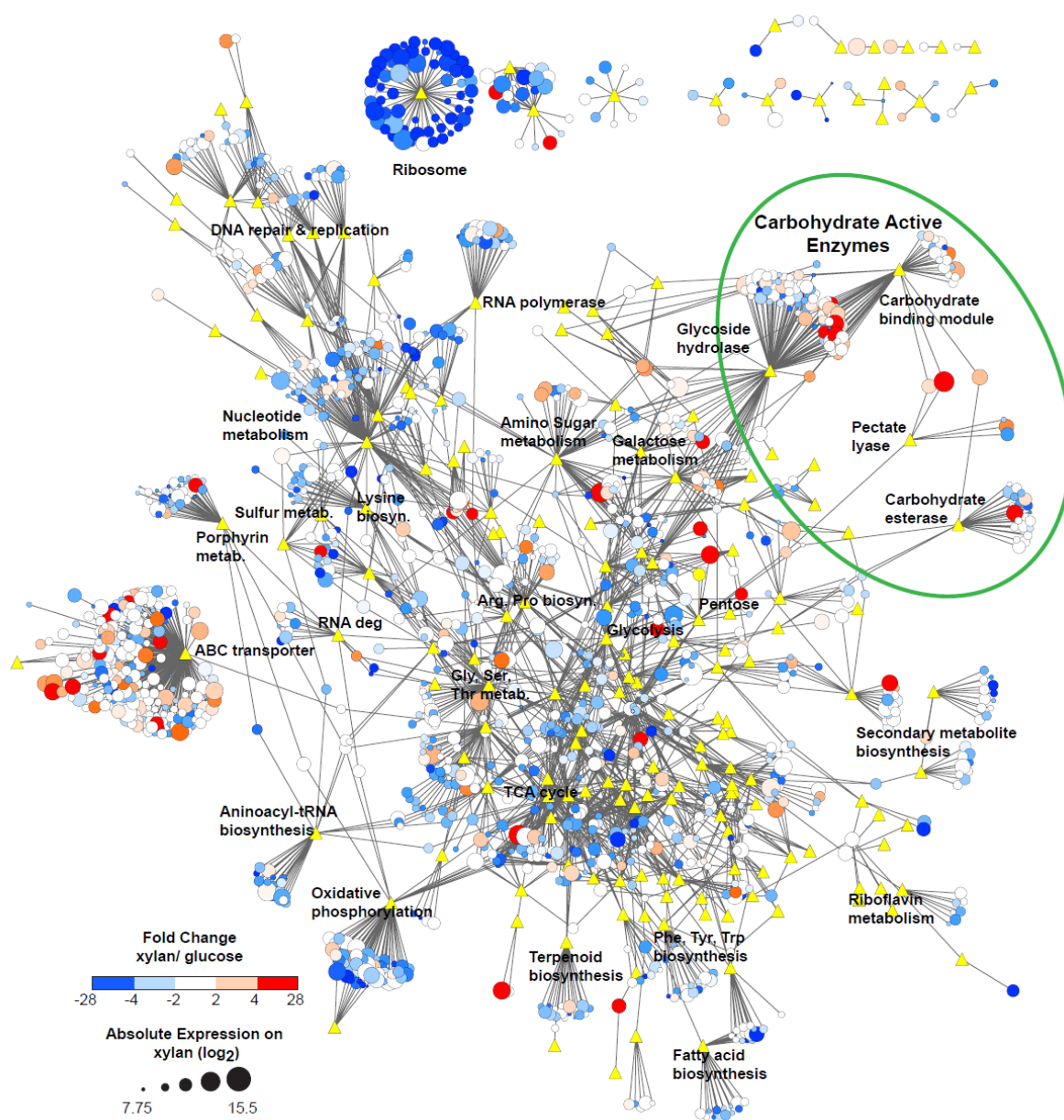


**Supplementary Figure A6.4. Genome-wide changes in expression during growth of ActE on substrate cellobiose versus glucose visualized as a Cytoscape interaction network.** Nodes are genes (circles) or KEGG/CAZy functional categories (yellow triangles); edges indicate that the gene belongs to the indicated functional group as defined by either KEGG or CAZy analysis. Gene node sizes reflect expression intensity determined by microarray from growth on substrate as a log<sub>2</sub> ratio. Node colors represent expression changes as the log<sub>2</sub> ratio of substrate/glucose transcript intensities, where the genome-wide average transcriptional intensity was ~10.5 for both substrate and glucose. Transcripts with less than two-fold changes in expression intensity are colored white; transcripts with greater than two-fold increase in expression intensity during growth on substrate are shown as a red gradient; transcripts with greater than two-fold increase in expression intensity during growth on glucose are shown as a blue gradient.

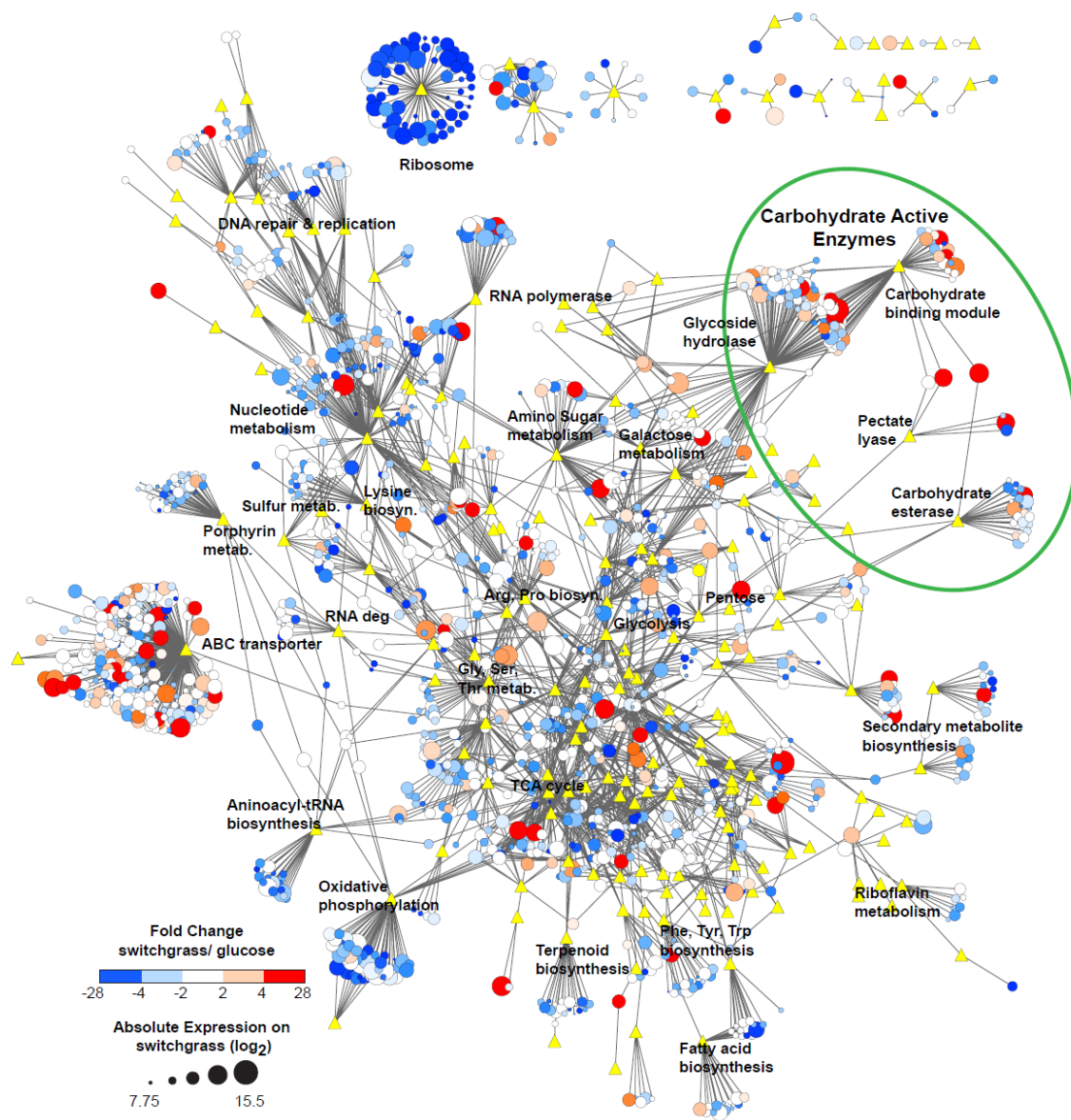




**Supplementary Figure A6.5. Genome-wide expression changes for growth on the substrate cellulose versus glucose visualized as a Cytoscape interaction network.** Other information provided in the caption to Supplementary Figure A6.4.

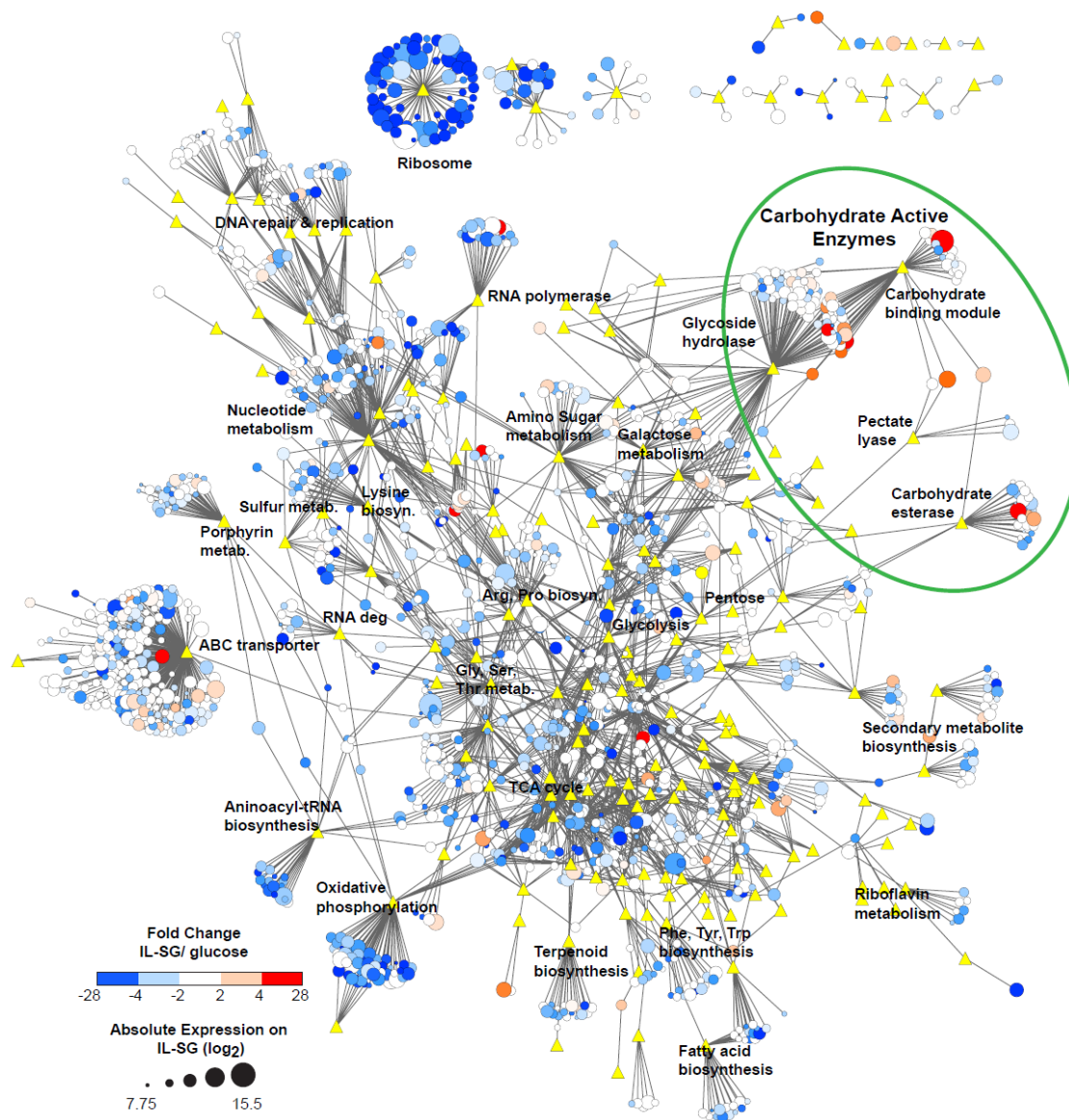


**Supplementary Figure A6.6. Genome-wide expression changes for growth on the substrate xylan versus glucose visualized as a Cytoscape interaction network.** Other information provided in the caption to Supplementary Figure A6.4.



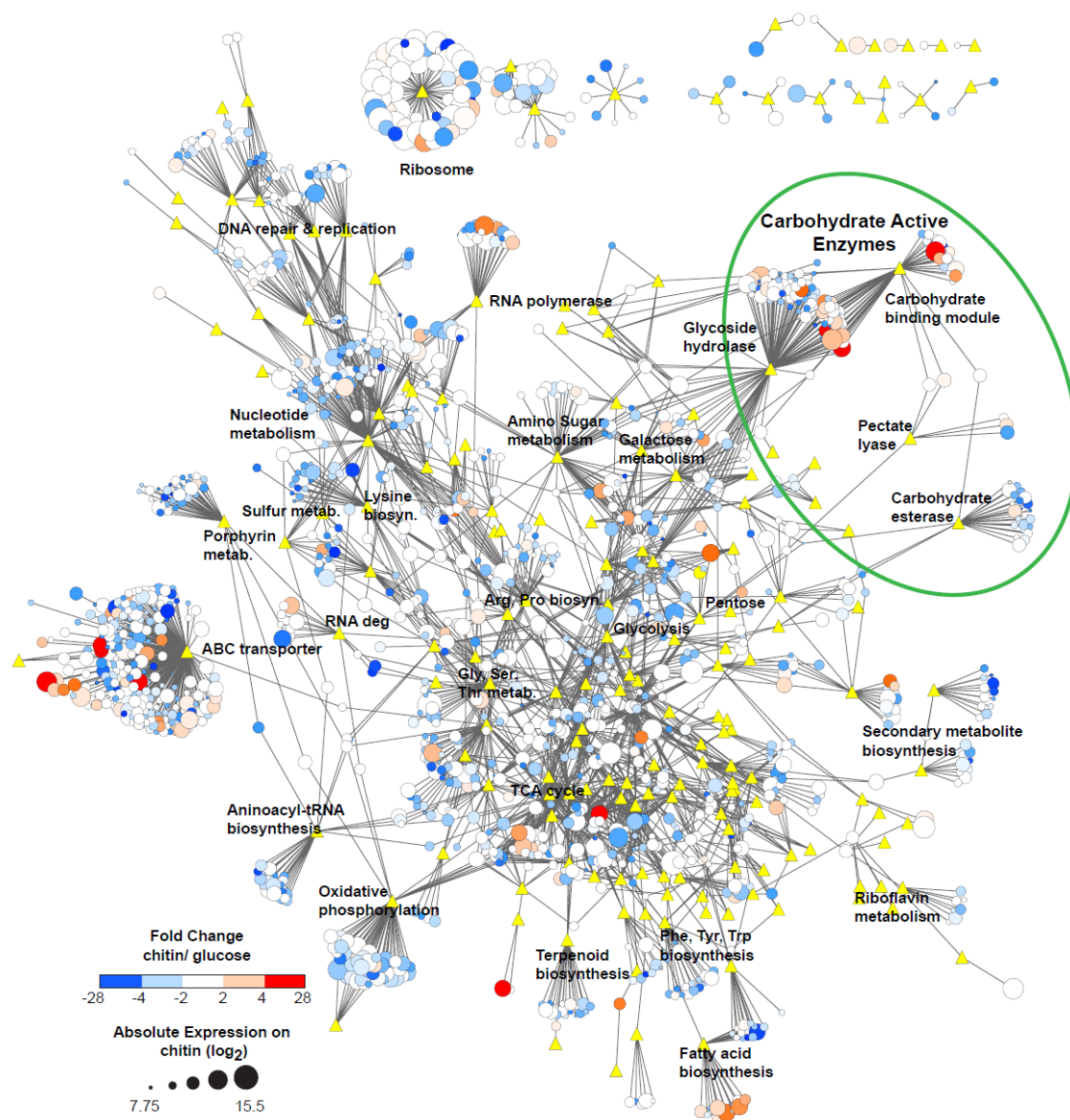
**Supplementary Figure A6.7. Genome-wide expression changes for growth on the substrate switchgrass versus glucose visualized as a Cytoscape interaction network.** Other information provided in the caption to Supplementary Figure A6.4.



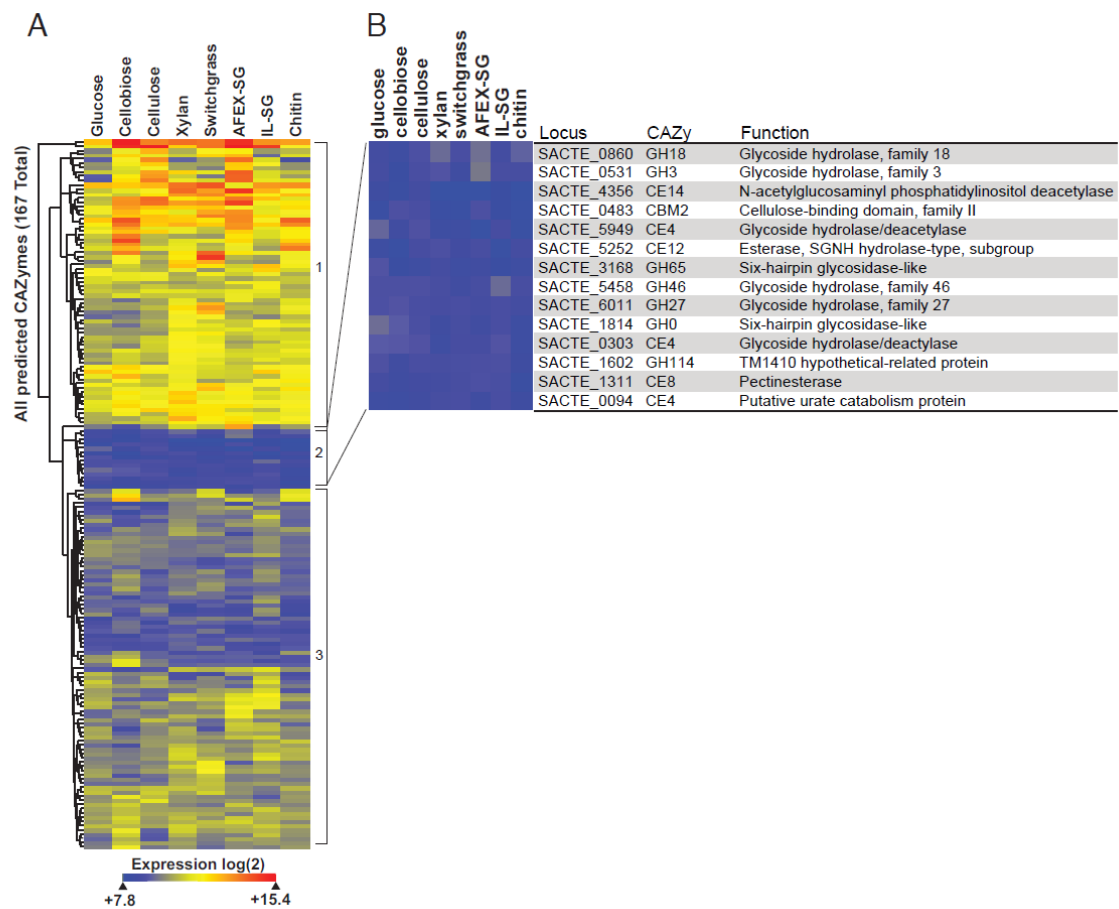


**Supplementary Figure A6.8. Genome-wide expression changes for growth on the substrate IL-treated switchgrass versus glucose visualized as a Cytoscape interaction network.** Other information provided in the caption to Supplementary Figure A6.4.

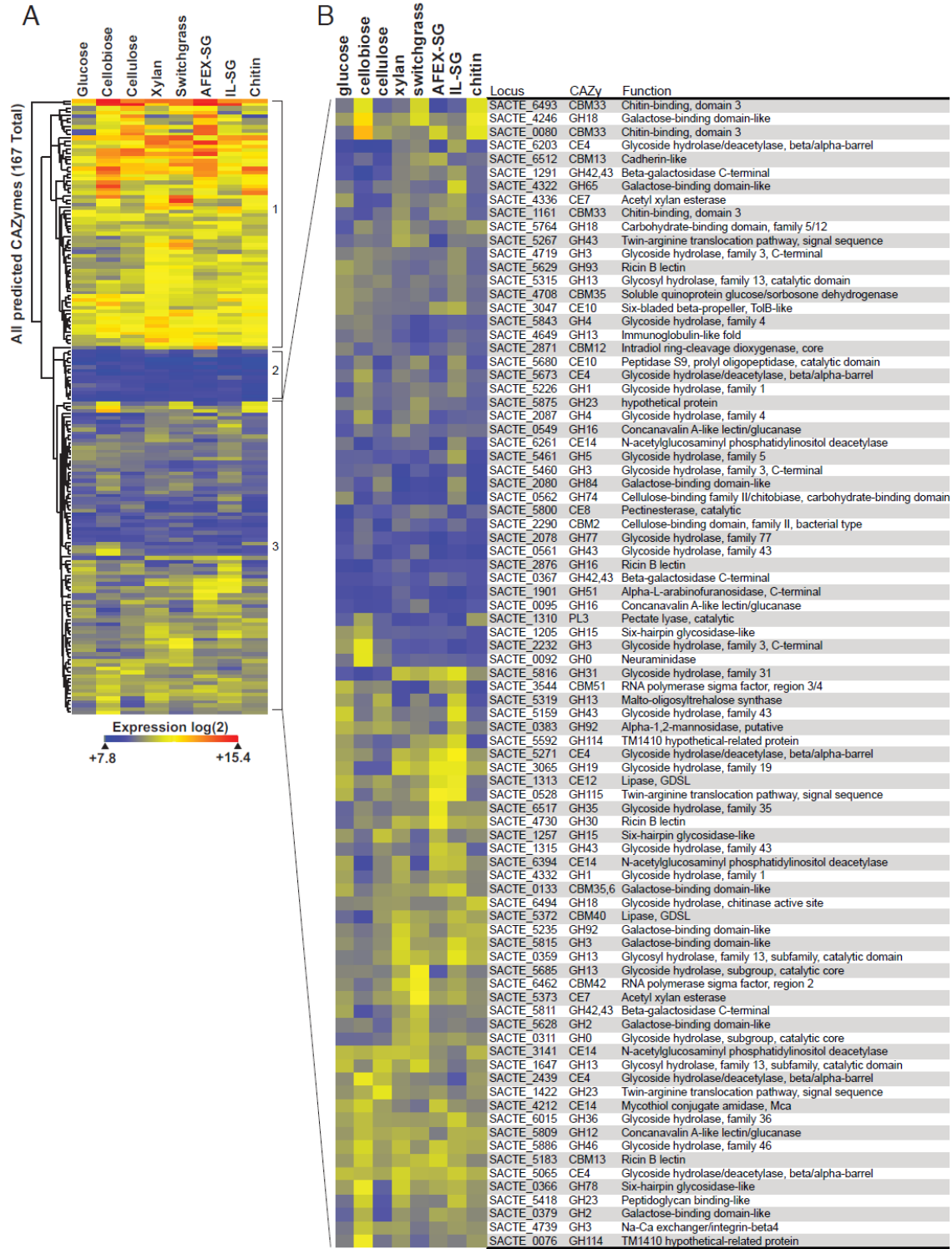




**Supplementary Figure A6.9. Genome-wide expression changes for growth on the substrate chitin versus glucose visualized as a Cytoscape interaction network.** Other information provided in the caption to Supplementary Figure A6.4.



**Supplementary Figure A6.10. Expression of 167 predicted CAZy genes in ActE, highlighting group 2 genes.** These genes showed no signal above the average genomic expression intensity ( $\log_2 = 10.5$ ). (A) Clustering of genes with similar expression profiles. (B) Additional information on group 2 genes including expression profile, SACTE\_locus ID, CAZy family, and annotated function.



**Supplementary Figure A6.11. Expression of 167 predicted CAZy genes in ActE, highlighting group 3 genes.** (A) Clustering of genes with similar expression profiles. (B) Additional information on group 3 genes including expression profile, SACTE\_locus ID, CAZy family, and annotated function.

**Supplementary Table A6.1.** Proteins separated by ion exchange chromatography and identified by mass spectrometry.

| Protein band <sup>a</sup> | Locus      | Catalytic domain | CBM  | Functional class <sup>b</sup> |
|---------------------------|------------|------------------|------|-------------------------------|
| a                         | SACTE_3159 | CBM33            | CBM2 | cellulase                     |
| b                         | SACTE_0265 | GH10             | CBM2 | xylanase                      |
| c                         | SACTE_4755 | GH64             |      | beta-1,3-glucanase            |
| d                         | SACTE_0482 | GH5              | CBM2 | cellulase                     |
| e                         | SACTE_0237 | GH6              | CBM2 | cellulase                     |
| f                         | SACTE_0236 | GH48             | CBM2 | cellulase                     |
| g                         | SACTE_3717 | GH9              | CBM2 | cellulase                     |
| h                         | SACTE_2347 | GH5              | CBM2 | mannanase                     |

<sup>a</sup> Protein bands labeled in Figure 3A were identified by MALDI-TOF mass spectrometry. <sup>b</sup> Function identified by assays of individual fractions from ion exchange chromatography.

**Supplementary Table A6.2.** Spectra count of proteins identified on each substrate, where top 95 % spectra covered were highlighted green, light purple, purple, blue, orange, pink, light blue and yellow on glucose, cellobiose, cellulose, xylan switchgrass, AFEX-SG, IL-SG and chitin, respectively. Available at <http://www.nature.com/article-assets/npg/srep/2013/130107/srep01030/extref/srep01030-s2.xls>.

## **Appendix 7: Microbial community analysis of leaf-cutter ant refuse dumps reveals patterns of community succession and strata-specific coexisting modules**

**Publication in preparation:** Jarrod J. Scott, Gina R. Lewin, Garret Suen, Susannah G. Tringe, Kerrie W. Barry, Justin C. Touchon, and Cameron R. Currie. Microbial community analysis of leaf-cutter ant refuse dumps reveals patterns of community succession and strata-specific coexisting modules.

**Author contributions:** Conceived and designed the experiments: JJS, GS, SGT, JCT, CRC. Performed the experiments: JJS, KWB. Wrote the paper: JJS, GRL, CRC.

### **A7.1 Abstract**

Colonies of the leaf-cutter ant *Atta colombica* cultivate fungus using plant material. The ants generate large waste deposits of partially-degraded, recalcitrant plant fragments on the forest floor. These refuse dumps are enriched in nutrients compared to surrounding soil and develop distinct vertical stratification over time. To better understand community dynamics within these nutrient rich reservoirs, we sequenced the 16S rRNA genes from five strata within dumps of three *A. colombica* colonies and the 18S rRNA gene from fungus gardens and refuse dumps. Bacterial diversity was dominated by Proteobacteria, Actinobacteria, and Bacteroidetes. Community structure was more similar among samples collected from equal depths, rather than from the same dump, suggesting a succession in community structure. Network analysis of co-occurring operational taxonomic units revealed numerous inter-lineage associations, ranging from solitary pairs to larger assemblages, some of which were specific to particular strata. Our findings indicate that refuse dumps host a diverse and specialized community, likely evolved to exploit this unusual niche. This work represents a significant step in understanding the role

microbial communities play in degrading recalcitrant plant biomass within leaf-cutter ant refuse dumps and highlights the potential of refuse dumps as a model for exploring patterns in microbial ecology.

## A7.2 Introduction

Resource islands are localized concentrations of organic matter that exhibit elevated nutrient levels compared to adjacent areas (Herman *et al.*, 1995; Schlesinger *et al.*, 1996; Schade & Hobbie, 2005). Animals alter nutrient distributions and create resource islands through soil redistribution (e.g., burrowing, mound construction) (Ayarbe, 2000; Willott *et al.*, 2000) and resource concentration (e.g., food caching) (Hagenah & Bennett, 2013). Additionally, animal waste contains undigested biomass, and though not common, some animals concentrate large amounts of waste over long time periods (e.g., Farji-Brener & Medina, 2000; Hanley, 2002; Sternberg *et al.*, 2007). In environments of low nutrient availability, localized waste deposits become an important resource for other organisms, thus functioning as resource islands (Willott *et al.*, 2000; Hanley, 2002; Chronáková *et al.*, 2009; Hudson *et al.*, 2009).

Throughout Central and South America, leaf-cutter ants (*Atta* and *Acromyrmex*) are dominant herbivores in forest, savannah, and grassland ecosystems (Weber, 1972; Hölldobler & Wilson, 1990). Leaf-cutter ants use fresh plant material, primarily leaves but also flowers, fruits, and seeds, as growth substrate for a saprophytic basidiomycetous fungal mutualist (Belt, 1888; Weber, 1972) (Figure A7.1a). This fungal cultivar is grown as a garden matrix within interconnected subterranean chambers (Figure A7.1b). The fungus produces specialized protein-rich hyphal swellings (Möller, 1893) that serve as the colony's primary food source (Weber,

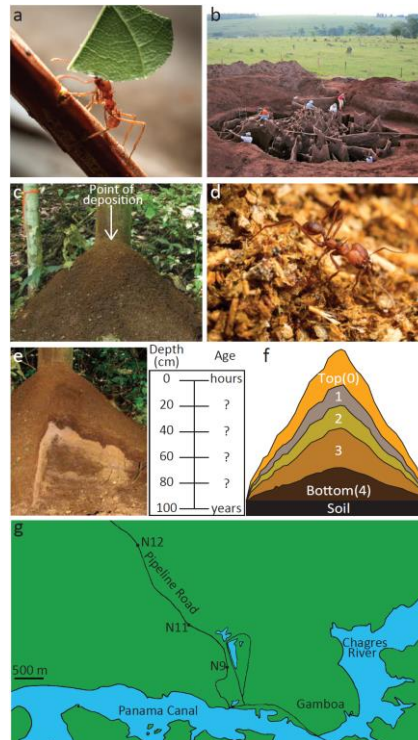


**Figure A7.1. The leaf-cutter ant refuse system.** **a)** A foraging leaf-cutter ant (*Atta sexdens*) harvesting fresh leaf material that will subsequently be used as growth substrate in the fungus gardens. **b)** Cement casting and excavation reveals the massive and intricate structure of a mature leaf-cutter ant nest. **c)** A refuse dump from a mature colony of *A. colombica* (N11, on map in **g**) showing the point of refuse deposition. In this case, workers ascend the base of the tree and deposit refuse particles on the top of the dump where other workers then redistribute the material. **d)** An *Atta sexdens* worker tends the refuse dump (colony at University of Wisconsin–Madison). **e)** Cross section of the same dump in **c** revealing vertical stratification. **f)** Simplified diagram of a refuse dump indicating the sampling scheme used in this study. A total of 5 g (wet weight) of pooled refuse material was collected from each stratum, about 20 cm apart. Strata were designated according to depth by the suffix D0, D1, D2, D3, or D4; D0 represented the top (freshest) of the dump, D1–D3 corresponded to intermediate strata, and D4 represented the bottom (oldest) stratum at the dump/soil interface. Scale bar between **e** and **f** indicates the relative depth and theoretical age of each stratum. **g)** Location of the primary nests used in this study. N9 (09°08′57.0″N 079°43′52.7″W), N11 (09°08′03.1″N 079°43′18.6″W), and N12 (09°07′34.8″N 079°42′52.3″W). (Photo credits: a and d, Don Parsons; b, Wolfgang Thaler; c and e, Jarrod J. Scott).

1972). Mature colonies of leaf-cutter ants, especially in the genus *Atta*, contain millions of

workers and can survive for over 10 years (Hölldobler & Wilson, 1990). Leaf-cutter ants significantly modify the bioavailability of resources within the nutrient-depleted tropical soil through nest excavation (Moutinho & Nepstad, 2003; Verchot *et al.*, 2003), foraging, and waste deposition. A single colony of leaf-cutter ants can harvest 50–500 kg (dry weight) of fresh plant material per year (Herz, Beyschlag & Hölldobler, 2007b) from a wide range of plant species (Wirth *et al.*, 2003), thus mediating net primary productivity and nutrient redistribution. However, approximately one-half of all harvested material, including the majority of the recalcitrant cellulose and lignin in leaves, is not utilized in the fungus garden and instead is discarded as waste (Herz, Beyschlag & Hölldobler, 2007a) (Figure A7.1c).

A significant percentage of workers are devoted to waste management in leaf-cutter ant colonies (Bot *et al.*, 2001; Hart & Ratnieks, 2002). Workers remove exhausted substrate, spent fungal biomass, and dead ants from the colony and transport this material to specific dumping sites. Then, other workers redistribute waste particles around the dump (Figure A7.1d). Most species of *Atta* deposit their waste in deep belowground chambers (Stahel & Geijskes, 1939; Moser, 1963; Bot *et al.*, 2001; Moreira *et al.*, 2004) where the concentration of organic matter may influence deep soil horizon nutrient dynamics (Verchot *et al.*, 2003). *Atta colombica*, native to lowland rainforests of Costa Rica, Colombia, and Panama, is different from most other *Atta* species in that their waste is deposited in a single aboveground heap located downslope of the colony often at the base of a tree (Haines, 1978; Hart & Ratnieks, 2001) (Figure A7.1c). Compared to surrounding forest soils, external refuse dumps of *A. colombica* colonies have higher pH and significantly greater levels of bulk carbon, nitrogen, and phosphorus including soluble nutrients (e.g., phosphate, nitrate, and ammonium) (Hudson *et al.*, 2009). Therefore,



these dumps act as nutrient reservoirs within nutrient deficient tropical soil systems (Hudson *et al.*, 2009). In fact, plants, nematodes, and arthropods clearly exploit these resource patches (Farji-Brener & Medina, 2000; Hanley, 2002; Hudson *et al.*, 2009).

Previous work has demonstrated that microbial diversity is lower in tropical forest soils than other soil types (Fierer & Jackson, 2006). Given that refuse dumps are highly localized depositions of organic matter exhibiting distinct characteristics from surrounding soils (Hudson *et al.*, 2009), these habitats are attractive candidates for the establishment of unique and diverse microbial communities. For these elevated nutrient levels to be maintained, refuse dumps require continual input of fresh waste material from the colony. Once input ceases (either through nest abandonment or colony death), nutrient concentrations return to base levels within a year (Hudson *et al.*, 2009). Additionally, refuse dumps develop distinct vertical stratification likely due to the extended periods of dry conditions (January–April), when material accumulates, followed by a rainy season (May–December), when material is partially degraded with the increase in water potential due to the heavy rainfall (Figure A7.1e, f) (Haines, 1978).

In a previous study, we used a phospholipid fatty acid/fatty acid methyl ester analysis to evaluate the microbial communities of fungus gardens and refuse dumps (laboratory-maintained and field-collected colonies) (Scott *et al.*, 2010). We found that community structure changed significantly between gardens and dumps, with gardens dominated by Gram-negative bacteria and dumps enriched for Gram-positive bacteria. Additionally, we demonstrated that refuse material from field-collected samples exhibited the highest variation in community structure as well as the greatest diversity. In subsequent investigations, Suen *et. al.*, (Suen *et al.*, 2010) and

Aylward et. al., (Aylward *et al.*, 2012) used sequence-based approaches to show that Gammaproteobacteria dominated fungus garden communities of *A. colombica*. To date however, no study has used sequence-based approaches to explore microbial community composition in leaf-cutter ant refuse dumps.

In this study, we generated clone libraries (16S rRNA and 18S rRNA) and pyrotag datasets (16S rRNA only) from five distinct strata from refuse dumps of three mature *A. colombica* colonies collected along Pipeline Road in the canal zone of Central Panama (Figure A7.1g). These data were combined with previously published fungus garden community datasets to describe taxonomic composition and estimate overall alpha and beta community diversity across the leaf-cutter ant system. We then conducted co-occurrence analyses to determine significant pairwise operational taxonomic unit (OTU) combinations and to construct networks of coexisting OTUs overlaid with taxonomic classification and habitat enrichment data. Our results show that similar strata of leaf-cutter ant dumps have more similar microbial communities, and leaf-cutter ant refuse dumps contain modules of coexisting OTUs with a specific stratum preference.

### **A7.3 Results**

In this study we generated 15,917 16S rRNA sequences 220 (*Ba*), 507,958 16S rRNA pyrosequencing reads (*Py*), and 28,344 18S rRNA sequences (*Eu*, gardens and dumps). After processing and quality control, the *Ba* dataset contained 14,618 high-quality sequences (1233 bp average length), ranging from 549 (N9D1) to 2617 (N9D4) sequences per sample (mean of 975 [SD, 532]). The *Py* dataset contained 391,893 quality reads (231 bp average length), ranging from 462 (N12D3) to 53,534 (N9D1) reads per sample (mean of 18,489 [SD, 11,970]). The *Eu*

dataset contained 18,732 sequences (849 bp average length) and sequences per sample ranged from 305 (N11D4) to 1811 (N12GB) (mean of 895 [SD, 436]) (Supplementary Table A7.1).

Distance matrix clustering revealed the presence of 2905 OTUs (97% similarity) in the *Ba* dataset, 2719 in the *Eu* dataset, and 7652 in the *Py* dataset (Supplementary Table A7.2).

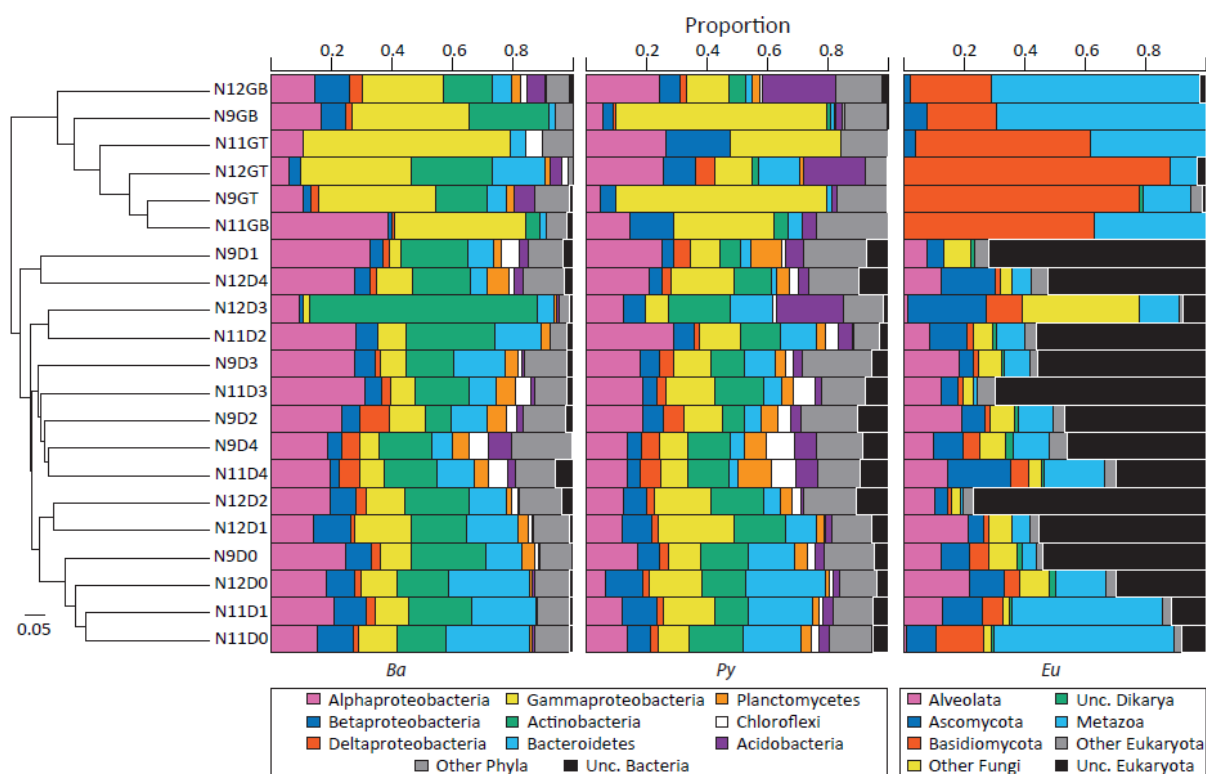


**Figure A7.2. Clustering analysis and taxonomic distribution of garden and dump samples.** OTU-based community clustering analysis derived from the Bray-Curtis similarity coefficient for the combined 16S/18S clone libraries and 16S pyrotag data (for the individual clustering analyses, see Figure S1). Sample names indicate the nest of origin followed by dump stratum (D) or garden samples (G). Overlaid are relative proportions of sequences from major taxonomic lineages present in the 16S clone libraries (*Ba*, left), 16S pyrotag libraries 620 (*Py*, center), and the 18S clone libraries (*Eu*, right). Color scheme based on the Color Universal Design (<http://jfly.iam.u-tokyo.ac.jp/color/>, last accessed 8.8.2014).

### ***A7.3.1 Taxonomic description of leaf-cutter ant colonies***

At 97% OTU identity, dominant members of refuse dump bacterial communities included Alphaproteobacteria (*Ba*, 21.1%; *Py*, 16.1%), Betaproteobacteria (*Ba*, 7.1%; *Py*, 6.9%), Gammaproteobacteria (*Ba*, 9.8%; *Py*, 14.1%), Actinobacteria (*Ba*, 22.8%; *Py*, 13.9%), and Bacteroidetes (*Ba*, 13.9%; *Py*, 10.6%) (Figure A7.2; Supplementary Tables A7.1 and A7.2). Although we found considerable variation across samples, these groups accounted for over 70% of total OTU diversity. Other groups were well represented in particular samples, but present in low amounts overall. In addition, both datasets revealed the presence of a broad range of microbial groups. The *Ba* dataset contained representatives from 39 phyla, and the *Py* dataset contained representatives from 45 phyla. Unclassified bacteria comprised 2.2% and 6.7% of the *Ba* and *Py* datasets, respectively. In contrast, a major portion of OTUs in the *Eu* dataset from refuse dumps remained unclassified (44.8 %). The eukaryotic lineages identified in dumps included Metazoa (14.7%), Alveolata (12.1%), Ascomycota (11.0%), and Basidiomycota (4.6%). Our survey of fungus gardens revealed high richness of both Basidiomycota (56.1%; the ants' cultivar is a basidiomycete fungus) and Metazoa (39.8%), however we found a low OTU richness of Ascomycota (2.2%), and we did not detect Alveolata or other types of Fungi (Figure A7.2; Supplementary Tables A7.1 and A7.2).

### ***A7.3.2 Alpha and beta diversity estimates***

OTU-based beta diversity estimates using the Bray-Curtis Index revealed that garden samples formed a distinct cluster to the exclusion of dump samples (Figure A7.2), especially for both 16S datasets (Supplementary Figure A7.1). Additionally, community structure was more similar among samples collected from similar depths, rather than samples originating from the same

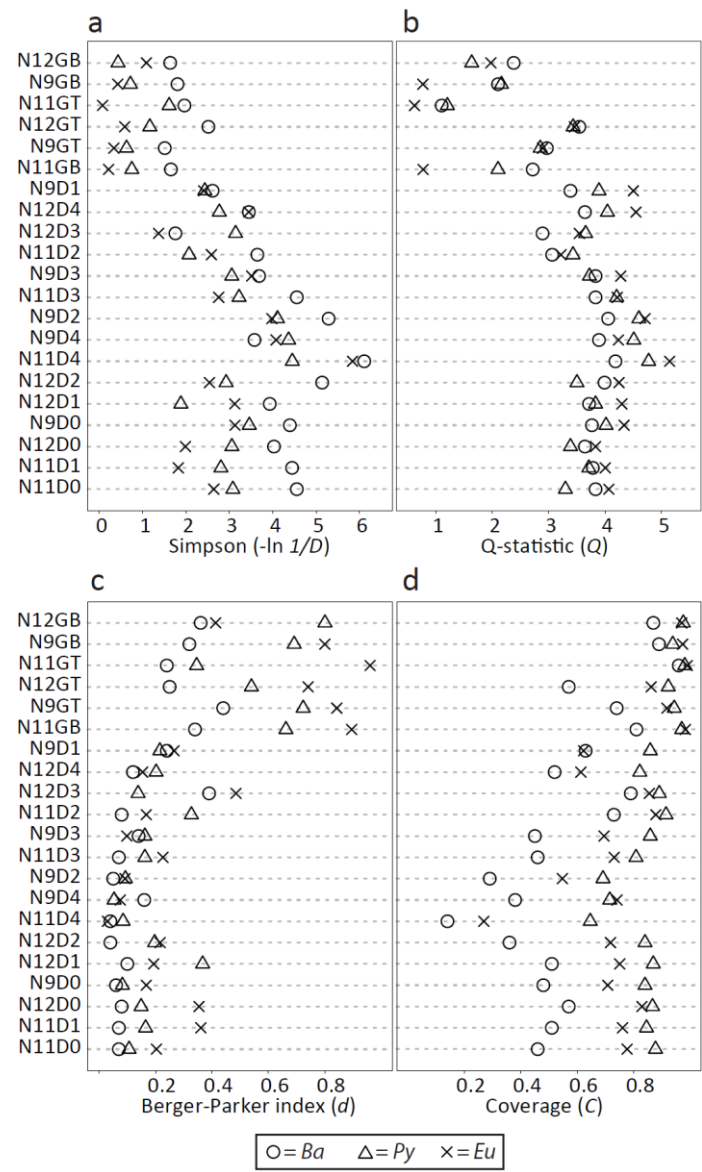


**Figure A7.3. Alpha diversity estimates of garden and dump communities.** Samples are ordered according to the results of the clustering analysis (Figure 2). **a)** Negative natural logarithm transformed (Pielou, 1975) Simpson index ( $1/D$ ), **b)** Q-statistic ( $Q$ ), **c)** Berger-Parker Index ( $d$ ), **d)** Good's coverage ( $C$ ). Data is shown for 16S clone libraries (*Ba*, circles), 16S pyrosequencing libraries (*Py*, triangles), and 18S clone libraries (*Eu*, X).

nest. Specifically all top dump strata (D0) clustered together along with two of the three D1 samples (Figure A7.2). Based on alpha diversity estimates of the *Ba*, *Py*, and *Eu* datasets, gardens were less diverse and better sampled than dumps (Figure A7.3). Calculations of the

inverse Simpson's index indicated that, on average, dump samples (mean [SD]  $1/D = 56.6$  [85.3]; range, 3.94–450) contained higher diversity than garden samples (mean [SD]  $1/D = 3.69$  [2.88]; range, 1.08–12.41). Similarly, Q-statistic values showed greater diversity in dump samples (mean [SD]  $Q = 57.3$  [29.2]; range, 18.0–171.7) compared to garden samples (mean [SD]  $Q = 12.7$  [10.7]; range, 1.86–34.6). Dominant OTUs were proportionally more abundant in gardens (mean [SD]  $d = 0.577$  [0.240]; range, 0.240–0.961) than dumps (mean [SD]  $d = 0.162$  [0.107]; range, 0.030–0.485) using the Berger-Parker index. Finally, Good's coverage estimates for gardens (mean [SD]  $C = 0.904$  [0.107]; range, 0.570–0.990) were greater than for dumps (mean [SD]  $C = 0.669$  [0.193]; range, 0.140–0.915) (Figure A7.3).

### A7.3.3 Co-occurrence analysis

We performed co-occurrence analyses to identify pairs of OTUs whose presence were significantly correlated. Approximately 0.65% and 3.5% of the pairwise comparisons were significant (97% OTU definition) in the *Ba/Eu* and *Py* datasets, respectively (Table A7.1, which includes a summary of the findings at multiple OTU definitions). Analysis of simulated data indicated a FDR of 2.8% in the *Ba/Eu* dataset and 2.6% in the *Py* dataset 270 at a 97% OTU definition (Table A7.1). Co-occurrence analyses revealed 1162 significant interactions (*edges*, *m*) encompassing 428 OTUs (*nodes*, *n*) from the *Ba/Eu* dataset (Figure A7.4a) and 23,005 significant interactions (*edges*) from 1135 OTUs (*nodes*) in the *Py* dataset (to simplify network visualization of the *Py* dataset we only considered interactions where  $q < 0.01$  rather than 0.05) (Figure A7.4b).

**Table A7.1.** Results of co-occurrence analysis

| Dataset                             | Parameter                                                            | OTU Definition (%) |         |         |         |         |         |
|-------------------------------------|----------------------------------------------------------------------|--------------------|---------|---------|---------|---------|---------|
|                                     |                                                                      | 90                 | 93      | 95      | 97      | 98      | 99      |
| Clone libraries<br>( <i>Ba/Eu</i> ) | No. of original OTUs <sup>1</sup>                                    | 1891               | 2901    | 3962    | 5848    | 7534    | 10 697  |
|                                     | No. of OTUs after filtering <sup>2</sup>                             | 378                | 479     | 524     | 602     | 625     | 582     |
|                                     | No. of co-occurrence tests                                           | 70 892             | 114 502 | 137 044 | 180 916 | 195 008 | 169 080 |
|                                     | No. of significant OTU pairs <sup>3</sup>                            | 420                | 666     | 813     | 1162    | 1387    | 1247    |
|                                     | No. of OTUs <sup>4</sup>                                             | 237                | 324     | 362     | 428     | 462     | 429     |
|                                     | No. strict occurring pairs ( $J = 1.0$ ) <sup>5</sup>                | 57                 | 94      | 115     | 187     | 314     | 372     |
|                                     | No. of checkerboard pairs ( $J = 0.0$ ) <sup>6</sup>                 | 0                  | 2       | 4       | 6       | 4       | 2       |
|                                     | Random data: no. of coexisting OTU pairs ( $q < 0.05$ ) <sup>7</sup> | 34                 | 55      | 50      | 33      | 45      | 21      |
| Pyrotag<br>( <i>Py</i> )            | FDR (empirical estimates) <sup>8</sup>                               | 0.081              | 0.083   | 0.062   | 0.028   | 0.032   | 0.017   |
|                                     | No. of original OTUs <sup>1</sup>                                    | 1089               | 2332    | 4074    | 8162    | 12,547  | 17,901  |
|                                     | No. of OTUs after filtering <sup>2</sup>                             | 309                | 543     | 792     | 1172    | 1371    | 1355    |
|                                     | No. of co-occurrence tests                                           | 47 586             | 147 153 | 313 236 | 686 206 | 939 135 | 917 335 |
|                                     | No. of significant OTU pairs <sup>3</sup>                            | 4113               | 9828    | 15 534  | 23 005  | 29 557  | 28 025  |
|                                     | No. of OTUs <sup>4</sup>                                             | 291                | 522     | 764     | 1135    | 1339    | 1310    |
|                                     | No. strict occurring pairs ( $J = 1.0$ ) <sup>5</sup>                | 74                 | 258     | 503     | 979     | 1622    | 1822    |
|                                     | No. of checkerboard pairs ( $J = 0.0$ ) <sup>6</sup>                 | 3                  | 41      | 105     | 91      | 121     | 69      |
|                                     | Random data: no. of coexisting OTU pairs ( $q < 0.05$ ) <sup>7</sup> | 1964               | 776     | 1530    | 595     | 577     | 445     |
|                                     | FDR (empirical estimates) <sup>8</sup>                               | 0.478              | 0.079   | 0.099   | 0.026   | 0.020   | 0.016   |

<sup>1</sup>Total number of observed OTUs prior to filtering.

<sup>2</sup>Total number of OTUs after removing those that occurred in less than three samples.

<sup>3</sup>Number of significant pairwise associations after correcting for False Discovery Rate (FDR) ( $q < 0.05$ ).

<sup>4</sup>Number of OTUs from significant pairwise associations.

<sup>5</sup>Number of OTU pairs that only co-occurred.

<sup>6</sup>Number of OTU pairs that never co-occurred.

<sup>7</sup>Number of significant OTU associations observed in a randomized dataset.

<sup>8</sup>Ratio of significant OTU associations in the randomized to the actual dataset.

Of these OTUs, 16% and 5% were strictly co-occurring ( $J = 1.0$ ) in the *Ba/Eu* and *Py* datasets, respectively, and aside from a few examples of checkerboard associations ( $J = 0.0$ ), we did not observe significant values below 0.7 (Supplementary Figure A7.2). To determine if any pairwise interactions exhibited habitat specificity, we calculated the proportion of instances where two OTUs co-occurred more often in a given habitat (garden, dump top, or dump bottom) or exhibited no habitat preference (mixed). In the *Ba/Eu* dataset, 38% and 18% of significant interactions showed habitat preference to the dump top and dump bottom, respectively.

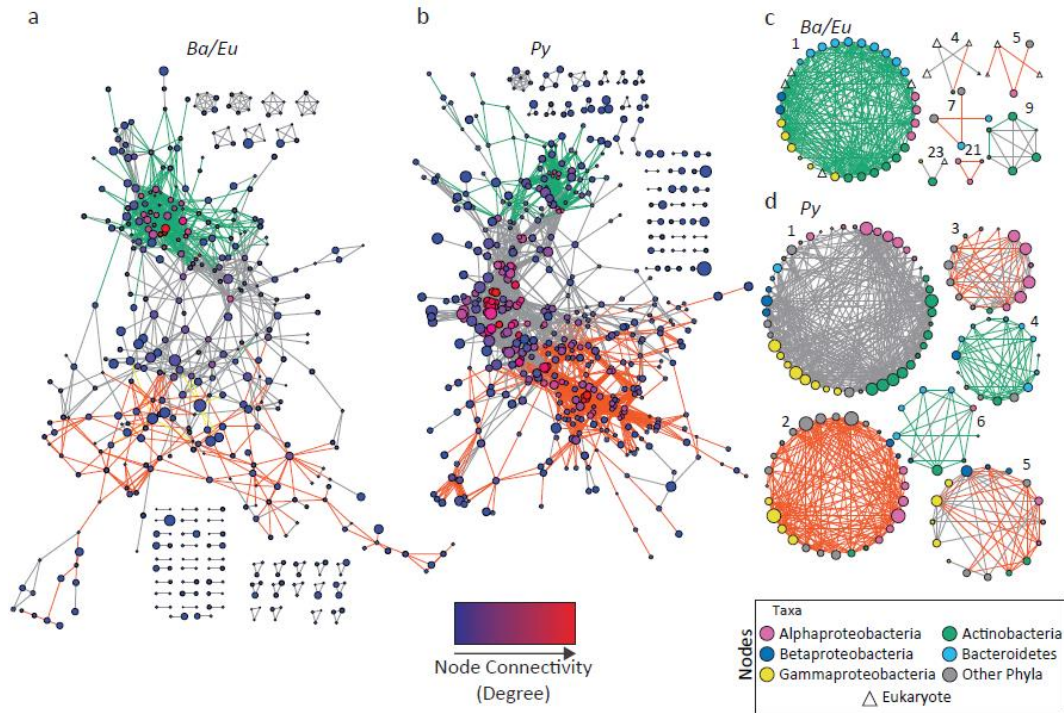


**Figure A7.4. Co-occurrence networks.** Results of the co-occurrence analyses visualized as networks of significant OTU interactions from the **a**) *Ba/Eu* and **b**) *Py* datasets. To simplify network visualization only interactions where  $q < 0.01$  are displayed for the *Py* dataset. Nodes represent unique OTUs and edges correspond to significant associations. Node size is proportional to abundance (natural log transformed) and node color denotes degree (number of connections) (blue, low connectivity; red, high connectivity). Edge color denotes the habitat specificity of each interaction (green, dump top; grey, mixed (top and bottom); red, bottom). Panels **c** (*Ba/Eu*) and **d** (*Py*) represent distinct modules of coexisting OTUs identified using MCL of the co-occurrence networks in **a** and **b**, respectively. This represents the same analysis as in Figure S4 but here modules only include OTUs that were both significantly enriched (based on METASTATS analysis) and found across all three nests (see supporting information). Node shape denotes domain, size is proportional to the number of connections, and color represents taxa. Edge color denotes the habitat specificity of each interaction as described above. For a list of OTUs, cluster numbers, and associated details, see Supplementary Table A7.3.

#### A7.3.4 Network topology

We did not find a significant correlation between the number of sequences per OTU and degree ( $k$ ) for either the *Ba/Eu* (Pearson's,  $r[426] = 0.055$ ,  $p = 0.255$ ) or the *Py* (Pearson's,  $r[510] =$



0.062,  $p = 0.164$ ) datasets (Supplementary Figure A7.3a, b), meaning that OTU abundance was not correlated with OTU connectivity within the networks. In fact, we observed numerous instances where degree and the number of sequences were at opposite extremes for a given OTU (Supplementary Figure A7.3a, b; Supplementary Table A7.3). In contrast, if OTUs were grouped by phyla, a positive correlation was observed between the number of sequences per phyla and degree (*Ba/Eu*,  $r[28] = 0.665$ ,  $p < 0.001$ ; *Py*,  $r[23] = 0.800$ ,  $p < 0.001$ ) (Supplementary Figure A7.3c, d). Alphaproteobacteria, Actinobacteria, Gammaproteobacteria, and Bacteroidetes were the most highly connected taxa in the co-occurrence analysis. Supplementary Tables A7.3 and A7.4 summarize the taxon distribution in the full datasets versus the co-occurrence analysis. See the results section of supporting information and Supporting Table A7.6 for a summary of the topological properties for the *Ba/Eu* and *Py* networks.

### ***A7.3.5 Markov clustering and coexisting modules***

To identify modules of co-occurring OTUs, we clustered both co-occurrence networks using the MCL algorithm (based on the *Jaccard* similarity coefficient). MCL analysis revealed 105 distinct modules in the *Ba/Eu* dataset and 99 in the *Py* dataset, ranging in size from more than 40 OTUs per module to modules containing only 2 OTUs (see Supplementary Figure A7.4 and Supplementary Table A7.3 for all clusters). Larger modules, in particular, contained a wide diversity of phylum-level interlineage associations. Some smaller clusters had diverse associations, while others contained closely related OTUs (Supplementary Figure A7.4; Supplementary Table A7.3).

**Table A7.2.** Putative core microbes in leaf-cutter ant refuse dumps

| Habitat | Network | OTU                 | Taxa <sup>1</sup>   |                      |                         | k <sup>2</sup> |
|---------|---------|---------------------|---------------------|----------------------|-------------------------|----------------|
| Top     | Ba      | 410Ba               | Betaproteobacteria  | Comamonadaceae       | <i>Rhodoferrax</i>      | 43             |
|         |         | 222Ba               | Actinobacteria      | Microbacteriaceae    | <i>Leucobacter</i>      | 43             |
|         |         | 733Ba               | Actinobacteria      | Microbacteriaceae    | <i>Microbacterium</i>   | 30             |
|         |         | 407Ba               | Gammaproteobacteria | Moraxellaceae        | <i>Acinetobacter</i>    | 43             |
|         |         | 406Ba               | Gammaproteobacteria | Xanthomonadaceae     | <i>Thermomonas</i>      | 28             |
|         |         | 143Ba               | Bacteroidetes       | Sphingomonadaceae    | <i>Sphingobium</i>      | 26             |
|         |         | 678Ba               | Bacteroidetes       | Sphingobacteriaceae  | <i>Sphingobacterium</i> | 22             |
|         | Py      | 77Py                | Gammaproteobacteria | Xanthomonadaceae     | <i>Dokdonella</i>       | 41             |
|         |         | 84Py                | Actinobacteria      | Mycobacteriaceae     | <i>Mycobacterium</i>    | 38             |
|         |         | 147Py               | Actinobacteria      | Microbacteriaceae    | unclassified            | 38             |
| 153Py   |         | Actinobacteria      | Microbacteriaceae   | unclassified         | 30                      |                |
| 196Py   |         | Gammaproteobacteria | unclassified        | unclassified         | 30                      |                |
| Bottom  | Ba      | 670Ba               | Bacteroidetes       | Flexibacteraceae     | <i>Cytophaga</i>        | 12             |
|         |         | 761Ba               | Crenarchaeota       | Nitrososphaerales    | Nitrososphaera          | 9              |
|         |         | 328Ba               | Gemmatimonadetes    | Gemmatimonadaceae    | unclassified            | 8              |
|         |         | 1064Ba              | Chloroflexi         | Caldilineaceae       | <i>Caldilinea</i>       | 7              |
|         |         | 646Ba               | Gemmatimonadetes    | Gemmatimonadaceae    | unclassified            | 6              |
|         |         | 133Ba               | Alphaproteobacteria | Rhodospirillaceae    | unclassified            | 6              |
|         | Py      | 1479Py              | Chlamydiae          | Rhabdochlamydiaceae  | unclassified            | 45             |
|         |         | 1460Py              | Chloroflexi         | Thermomicrobia       | unclassified            | 45             |
|         |         | 1519Py              | Gammaproteobacteria | Coxiellaceae         | unclassified            | 45             |
|         |         | 1485Py              | Alphaproteobacteria | Rhizobiales          | unclassified            | 45             |
|         |         | 462Py               | Actinobacteria      | Acidimicrobiales     | unclassified            | 41             |
|         |         | 514Py               | Alphaproteobacteria | Rhizobiales          | unclassified            | 38             |
|         |         | 228Py               | Actinobacteria      | Solirubrobacteraceae | <i>Solirubrobacter</i>  | 38             |
|         |         | 288Py               | Gammaproteobacteria | Xanthomonadaceae     | <i>Lysobacter</i>       | 38             |

The most highly connected, enriched, and shared OTUs in the co-occurrence networks

<sup>1</sup>Taxa are classified to genus. In the case of 'Unclassified' genera the next identified taxonomic rank is given.

<sup>2</sup>Total number of connections (degrees) in the co-occurrence network.

Roughly half of the OTUs in each dataset were shared across all three nests (212 of 428 OTUs in the *Ba/Eu* co-occurrence analysis; 355 of 512 OTUs in the *Py* co-occurrence analysis [ $q < 0.01$ ]). The majority of these OTUs were representatives from dominant phyla, notably Proteobacteria, Actinobacteria, and Bacteroidetes (Supplementary Table A7.3; Supplementary Figure A7.4). We also observed numerous cases of coexisting OTUs from underrepresented phyla (< 1% of total OTU richness) shared across all three nests (Supplementary Tables A7.1, A7.2, A7.3). Of all shared OTUs, 44% from the *Ba/Eu* dataset (94 OTUs) and 66% from the *Py* dataset (236 OTUs) were enriched in a given habitat (upper dump or lower dump; Supplementary Table A7.3) based

on METASTATS analysis (see supporting information methods for detailed methods). Several of the large modules of coexisting OTU assemblages were dominated by OTUs shared across the three nests and enriched in either the top or bottom of the refuse dump. These included module 1 in the *Ba/Eu* dataset and modules 1–6 in the *Py* dataset (Figure A7.4c, d). Module 1 from the *Ba/Eu* dataset was a highly connected, inter-lineage cluster dominated by microbes enriched in the upper dump strata (Figure A7.4c). In the *Py* dataset, we found two inter-lineage modules (4 and 6) enriched in the upper dump strata and three specific to the lower dump strata (2, 3, and 5) (Figure A7.4d). Additionally, we found a large module with no habitat specificity (module 1 in the *Py* dataset). We did not observe any large modules that were exclusively found in garden habitats (Supplementary Table A7.3).

#### ***A7.3.6 Putative core community members of refuse dumps***

To identify putative core microbes in leaf-cutter ant refuse dumps, we selected OTUs that were a) significantly enriched in the upper or lower strata of the dump, b) shared across all nests, and c) highly connected network elements (Table A7.2, Supplementary Table A7.3). OTUs fitting these criteria in the top dump strata included *Acinetobacter*, *Rhodoferrax*, and members of the Microbacteriaceae, especially *Microbacterium* and *Leucobacter* (*Ba* dataset), as well as multiple Actinomycetales, Xanthomonadales, and Flavobacteriaceae (*Py* dataset). Arthropoda and Basidiomycota were also identified as core members in the upper strata (*Eu*). In contrast, putative core members from the bottom dump strata included the archaeon Nitrososphaeraceae, Gemmatimonadetes, Chloroflexi, and several Rhodospirillales, and Rhizobiales (*Ba*), as well as representatives from the Thermomicrobia, Rhabdochlamydiaceae, Acidimicrobiales, and Rhizobiales (*Py*). Oribatida (mites) and Hypocreales (ascomycete fungus) were identified as core

members in the lower strata (*Eu*). Multiple unclassified OTUs were abundant and enriched in both the upper and lower strata.

#### **A7.4 Discussion**

Waste management is an integral part of animal biology because waste material often harbors agents that may adversely affect host fitness (Mawdsley *et al.*, 1995; Hart & Ratnieks, 2002; Pichtel, 2005). However, animal waste contains undigested material, so it can also serve as an important nutrient resource for other organisms (Bernath & Kunz, 1981; Farji-Brener & Medina, 2000; Hanley, 2002; Sternberg *et al.*, 2007; Chronáková *et al.*, 2009). Leaf-cutter ant colonies are keystone components of the Neotropical ecosystem, partially because they concentrate large amounts of waste material in areas of relatively low nutrient availability. This behavior has important consequences for the establishment of novel food webs and ultimately evolutionary specialization (Hanley, 2002; Sternberg *et al.*, 2007; Hudson *et al.*, 2009). To date, little is known about microbial communities in leaf-cutter ant refuse dumps, or how they compare to fungus gardens or tropical soil. To initiate our understanding of this unique ecosystem, we performed the first large-scale analysis of microbial community composition of leaf-cutter ant refuse dumps using a combination of full-length and pyrosequencing analyses. We identified a highly diverse, and seemingly distinct, microbial assemblage dominated by Proteobacteria (Alphaproteobacteria, Gammaproteobacteria, and Betaproteobacteria), Actinobacteria, and Bacteroidetes.

Our comparison of dump communities to Neotropical soils revealed considerable differences in community composition. Soils contained a significantly higher richness of Acidobacteria and

Verrucomicrobia (> 30% on average) (Navarrete *et al.*, 2010; Nemergut *et al.*, 2010) than dump communities (< 2%), but a lower richness of Actinobacteria, Bacteroidetes, and Gammaproteobacteria (Figure A7.2; Supplementary Figure A7.5). Refuse dumps have higher pH than tropical soils and contain greater concentrations of many important nutrients, most notably nitrate, ammonium, and phosphate (Hudson *et al.*, 2009), as well as numerous mineral elements (Haines, 1978). Given that pH and nutrient availability are important drivers of microbial diversity (Fierer & Jackson, 2006; Fierer *et al.*, 2007; Hansel *et al.*, 2008; Jones *et al.*, 2009; Nemergut *et al.*, 2011), it is likely that the differences in community structure between soil and refuse dumps were partially due to differences in abiotic conditions.

Comparison of refuse dumps to fungus gardens indicated a clear transition of community composition and structure, expanding on earlier findings (Scott *et al.*, 2010; Suen *et al.*, 2010) (Figure A7.2). Although Proteobacteria dominated dump communities (> 50%), the relative abundance of Gammaproteobacteria significantly decreased relative to the gardens. Furthermore, we found an increase in both Actinobacteria and Bacteroidetes diversity within refuse dumps. These changes were likely the result of the large percentage of colony resources devoted to maintaining the ants' fungal cultivar and a low-diversity bacterial community in the gardens. Additionally, the microbial community may have shifted between the garden and the refuse dump because the leaf material is more degraded in the dumps, and external dumps are exposed to continual environmental inputs (Weber, 1972; Currie *et al.*, 1999; Currie & Stuart, 2001; Hart & Ratnieks, 2002). We originally hypothesized that waste material was removed from the garden and immediately transferred to the dump and thus predicted that communities in the top dump layer (freshest material) would be most similar to those in the bottom layer of the fungus

gardens. However, the microbial communities in the top of the dump clustered more with other dump communities, not fungus garden communities, especially in the *Ba* and *Py* datasets (Figure A7.2, Supplementary Figure A7.1). An explanation that supports these results is waste caching, a behavior where workers deposit waste material in a cache before it is transported to dumps. This behavior is not well studied, but it has been observed occasionally in wild colonies of *A. colombica* (Anderson & Ratnieks, 2000; Hart & Ratnieks, 2002; Hart *et al.*, 2002) and in a small number of lab-maintained *Atta cephalotes* colonies (an internal waste dumper) (Hart & Ratnieks, 2000). These findings raise a number of interesting questions including the purpose of this behavior, how long material is cached, and what effects caching has on microbial community development prior to refuse deposition.

We found a diverse array of microbes shared across all three dumps (Supplementary Tables A7.7 and A7.8) including members of uncommon and even rare taxa (< 0.5% total sequences). However, it is important to note that many of the OTUs were present as either singletons or doubletons, similar to other highly diverse environments (Table S9). Therefore, the proportion of shared diversity identified in refuse dumps may increase with deeper sequencing and/or more targeted sampling. Despite this, our findings suggest that refuse dumps contain a core microbial community that is consistent across spatiotemporal scales. Within dumps, communities were more similar based on depth rather than nest of origin (Figure A7.2). We found very few OTUs shared across all strata within dumps, even at a considerably relaxed OTU cutoff (Supplementary Table A7.10), and the abundance of numerous microbial groups changed significantly across strata (Figure S7). These results suggest a broad-scale succession in microbial communities within individual dumps. As might be expected, the OTUs enriched in the upper layers were

dominated by aerobes including *Microbacterium* and members of the phyla Bacteroidetes. In contrast, many of the OTUs enriched in the lower levels of the dump were indicative of soil or root-associated communities, including several anaerobic organisms that use a diverse suite of complex molecules as energy sources, (e.g., anaerobic steroid-degrading *Steroidobacter* and ammonia oxidizing archaea) (Supplementary Table A7.3). These results support the hypothesis that the lower levels of the dump are more anoxic and nutrient-deplete than the upper layers and may act as transitional communities between the upper dump layers and surrounding soil. Furthermore, the co-occurrence and Markov-based clustering analyses identified distinct and taxonomically diverse clusters of co-occurring OTUs that were specific to either the top or bottom strata of the dump. This enrichment of specific clusters could result from interactions between the OTUs or from similar environmental preferences (Figure A7.4).

Over the last several decades, the integration of microbial and molecular research into the study of the leaf-cutter ant system has led to many insights including the identification of symbiotic bacteria associated with both the ants and their fungus gardens (Currie *et al.*, 1999; Van Borm *et al.*, 2002; Pinto-Tomás *et al.*, 2009; Aylward *et al.*, 2013). Here, we show that a specialized microbial community is present in leaf-cutter ant refuse dumps—which are dissimilar from either gardens or soil—that likely evolved to exploit this unique habitat. The distinct transitions in microbial communities from gardens to dump, and from the upper dump strata to the lower strata, likely mirror changes in plant biomass composition and abiotic conditions. The results discussed here demonstrate the utility of leaf-cutter ant dumps as a model for addressing fundamental issues in microbial ecology including biogeography, community assembly, and

evolution of niche specialization. This study also contributes to our growing appreciation for the role that leaf-cutter ants play as drivers of microbial diversity.

## **A7.5 Materials and Methods**

### ***A7.5.1 Site description, sampling, and study overview***

Between May 30 and June 10, 2008, at the beginning of the rainy season, we sampled fungus gardens and refuse dumps from three mature colonies (Nest 9, Nest 11, and Nest 12, here abbreviated N9, N11, and N12) of the leaf-cutter ant, *A. colombica*, all located along Pipeline Road in the Canal Zone of central Panama (Figure A7.1g). We aseptically collected triplicate 30 x 115 mm cores (~5 g wet weight) from the fungus gardens and from five discernable dump strata (refuse dumps were vertically cross-sectioned to reveal internal stratification, identified by visual distinction) (Figure A7.1e, f). In this study, fungus garden samples were assigned names based on their nest and whether they were from the top or bottom half of the garden (top or bottom, e.g., N9GT), as are dump samples, with D0 referring to the top, or freshest refuse material, and D4 indicating the bottom, oldest level of the dump (e.g., N11D4, see Figure A7.1f). Samples were stored at -80°C then transported on dry ice to the University of Wisconsin-Madison for microbial community analysis. In the first of a series of studies, we used phospholipid fatty acid/fatty acid methyl ester analysis to assess broad-scale microbial community structure in gardens and dumps from the three colonies described above as well as two additional field colonies and nine lab maintained colonies (Scott *et al.*, 2010). We then used 16S rRNA clone library and 16S pyrotag analyses to further explore the microbiome of fungus gardens from the three nests (Suen *et al.*, 2010). In the current study, we present 16S clone library, 16S pyrotag, and 18S clone library data of dump strata. For comparative purposes, we



also include the 16S data generated previously from fungus garden samples (Suen *et al.*, 2010) and present 18S clone libraries generated for the garden samples. We use the following abbreviations to denote the three different datasets used in this study: 16S clone libraries, *Ba*; 18S clone libraries, *Eu*; and 16S pyrotag libraries, *Py*. OTUs were each named with a unique number followed by the dataset designation (e.g., 119*Py*). See supporting information for detailed methods on DNA extraction, sequencing, analysis, and taxonomic classification. All sequences generated in this study were deposited in the European Nucleotide Archive (ENA) with the following accessions numbers: 16S clone libraries, LN560104–LN574657; 18S clone libraries, LN574658–LN589471. 16S pyrotag reads were deposited in the ENA Sequence Read Archive, under project accession number PRJEB7274, sample accession numbers ERS543652–ERS543663.

#### ***A7.5.2 Estimates of community diversity***

To reduce the effects of unequal sampling on diversity assessment, we randomly subsampled our datasets based on the number of sequences in the smallest sample. We estimated alpha diversity for each sample using the following indices: the Simpson index ( $D$ , nonparametric index sensitive to abundant OTUs) (Simpson, 1949); Berger-Parker Index ( $d$ , measures importance of dominant OTU) (Berger & Parker, 1970); the  $Q$ -statistic ( $Q$ , parametric index not skewed by very rare/abundant OTUs) (Kempton & Taylor, 1976; 1978); and Good's coverage ( $C$ , an estimate of sample coverage) (Good & Toulmin, 1956). We estimated beta diversity using the Bray-Curtis similarity coefficient (Bray & Curtis, 1957) to calculate the community similarity between all pairwise comparisons for the *Ba*, *Eu*, and *Py* datasets. We clustered each resulting similarity matrix using the UPGMA algorithm (Sokal & Michener, 1958), creating newick-

formatted dendrograms that were then visualized in FIGTREE v.1.3.1

(<http://tree.bio.ed.ac.uk/software/figtree/>, last accessed 07.22.2014). To identify OTUs enriched in the fungus garden, upper dump strata, or lower dump strata, we used METASTATS (White *et al.*, 2009) (<http://metastats.cbcb.umd.edu/detection.html>; last accessed 7.22.2014; see supporting information for additional details).

### **A7.5.3 Co-occurrence analysis**

To explore potentially important OTU associations, we performed co-occurrence analyses to identify OTU pairs that occurred together more often than expected by chance, using methods described previously by Chaffron *et al.*, 2010. We generated an OTU-by-sample table containing presence/absence information and removed all OTUs present in less than three samples. The *Ba/Eu* datasets were combined, and the *Py* dataset was analyzed separately. We then tested the co-occurrence significance for all possible pairwise OTU combinations (at multiple OTU definitions) using Fisher's exact test (FET). For each pairwise comparison, the four cells in the contingency table denoted the number of samples containing 1) both OTUs, 2 and 3) only one of the two OTUs, or 4) neither OTU. We then adjusted all *p*-values to account for multiple testing using the positive false discovery rate (pFDR) controlling procedure (Storey, 2002; 2003; Carlson *et al.*, 2009). We used an online False Discovery Rate Calculator to calculate *q*-values (i.e., adjusted *p*-values) for all OTU combinations (<http://research.microsoft.com/en-us/um/redmond/projects/MSCompBio/>, last accessed 8.8.2014) (Carlson *et al.*, 2009). Next, we verified these results by generating simulated communities from the original datasets (*Ba/Eu* and *Py*) using *commsimulator* as implemented in the *vegan* library (Oksanen *et al.*, 2012) in R (R Core Team, 2014) (<http://www.r-project.org>, last accessed 8.2.2014). Here OTUs were randomly

reassigned to the different samples while maintaining row (OTU) and column (samples) marginal totals, effectively shuffling the connections between OTUs and samples. We then calculated the number of significant OTU associations in the simulated datasets (as above) and hence empirically estimated the false discovery rate of the co-occurrence analysis (Chaffron *et al.*, 2010).

#### ***A7.5.4 Network analysis of coexisting OTUs***

Based on the results of the co-occurrence analysis, we then constructed networks of significant interactions between OTUs and visualized the results in CYTOSCAPE (v2.8.2) (Smoot *et al.*, 2011) using the *yFiles organic* layout (<http://www.yworks.com>, last accessed 7.22.2014). Here, nodes ( $n$ ) represented OTUs and edges ( $m$ ) represented significant interactions based on the adjusted FET values ( $q < 0.05$ ). To assess whether these co-occurrence networks exhibited nonrandom community structure, we compared the *Ba/Eu* and *Py* networks to randomly generated networks using the Erdos-Renyi Model (Erdős & Rényi, 1959) as implemented in the *RandomNetwork* plugin (v.1.0) for CYTOSCAPE. Random networks (1000 randomizations) were constructed using the same number of  $n$  and  $m$  as in the originals. This method also maintained the average number of neighbors and network density. See supporting information for details for topological analyses of co-occurrence networks.

#### ***A7.5.5 Identification of coexisting OTU clusters***

Next, to identify modules of coexisting OTUs from both co-occurrence networks, each network was clustered using a Markov Clustering (MCL) algorithm (Enright *et al.*, 2002) as implemented in the CYTOSCAPE plugin, *clusterMaker* (v.1.9) (Morris *et al.*, 2011). Edge-weight clustering

was based on the correlation coefficient between significant OTU associations, calculated using the *Jaccard* similarity coefficient ( $J = SAB / (SA + SB - SAB)$  where *SAB* was the number of samples in which OTUs *A* and *B* co-occurred, *SA* was the total number of samples containing OTU *A*, and *SB* was the total number of samples containing OTU *B*) (Chaffron *et al.*, 2010). Thus, OTUs that always co-occurred and were never found alone returned an edge weight value of 1.0. In contrast, those OTUs that never co-occurred returned an edge weight value of 0.0 (Jaccard, 1901).

## A7.6 References

Anderson C, Ratnieks F. (2000). Task partitioning in insect societies: novel situations. *Insect Soc* 47:198–199.

Ayarbe J. (2000). Mammal mounds stimulate microbial activity in a semiarid shrubland. *Ecology* 81:1150–1154.

Aylward FO, Burnum KE, Scott JJ, Suen G, Tringe SG, Adams SM, *et al.* (2012). Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens. *ISME J* 1688–1701.

Aylward FO, Burnum-Johnson KE, Tringe SG, Teiling C, Tremmel DM, Moeller JA, *et al.* (2013). *Leucoagaricus gongylophorus* produces diverse enzymes for the degradation of recalcitrant plant polymers in leaf-cutter ant fungus gardens. *Appl Environ Microbiol* 79:3770–3778.

Belt T. (1888). *The Naturalist in Nicaragua*. J. Murray: London.

Berger WH, Parker FL. (1970). Diversity of planktonic Foraminifera in deep sea sediments. *Science* 168:1345–1347.

Bernath R, Kunz T. (1981). Structure and dynamics of arthropod communities in bat guano deposits in buildings. *Can J Zool* 59:260–270.

Bot A, Currie CR, Hart AG, Boomsma J. (2001). Waste management in leaf-cutting ants. *Ethol Ecol Evol* 13:225–237.

Bray RJ, Curtis JT. (1957). An ordination of upland forest communities of southern Wisconsin. *Ecol Monogr* 27:325–349.

Carlson JM, Heckerman D, Shani G. (2009). Estimating false discovery rates for contingency tables. Microsoft Research: Redmond, WA.

Chaffron S, Rehrauer H, Pernthaler J, Mering von C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res* 20:947–959.

Chronáková A, Horák A, Elhottová D, Kristůfek V. (2009). Diverse archaeal community of a bat guano pile in Domica Cave (Slovak Karst, Slovakia). *Folia Microbiol (Praha)* 54:436–446.

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

Currie CR, Stuart AE. (2001). Weeding and grooming of pathogens in agriculture by ants. *Proc Biol Sci* 268:1033–1039.

Enright AJ, Van Dogen S, Ouzounis CA. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584.

Erdős P, Rényi A. (1959). On random graphs. I. *Publications Matheaticae* 6:290–297.

Farji-Brener AG, Medina CA. (2000). The importance of where to dump the refuse: seed banks and fine roots in nests of the leaf-cutting ants *Atta cephalotes* and *A. colombica*. *Biotropica* 32:120–126.

Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, *et al.* (2007). Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl Environ Microbiol* 73:7059–7066.

Fierer N, Jackson RB. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA* 103:626–631.

Good IJ, Toulmin GH. (1956). The number of new species, and the increase in population coverage, when a sample is increased. *Biometrika* 43:45–63.

Hagenah N, Bennett NC. (2013). Mole-rats act as ecosystem engineers within a biodiversity hotspot, the Cape Fynbos. *J Zool* 289:19–26.

Haines B. (1978). Element and energy flows through colonies of the leaf-cutting ant, *Atta colombica*, in Panama. *Biotropica* 10:270–277.

Hanley RS. (2002). A new species of Mexican *Tinotus* from the refuse piles of *Atta* ants, including an annotated world catalog of *Tinotus* (Coleoptera: Staphylinidae: Aleocharinae: Aleocharini). *Coleopt Bull* 56:453–471.

Hansel CM, Fendorf S, Jardine PM, Francis CA. (2008). Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl Environ Microbiol* 74:1620–1633.

Hart AG, Anderson C, Ratnieks FLW. (2002). Task partitioning in leafcutting ants. *Acta Ethol* 5:1–11.

Hart AG, Ratnieks F. (2000). Leaf caching in *Atta* leafcutting ants: discrete cache formation through positive feedback. *Anim Behav* 59:587–591.

Hart AG, Ratnieks F. (2001). Task partitioning, division of labour and nest compartmentalization collectively isolate hazardous waste in the leafcutting ant *Atta cephalotes*. *Behav Ecol Sociobiol* 49:387–392.

Hart AG, Ratnieks F. (2002). Waste management in the leaf-cutting ant *Atta colombica*. *Behav Ecol* 13:224–231.

Herman RP, Provencio KR, Herrera-Matos J, Torrez RJ. (1995). Resource islands predict the distribution of heterotrophic bacteria in Chihuahuan Desert soils. *Appl Environ Microbiol* 61:1816–1821.

Herz H, Beyschlag W, Hölldobler B. (2007a). Assessing herbivory rates of leaf-cutting ant (*Atta colombica*) colonies through short-term refuse deposition counts. *Biotropica* 39:476–481.

Herz H, Beyschlag W, Hölldobler B. (2007b). Herbivory rate of leaf-cutting ants in a tropical moist forest in Panama at the population and ecosystem scales. *Biotropica* 39:482–488.

Hölldobler B, Wilson E. (1990). *The Ants*. Belknap Press of Harvard University Press: Cambridge, Mass.

Hudson TM, Turner BL, Herz H, Robinson JS. (2009). Temporal patterns of nutrient availability around nests of leaf-cutting ants (*Atta colombica*) in secondary moist tropical forest. *Soil Biol Biochem* 41:1088–1093.



Jaccard P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin de la Société Vaudoise des Sciences Naturelles 37:547–579.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J 3:442–453.

Kempton RA, Taylor LR. (1976). Models and statistics for species diversity. Nature 262:818–820.

Kempton RA, Taylor LR. (1978). The Q-statistic and the diversity of floras. Nature 275:252–253.

Mawdsley J, Bardgett R, Merry R, Pain B, Theodorou M. (1995). Pathogens in livestock waste, their potential for movement through soil and environmental pollution. Appl Soil Ecol 2:1–15.

Moreira A, Forti LC, Andrade AP, Boaretto MA, Lopes J. (2004). Nest architecture of *Atta laevigata* (F. Smith, 1858) (Hymenoptera: Formicidae). Stud Neotrop Fauna Environ 39:109–116.

Morris JH, Apeltsin L, Newman AM, Baumbach J, Wittkop T, Su G, *et al.* (2011). *cluster-Maker*: a multi-algorithm clustering plugin for Cytoscape. BMC Bioinformatics 12:436.

Moser JC. (1963). Contents and structure of *Atta texana* nest in summer. *Ann Entomol Soc Am* 56:1–6.

Moutinho P, Nepstad D. (2003). Influence of leaf-cutting ant nests on secondary forest growth and soil properties in Amazonia. *Ecology* 84:1265–1276.

Möller A. (1893). *Die Pilzgärten einiger südamerikanischer Ameisen*. G. Fischer.

Navarrete AA, Cannavan FS, Taketani RG, Tsai SM. (2010). A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. *Diversity* 2:787–809.

Nemergut DR, Cleveland CC, Wieder WR, Washenberger CL, Townsend AR. (2010). Plotscale manipulations of organic matter inputs to soils correlate with shifts in microbial community composition in a lowland tropical rain forest. *Soil Biol Biochem* 42:2153–2160.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, *et al.* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* 13:135–144.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, OHara RB, *et al.* (2012). *Vegan*. Community Ecology Package. R package. <http://CRAN.R-project.org/package=vegan>.

Pichtel J. (2005). *Waste Management Practices: Municipal, Hazardous and Industrial*. Taylor & Francis: Florida, USA.

Pielou EC. (1975). *Ecological Diversity*. Wiley InterScience: New York.

Pinto-Tomás AA, Anderson MA, Suen G, Stevenson DM, Chu FST, Cleland WW, *et al.* (2009). Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science* 326:1120–1123.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Schade JD, Hobbie SE. (2005). Spatial and temporal variation in islands of fertility in the Sonoran Desert. *Biogeochemistry* 73:541–553.

Schlesinger W, Raikes J, Hartley A. (1996). On the spatial pattern of soil nutrients in desert ecosystems. *Ecology* 77:364–374.

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

Simpson EH. (1949). Measurement of diversity. *Nature* 163:688.

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 27:431–432.

Sokal RR, Michener CC. (1958). A statistical method for evaluating systematic relationships. *The University of Kansas Scientific Bulletin* 38:1409–1438.

Stahel G, Geizskes DC. (1939). Ueber den bau der nester von *Atta cephalotes* L. und *Atta sexdens* L. (Hym. Formicidae). *Revista de Entomologia, Rio de Janeiro* 10:27–78.

Sternberg LDSL, Pinzon MC, Moreira MZ, Moutinho P, Rojas EI, Herre EA. (2007). Plants use macronutrients accumulated in leaf-cutting ant nests. *Proceedings of the Royal Society B: Biological Sciences* 274:315–321.

Storey JD. (2002). A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol* 64:479–498.

Storey JD. (2003). The positive false discovery rate: A Bayesian interpretation and the  $q$  value. *Ann Stat* 31:2013–2035.

Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás AA, *et al.* (2010). An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet* 6:e1001129.

Van Borm S, Billen J, Boomsma J. (2002). The diversity of microorganisms associated with *Acromyrmex* leafcutter ants. *BMC Evol Biol* 2:9.

Verchot LV, Moutinho PR, Davidson EA. (2003). Leaf-cutting ant (*Atta sexdens*) and nutrient cycling: deep soil inorganic nitrogen stocks, mineralization, and nitrification in Eastern Amazonia. *Soil Biol Biochem* 35:1219–1222.

Weber N. (1972). *Gardening Ants: The Attines*. American Philosophical Society: Philadelphia.

White JR, Nagarajan N, Pop M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5:e1000352.

Willott S, Miller A, Incoll L. (2000). The contribution of rabbits (*Oryctolagus cuniculus* L.) to soil fertility in semi-arid Spain. *Biol Fertil Soils* 31:379–384.

Wirth R, Herz H, Ryel R, Beyschlag W, Hölldobler B. (2003). *Herbivory of Leaf-Cutting Ants: A Case Study on Atta colombica in the Tropical Rainforest of Panama*. Springer-Verlag: Berlin Heidelberg New York.

## **A7.7 Supporting Material**

### ***A7.7.1 Supplementary Methods***

#### *A7.7.1.a DNA extraction, clone library sequencing and bar-coded pyrosequencing*

Total DNA was extracted from 250 mg (wet weight) of each refuse material sample using the MoBio Power Soil DNA Extraction Kit (MoBio Laboratories, Carlsbad, CA, USA). Independent extractions were performed for each of the three culture-independent assays. For clone library construction, extracted DNA was PCR amplified (20 cycles) using universal bacterial (27F [5'-AGA GTT TGA TCC TGG CTC AG-3'] and 1391R [5'-GAC GGG CRG TGW GTR CA-3']), archaeal (4aF [5'-TCC GGT TGA TCC TGC CRG-3'] and 1391R [5'-GAC GGG CRG TGW GTR CA-3']), and eukaryotic (515F [5'-GTG CCA AGC AGC CGC GGT AA-3']) and 1209R [5'-GGG CAT CAC AGA CCT G-3']) primers and cloned into the pCR4-TOPO vector (Invitrogen). Clones were sequenced using capillary-based sequencing and assembled as previously described (Warnecke *et al.*, 2007). Samples were also PCR-amplified with bacterial-specific primers corresponding to the V6–V8 region (1492R [59-TAC GCY TAC CTT GTT ACG ACT T-39] and 926F [59-AAA CTY AAA KGA ATT GAC GG-39] fused to 5-base barcodes [reverse primer only] and 454-Titanium adapter sequences) and then sequenced on a Roche 454 FLX GS Titanium pyrosequencer (Margulies *et al.*, 2005).

#### *A7.7.1.b Sequence processing and taxonomic classification*

Unless otherwise noted, clone libraries (16S and 18S) and pyrotag (16S) datasets were processed similarly with a user-defined pipeline implemented in MOTHUR v.1.29.0 (Schloss *et al.*, 2009). We included the raw data from the previously published fungus garden study (Suen *et al.*, 2010) in this pipeline for comparative purposes. For pyrotag reads, we first removed oligos and barcodes (*trim.seqs*). We then removed exact duplicates from all datasets (*unique.seqs*) while

retaining abundance information. Global alignments and distance matrices were generated for pyrotag reads using the following steps. First, sequencing data from all 21 samples (gardens and dumps) were combined into a single fasta-formatted text file (one for each of the three datasets). We then screened the three datasets (*screen.seqs*) to remove sequences that were too short (16S clone library, < 1200 bp; 18S clone library, < 850 bp; 16S pyrotag, < 225 bp), contained more than eight homopolymers, or contained any ambiguities. Next, we generated initial alignments for the three datasets against the same SILVA-based reference alignment ([http://www.mothur.org/wiki/Silva\\_reference\\_files](http://www.mothur.org/wiki/Silva_reference_files); accessed 3.21.2012; 50 000 columns; 14 956 Bacterial, 2297 Archaeal, and 1238 Eukaryotic sequences) using the Needleman alignment method (k-tuple size = 8, match reward = +1, mismatch penalty = -3, gap extension penalty = -1, gap opening penalty = -5, search = kmer) (Schloss & Westcott, 2011; Schloss, 2009; White *et al.*, 2009; Schloss, 2010). Alignments were checked against a SILVA-based secondary structure mapping file (*align.check*) and then filtered to eliminate empty columns (*filter.seqs*, trump = '.', vertical = T, soft = 50). To reduce overestimation of richness due to pyrosequencing error (Storey, 2002; Kunin *et al.*, 2010), the pyrotag dataset was also preclustered (*pre.cluster*) (Storey, 2003; Huse *et al.*, 2010). This algorithm first ranks sequences in order of abundance and then merges rare sequences with more abundant sequences based on a user-defined threshold (mismatch = 3) (Kunin *et al.*, 2010; Huse *et al.*, 2010; Storey, 2002; 2003). Putative chimeras were then eliminated from the filtered alignments using UCHIME as implemented in MOTHRU (*chimera.uchime*; template = self, settings = default) (Edgar *et al.*, 2011). We also identified (using the classification methods outlined below) and removed (*remove.lineage*) possible contaminant sequences from each of the three datasets. Any eukaryotic sequences found in the 16S datasets were eliminated as were any bacterial sequences found in the 18S dataset. Finally,

we constructed phylip-formatted distance matrices (*dist.seqs*, *calc* = one gap, *count end penalty* = true) (Oksanen *et al.*, 2012; Schloss, 2010) and then clustered the matrices using the average-neighbor algorithm (*cluster*) (Schloss & Westcott, 2011). Unless otherwise noted, all OTU-based analyses were performed at 97% similarity for each dataset.

For taxonomic classification of our datasets, we used a MOTHUR-modified Greengenes reference taxonomy (last accessed 07.22.2014) (McDonald *et al.*, 2012) and a MOTHUR-modified SILVA reference taxonomy (last accessed 07.22.2014) under the following parameters (*classify.seqs*): *search parameter* = kmer; *search method* = bayesian; *ktuple size* = 8; *match reward* = +1; *mismatch penalty* = -1; *gap open penalty* = -2; *gap extension penalty* = -1; *bootstrap cutoff* = 80, 60, or 50 for 16S clone library, 18S clone library and 16S pyrotag sequences, respectively (Claesson *et al.*, 2009); *iterations* = 1000. Based on the resultant taxonomic summaries, we also calculated consensus taxonomies for individual OTUs (*classify.otu*).

#### A7.7.1.c Detecting differentially abundant OTUs

To reduce the number of OTUs present in the cluster analysis (Figure 4 c, d) we first used METASTATS to detect differentially abundant OTUs (97% cutoff) between habitat types (fungus gardens, upper dump strata, or lower dump strata) (<http://metastats.cbcb.umd.edu/detection.html>; accessed 8.8.2014). METASTATS uses a FDR to improve specificity in complex environments, as well as FET to deal with sparsely populated matrices (White *et al.*, 2009). Based on the results of the METASTATS analysis, OTUs that were not significantly enriched were eliminated. We also eliminated OTUs that were not shared across all three nests.



#### *A7.7.1.d Comparison of full-length and pyrosequencing datasets*

Not surprisingly, the larger pyrosequencing dataset was better suited to detect shared diversity across sample types, particularly from representatives of rare taxa ( $< 1\%$  total diversity), compared with the full-length dataset. Conversely, a much lower proportion of the full-length data remained unclassified, even though we employed a higher bootstrap cutoff in taxonomic analysis than in the pyrosequencing data. Despite the differences inherent to these two sequencing platforms, they revealed complimentary patterns of community diversity and community relatedness among samples.

#### *A7.7.1.e Changes in community composition within refuse dumps*

We modeled the relationship between stratum depth and diversity for 15 microbial groups (mean totals  $\geq 2\%$  from either the full-length or pyrotag datasets) representing a range of taxonomic ranks (phylum, class, order, and family). Briefly, we calculated the richness (total OTUs) and abundance (total clones) per taxonomic group for all samples. This allowed us to assess the relationships between taxonomic diversity and sample depth using a generalized linear model with underlying quasibinomial error function and logit link functions (also called logistic regression). A quasibinomial error function is similar to a binomial error function, except that it accounts for over dispersion of the model and penalizes  $P$  values accordingly (Pinheiro & Bates, 2000). We compared the fit of each generalized linear model using likelihood ratio test, often called an analysis of deviance because of the similar function of analysis of variance in ordinary linear models. Results were summarized as average taxon abundance (SD) by stratum (data pooled according to stratum,  $n = 3$ ) (Figure S6). Because of the small sample size we did not perform post-hoc analysis.

#### A7.7.1.f Changes in community

We used the CYTOSCAPE plugin *NetworkAnalyzer* (Assenov *et al.*, 2008) to calculate topological parameters for both the random and empirical networks, including the following: clustering coefficient ( $C$ , averaged for each node) (Albert & Barabási, 2002; Watts & Strogatz, 1998); average degree connectivity (the average of  $k_i$  for all  $n$ , where  $k_i$  is the number of edges connected to node  $i$ ) (Dorogovtsev & Mendes, 2002); degree distribution (fraction of  $n$  with degree  $k$ ) (Newman & Girvan, 2004); and mean shortest path (the average shortest-path length,  $\ell$ , between any two nodes) (Dorogovtsev & Mendes, 2002).

### A7.7.2 Supplementary Results

#### A7.7.2.a Network topology

The *Ba/Eu* and *Py* co-occurrence networks (Figure 4a, b) exhibited topological properties consistent with small world networks—specifically both networks had short average path lengths between nodes ( $\ell$ ) and high levels of clustering (clustering coefficient,  $C$ ) (Table S6) (Albert & Barabási, 2002; Watts & Strogatz, 1998). Characteristic path lengths ( $\ell$ ) were similar in scale to random networks of identical dimensions ( $n$  and  $m$ ), but the observed clustering coefficients were much greater than the same random networks (Table S6) indicating high levels of local clustering. The degree distribution of both networks approximated power law distributions,  $P_{\text{deg}}(k) \propto k^{-\gamma}$  (*Ba/Eu*,  $R^2 = 0.839$  and  $\gamma = 1.4$ ; *Py*,  $R^2 = 0.709$ , and  $\gamma = 1.0$ ) (Figure S7; Table S6), meaning that certain nodes were highly connected while the majority of nodes had relatively few connections,  $k$  (Figure 4a, b; Table S3). In both cases however,  $\gamma$  was smaller than reported values of other real-world networks (typically  $2 < \gamma < 3$  (Newman & Girvan, 2004; Chaffron *et al.*, 2010), but larger than the values derived from the random networks (Table S6).

### A7.7.3 References

Albert R, Barabási A-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**:47–97.

Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M. (2008). Computing topological parameters of biological networks. *Bioinformatics* **24**:282–284.

Chaffron S, Rehrauer H, Pernthaler J, Mering von C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Research* **20**:947–959.

Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, Smidt H, *et al.* (2009). Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4**:e6669.

Dorogovtsev SN, Mendes JFF. (2002). Evolution of networks. *Advances in Physics* **51**:1079–1187.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**:2194–2200.

Huse SM, Welch DM, Morrison HG, Sogin ML. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* **12**:1889–1898.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**:118–123.

- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
- McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, *et al.* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal* **6**:610–618.
- Navarrete AA, Cannavan FS, Taketani RG, Tsai SM. (2010). A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. *Diversity* **2**:787–809.
- Nemergut DR, Cleveland CC, Wieder WR, Washenberger CL, Townsend AR. (2010). Plot-scale manipulations of organic matter inputs to soils correlate with shifts in microbial community composition in a lowland tropical rain forest. *Soil Biology and Biochemistry* **42**:2153–2160.
- Newman MEJ, Girvan M. (2004). Finding and evaluating community structure in networks. *Physical Review E* **69**:026113.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, OHara RB, *et al.* (2012). vegan: Community Ecology Package. *R package*. <http://CRAN.R-project.org/package=vegan>.
- Pinheiro JC, Bates DM. (2000). Mixed-Effects Models in S and S-PLUS. Springer-Verlag: New York, NY.
- Schloss PD. (2009). A high-throughput DNA sequence aligner for microbial ecology studies. *PLoS ONE* **4**:e8230.
- Schloss PD. (2010). The effects of alignment quality, distance calculation method, sequence

filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol* **6**:e1000844.

Schloss PD, Westcott SL. (2011). Assessing and improving methods used in operational taxonomic unit-based approaches for 16Ss rRNA gene sequence analysis. *Applied and Environmental Microbiology* **77**:3219–3226.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**:7537–7541.

Storey JD. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B* **64**:479–498.

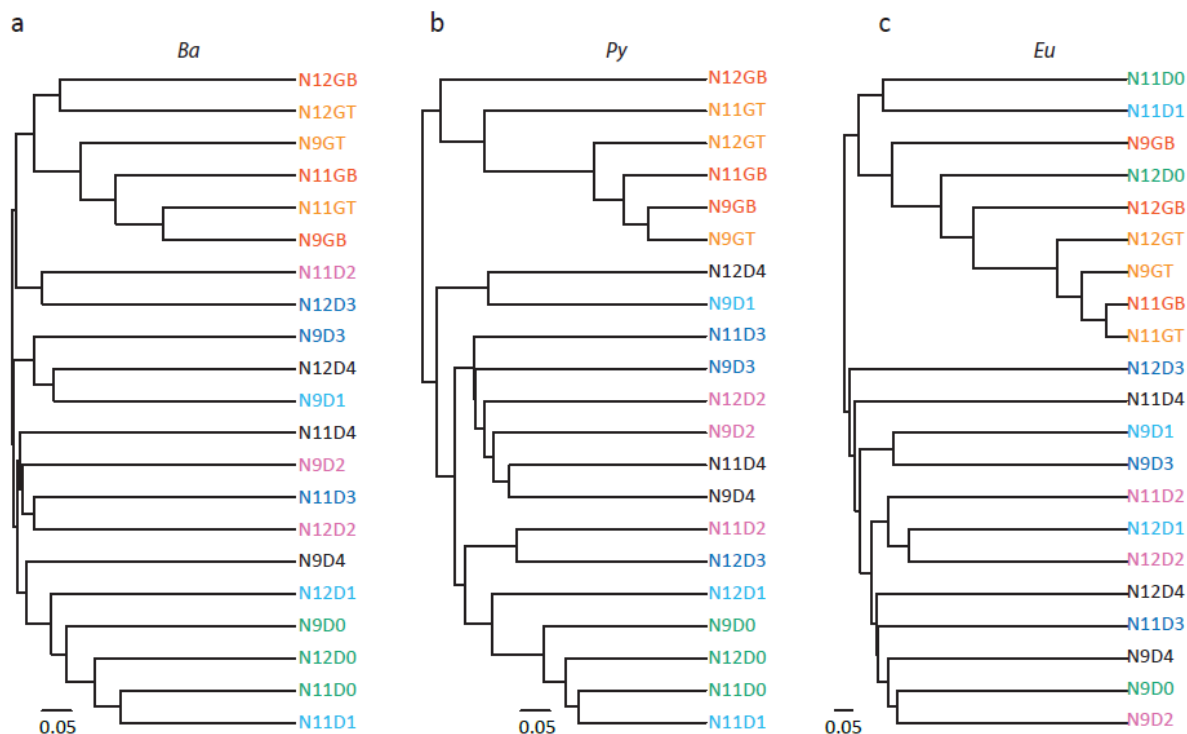
Storey JD. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics* **31**:2013–2035.

Suen G, Scott JJ, Aylward FO, Adams SM, Tringe SG, Pinto-Tomás AA, *et al.* (2010). An insect herbivore microbiome with high plant biomass-degrading capacity. *PLoS Genet* **6**:e1001129.

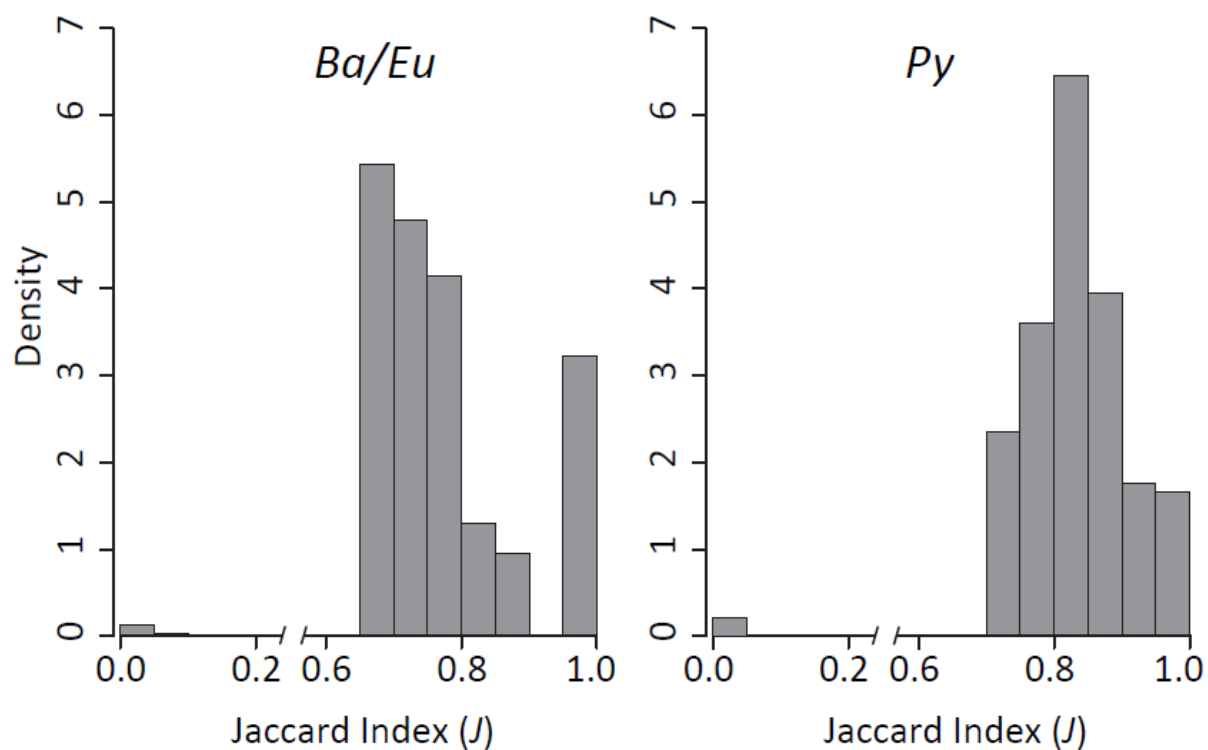
Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, *et al.* (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560–565.

Watts DJ, Strogatz SH. (1998). Collective dynamics of small-world networks. *Nature* **393**:440–442.

White JR, Nagarajan N, Pop M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5:e1000352.

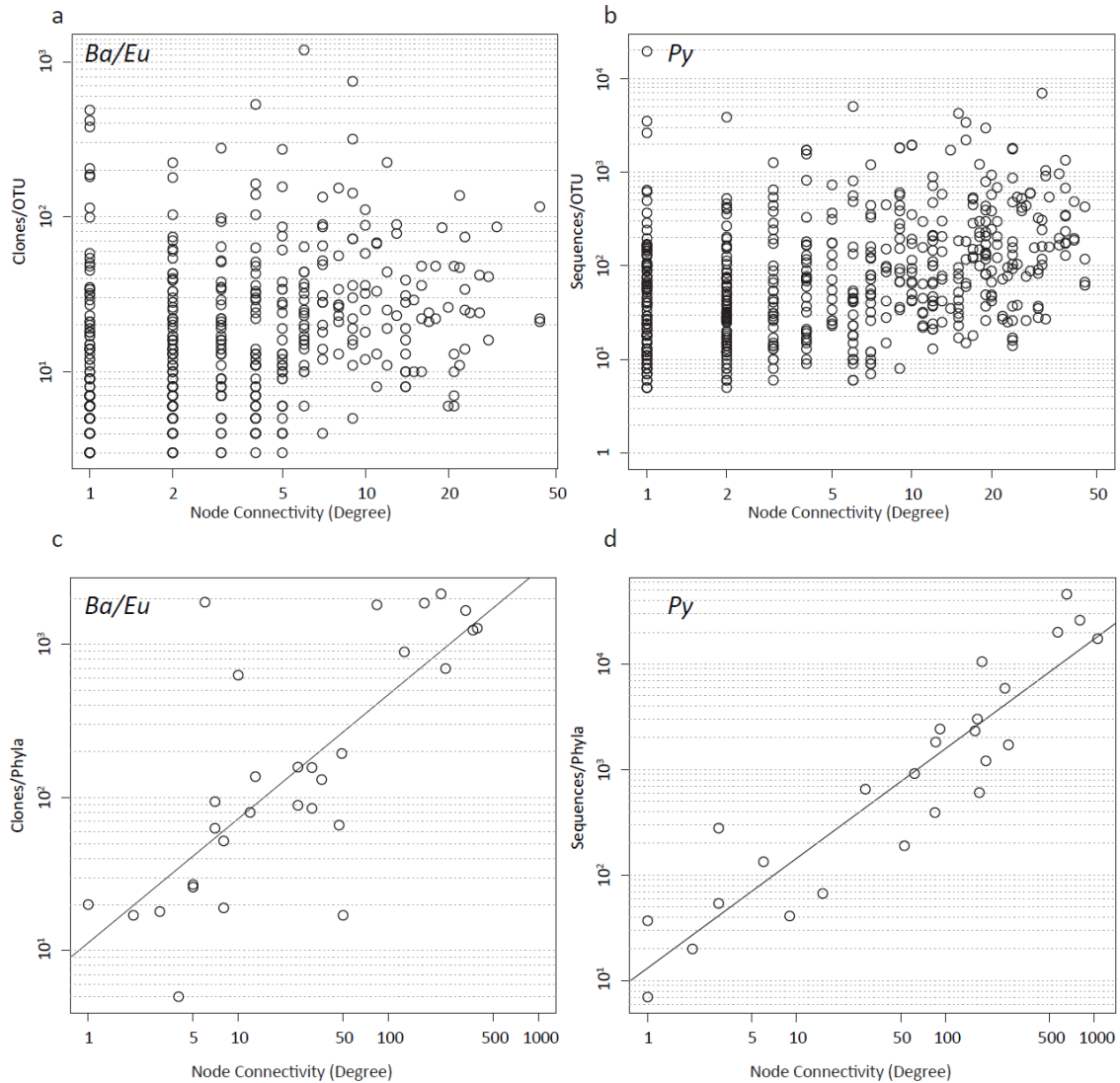


**Supplementary Figure A7.1. Individual clustering analyses of garden and dump samples.** Sample clustering is based on the Bray-Curtis similarity coefficient for the **a)** 16S clone libraries (*Ba*), **b)** 16S pyrotag libraries (*Py*), and **c)** 18S clone libraries (*Eu*). Sample names indicate the nest of origin followed by dump stratum (D) or garden sample (G). Colors indicate the layer of garden or dump from which each sample was collected.

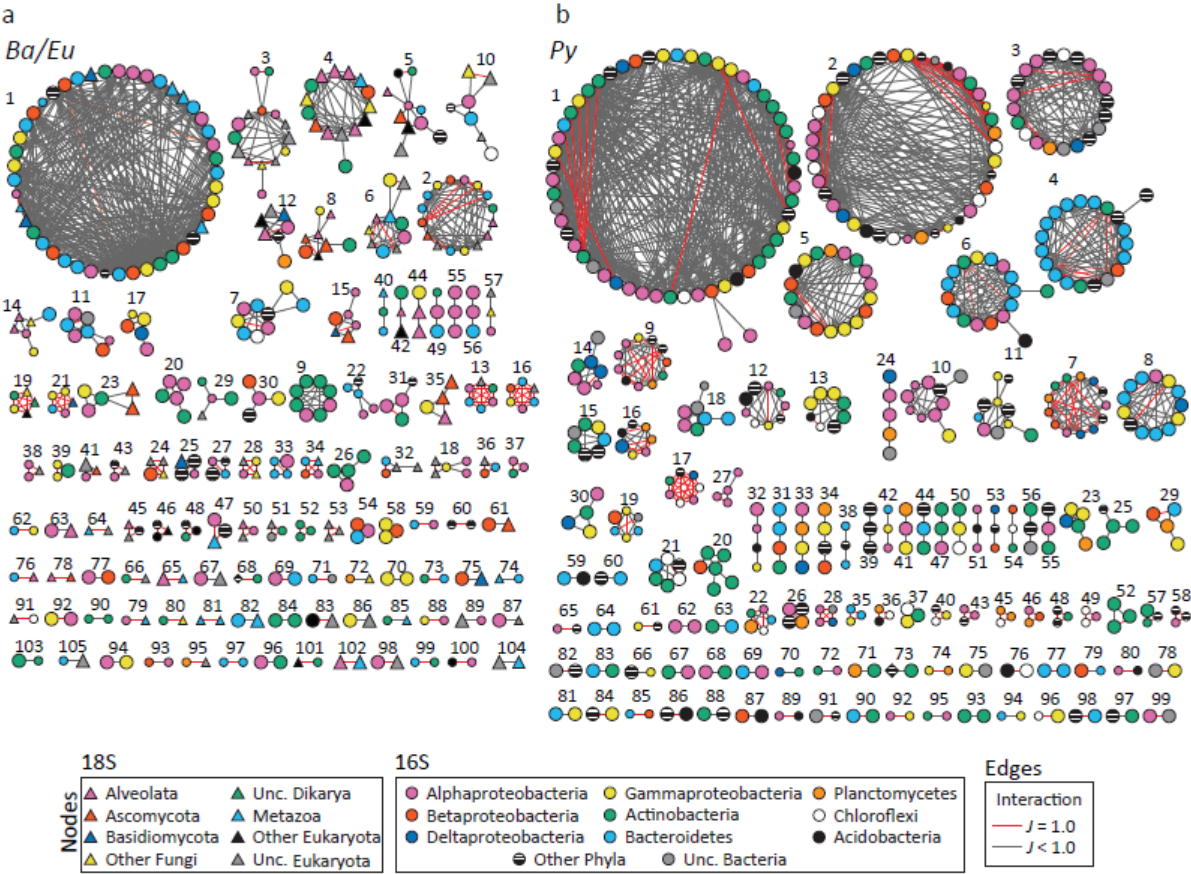


**Supplementary Figure A7.2. Frequency of co-occurrence association values.** Distribution of specificity scores between significant pairwise OTU associations from the *Ba/Eu* (left) and *Py* (right) datasets based on the *Jaccard* Index. Values of 1.0 indicate OTU pairs that always co-occurred while 0.0 indicates OTUs that always occurred separately.

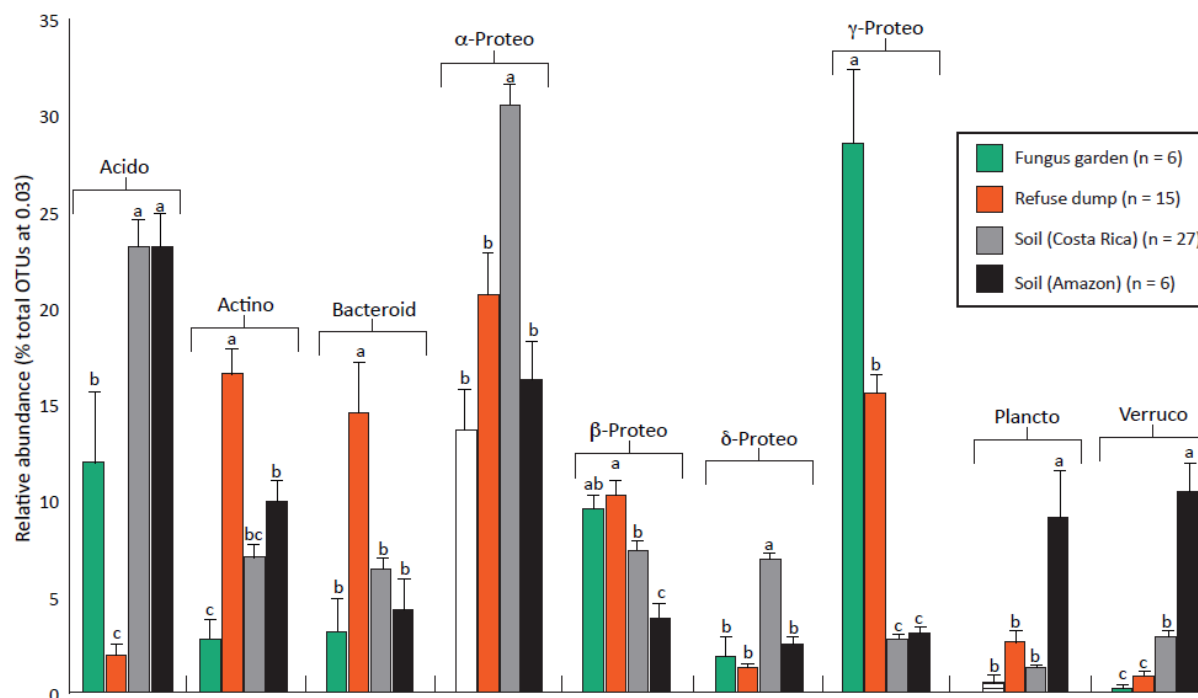




**Supplementary Figure A7.3. Assessment of the number of clones per OTU and node connectivity (degree) for each OTU from the a) *Ba/Eu* and b) *Py* co-occurrence networks.** In both cases, we did not observe a significant correlation between number of clones and number of connections. The same data but with OTUs grouped by phyla revealed a significant correlation between the number of clones/phyla and the number of connections for both the c) *Ba/Eu* and d) *Py* co-occurrence networks. Points in a and b represent individual OTUs and in c and d points represent OTUs grouped by phyla.

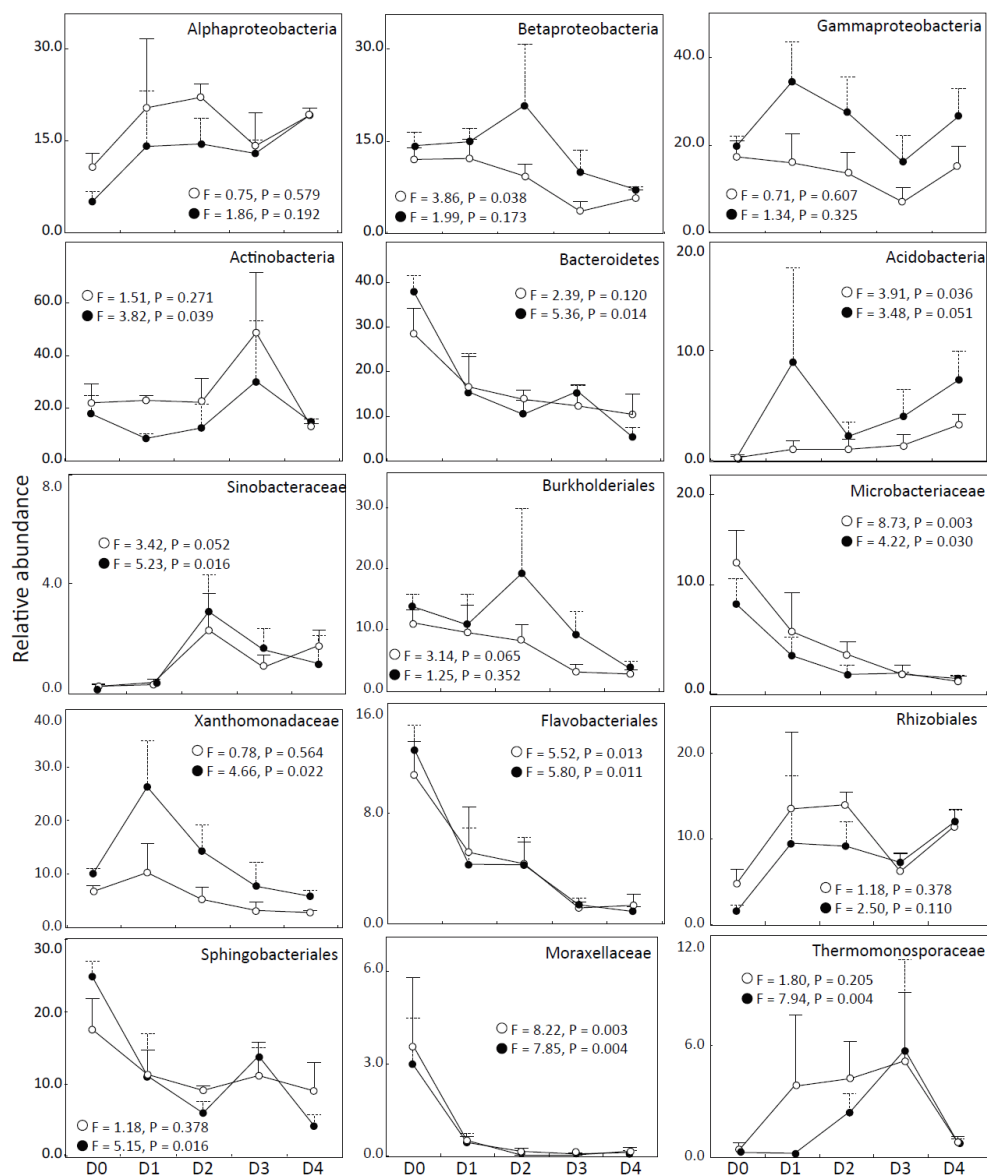


**Supplementary Figure A7.4. Results of MCL analyses of co-occurrence networks.** All modules of co-existing OTUs identified from the co-occurrence network in Figure A7.4, clustered using Markov Clustering Algorithm (MCL) based on similarity coefficients. Node shape and color denote taxonomic group. Larger nodes indicate OTUs shared across all three nests. Edge color denotes level of interaction according to the *Jaccard* Index. Values correspond to cluster number (for a list of OTUs, cluster numbers and associated details, see Supplementary Table A7.3). See Figure A7.4 for modules containing OTUs that were both enriched in a given habitat and found across all three nests. See also Supplementary Table A7.3 for a description of OTUs from each cluster.

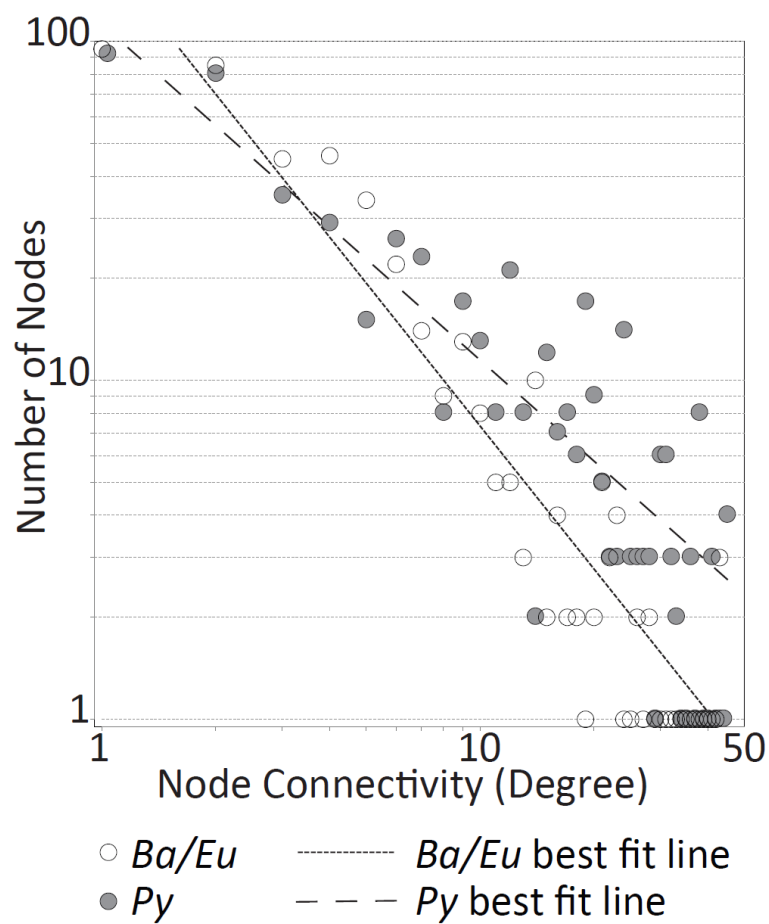


**Supplementary Figure A7.5. Taxonomic composition of fungus gardens, refuse dumps, and tropical soils.**

Here, we show the relative contribution of nine dominant phyla/subphyla (> 5% total OTU richness) at 97% OTU cutoff from the refuse dumps pyrotag dataset (*Py*) generated as part of this study as well as publically available datasets from leaf-cutter ant fungus gardens (Suen *et al.*, 2010), Costa Rican soil (Nemergut *et al.*, 2010), and Amazonian soils (Navarrete *et al.*, 2010). Columns represent group averages (SD). Groups not connected by the same letter (a–c) are significantly different ( $p < 0.05$ ) as calculated using the Tukey–Kramer post hoc analysis on one-way ANOVA. Abbreviations: Acido (Acidobacteria); Actino (Actinobacteria); Bacteroid (Bacteroidetes); α-Proteo (Alphaproteobacteria); β-Proteo (Betaproteobacteria); δ-Proteo (Deltaproteobacteria); γ-Proteo (Gammaproteobacteria); Plancto (Planctomycetes); Verruco (Verrucomicrobia). Values in parentheses indicate the number of individual samples generated for each study.



**Supplementary Figure A7.6. Assessment of taxonomic changes through dumps.** Relationship between relative abundance of 15 major bacterial clades and dump depth. Samples were pooled based on stratum of origin. Points represent mean (SD) for each refuse layer (D0–D4, horizontal axis). Open circles represent the full-length dataset (*Ba*) and filled circles represent the pyrotag dataset (*Py*). Solid error bars correspond to the full-length set and dashed error bars correspond to the pyrotag set. Significant changes were assessed using a quasibinomial generalized linear model (F) and subsequent ANOVA analysis (P). See supplementary methods.



**Supplementary Figure A7.7. Degree distribution from co-occurrence networks.** Connectivity degree distribution of nodes (OTUs) from the *Ba/Eu* (open circles) and the *Py* (closed circles) co-occurrence networks plotted on a log-log scale. Parameters were fitted with a power law (*Ba/Eu*, short dashes; *Py*, long dashes) to determine if the networks exhibited scale-free properties

**Supplementary Table A7.1.** Summary of taxonomic distribution (sequences). Total sequences by taxonomy for each sample in the 16S clone libraries, 16S pyrotag libraries, and 18S clone libraries. Phyla level taxa are highlighted in red followed by deeper levels of taxonomic resolution. Values indicate the number of sequences associated with a given taxonomic level. “Root” indicates the total number of sequences for a given sample. Separate **Excel** spreadsheet available on request.

**Supplementary Table A7.2.** Summary of taxonomic distribution (OTUs). Total OTUs (97%) by taxonomy for each sample in the 16S clone libraries, 16S pyrotag libraries, and 18S clone libraries. Phyla level taxa are highlighted in red followed by deeper levels of taxonomic resolution. Values indicate the number of OTUs associated with a given taxonomic level. “Root” indicates the total number of OTUs for a given sample. Separate **Excel** spreadsheet available on request.

**Supplementary Table A7.3.** Full results of the Metastats, co-occurrence, and clustering analyses. Summary information for each OTU in the clone library (*Ba/Eu*) and pyrotag (*Py*) co-occurrence networks. Column labels denote the following information for each OTU: **Library**, source library; **Dataset**, source dataset; **Size**, number of clones; **Proportion**, ratio of **Size** to total number of network sequences; **Metastats Analysis**, based on results of METASTATS analysis (**Enriched** indicates whether an OTU was significantly more abundant in the upper dump strata [DT], the lower dump strata [DB], gardens [G], or not enriched [NS]; **mean**, mean number of reads in each habitat; **stderr**, standard deviation of number of reads in each habitat; **q-value**, false discovery rate); **Degree**, number of connections in co-occurrence analyses; **MCL cluster**, associated cluster from Markov Clustering (MCL) analysis based on *Jaccard* similarity coefficient; **Shared**, indicates whether an OTU was shared across all three nests; **Taxa/Taxonomy**, describes the associated taxonomic lineage. Separate **Excel** spreadsheet available on request.

**Supplementary Table A7.4.** Summary of taxon distribution in the co-occurrence analysis (*Ba/Eu*)

| Domain           | Phylum                | Full dataset <sup>1</sup> |               | Co-occurrence analysis <sup>2</sup> |               |          |
|------------------|-----------------------|---------------------------|---------------|-------------------------------------|---------------|----------|
|                  |                       | No. of OTUs               | No. of clones | No. of OTUs                         | No. of clones | <i>k</i> |
| Bacteria/Archaea | Acidobacteria         | 95                        | 263           | 5                                   | 52            | 8        |
|                  | Actinobacteria        | 575                       | 3641          | 56                                  | 1667          | 328      |
|                  | Bacteroidetes         | 346                       | 2189          | 55                                  | 1240          | 367      |
|                  | Chloroflexi           | 123                       | 288           | 3                                   | 80            | 12       |
|                  | Crenarchaeota         | 44                        | 823           | 2                                   | 630           | 10       |
|                  | Firmicutes            | 72                        | 225           | 3                                   | 94            | 7        |
|                  | Gemmatimonadetes      | 60                        | 308           | 5                                   | 89            | 25       |
|                  | Planctomycetes        | 154                       | 282           | 3                                   | 26            | 5        |
|                  | Proteobacteria        |                           |               |                                     |               |          |
|                  | Alpha                 | 602                       | 2831          | 82                                  | 1277          | 392      |
|                  | Beta                  | 148                       | 1228          | 26                                  | 696           | 241      |
|                  | Delta                 | 167                       | 390           | 1                                   | 27            | 5        |
|                  | Gamma                 | 336                       | 4392          | 37                                  | 2140          | 225      |
|                  | SBR1093               | 2                         | 18            | 1                                   | 17            | 2        |
|                  | SC4                   | 8                         | 32            | 2                                   | 18            | 3        |
|                  | TM7                   | 90                        | 254           | 7                                   | 66            | 47       |
|                  | Verrucomicrobia       | 89                        | 222           | 3                                   | 63            | 7        |
|                  | Unclassified          | 92                        | 147           | 2                                   | 19            | 8        |
| Eukaryota        | Alveolata             | 285                       | 1443          | 26                                  | 894           | 128      |
|                  | Amoebozoa             | 14                        | 31            | 1                                   | 1890          | 6        |
|                  | Arthropoda            | 289                       | 3601          | 8                                   | 17            | 50       |
|                  | Nematoda              | 61                        | 325           | 5                                   | 157           | 31       |
|                  | Unclassified          |                           |               |                                     |               |          |
|                  | Metazoa               | 46                        | 198           | 6                                   | 137           | 13       |
|                  | Ascomycota            | 223                       | 2241          | 13                                  | 1816          | 84       |
|                  | Basidiomycota         | 222                       | 5466          | 7                                   | 194           | 49       |
|                  | Basal fungal lineages | 39                        | 276           | 1                                   | 20            | 1        |
|                  | Glomeromycetes        | 24                        | 211           | 4                                   | 158           | 25       |
|                  | Unclassified          | 30                        | 52            | 1                                   | 5             | 4        |
|                  | Dikarya               |                           |               |                                     |               |          |
|                  | Unclassified Fungi    | 97                        | 325           | 7                                   | 131           | 36       |
|                  | stramenopiles         | 76                        | 202           | 7                                   | 85            | 31       |
|                  | Unclassified          |                           |               |                                     |               |          |
|                  | Eukaryota             | 1284                      | 4244          | 49                                  | 1864          | 174      |
| Total            |                       | 5693                      | 36 148        | 428                                 | 15 569        | 2324     |

<sup>1</sup>Distribution of OTUs and total clones prior to filtering (*i.e.* removing OTUs present in less than 3 samples).<sup>2</sup>Distribution of OTUs, total clones, and connectivity after the co-occurrence analysis.

**Supplementary Table A7.5.** Summary of taxon distribution in the co-occurrence analysis (Py)

| Domain           | Phylum           | Full dataset <sup>1</sup> |              | Co-occurrence analysis <sup>2</sup> |              |          |
|------------------|------------------|---------------------------|--------------|-------------------------------------|--------------|----------|
|                  |                  | No. of OTUs               | No. of reads | No. of OTUs                         | No. of reads | <i>k</i> |
| Bacteria/Archaea | Acidobacteria    | 337                       | 24 211       | 20                                  | 10 539       | 176      |
|                  | Actinobacteria   | 892                       | 41 332       | 82                                  | 26 064       | 800      |
|                  | Bacteroidetes    | 599                       | 33 089       | 62                                  | 20 056       | 568      |
|                  | BRC1             | 5                         | 183          | 1                                   | 134          | 6        |
|                  | Chlamydiae       | 95                        | 833          | 7                                   | 392          | 85       |
|                  | Chloroflexi      | 417                       | 3033         | 25                                  | 1717         | 265      |
|                  | Crenarchaeota    | 8                         | 76           | 1                                   | 37           | 1        |
|                  | Fibrobacteres    | 9                         | 205          | 2                                   | 190          | 53       |
|                  | Firmicutes       | 332                       | 5675         | 13                                  | 1825         | 86       |
|                  | Gemmatimonadetes | 83                        | 3995         | 17                                  | 2427         | 92       |
|                  | Nitrospirae      | 12                        | 295          | 2                                   | 279          | 3        |
|                  | NKB19            | 3                         | 26           | 1                                   | 20           | 2        |
|                  | Planctomycetes   | 467                       | 2989         | 23                                  | 1211         | 187      |
|                  | Proteobacteria   |                           |              |                                     |              |          |
|                  | Alpha            | 1052                      | 38 293       | 101                                 | 17 406       | 1054     |
|                  | Beta             | 472                       | 37 552       | 26                                  | 5899         | 251      |
|                  | Delta            | 353                       | 5022         | 21                                  | 3018         | 164      |
|                  | Gamma            | 1436                      | 99 417       | 61                                  | 45 915       | 653      |
|                  | Unclassified     | 349                       | 2010         | 5                                   | 917          | 62       |
|                  | Spirochaetes     | 53                        | 1485         | 6                                   | 606          | 169      |
|                  | TG3              | 8                         | 82           | 2                                   | 67           | 15       |
|                  | Thermi           | 16                        | 139          | 1                                   | 54           | 3        |
|                  | TM6              | 37                        | 253          | 1                                   | 7            | 1        |
|                  | TM7              | 48                        | 135          | 3                                   | 41           | 9        |
|                  | Verrucomicrobia  | 105                       | 951          | 7                                   | 653          | 29       |
|                  | Unclassified     | 781                       | 6663         | 22                                  | 2331         | 158      |
| Total            |                  | 7969                      | 307 944      | 512                                 | 141 805      | 4892     |

<sup>1</sup>Distribution of OTUs and total reads prior to filtering (*i.e.* removing OTUs present in less than 3 samples).

<sup>2</sup>Distribution of significant OTUs, total reads, and connectivity after the co-occurrence analysis.



**Supplementary Table A7.6.** Topological parameters of the *Ba/Eu* and *Py* co-occurrence networks

| Dataset      | Actual <sup>1</sup> |       |                          |                       |                       |            | Randomized <sup>2</sup> |                     |          |
|--------------|---------------------|-------|--------------------------|-----------------------|-----------------------|------------|-------------------------|---------------------|----------|
|              | $n^3$               | $m^4$ | $\langle \ell \rangle^5$ | $\langle C \rangle^6$ | $\langle k \rangle^7$ | $\gamma^8$ | $\langle \ell \rangle$  | $\langle C \rangle$ | $\gamma$ |
| <i>Ba/Eu</i> | 428                 | 1162  | 5.18                     | .445                  | 5.43                  | 1.40       | 3.76                    | 0.012               | 0.692    |
| <i>Py</i>    | 512                 | 2446  | 4.74                     | .439                  | 9.56                  | 1.00       | 3.00                    | 0.019               | 0.012    |

<sup>1</sup>Results from the real co-occurrence networks and <sup>2</sup>from the randomized networks of identical dimension ( $n$  and  $m$ ).

<sup>3</sup>Total number nodes (OTUs).

<sup>4</sup>Total number of edges (significant connections).

<sup>5</sup>The mean shortest path length or geodesic distance between any two nodes.

<sup>6</sup>The clustering coefficient, a measure triangle density in a network.

<sup>7</sup>The average degree  $k$ , connectivity of the network defined as  $\langle k \rangle = 2m/n$ .

<sup>8</sup>The exponential value from power law distribution equation,  $P_{\text{deg}}(k) \propto k^{-\gamma}$ .

**Supplementary Table A7.7.** Total, unique, and shared OTUs from refuse dumps at various definitions for the full-length (*Ba*) and pyrotag (*Py*) datasets.

| Dataset   | OTU Cutoff          | Total OTUs <sup>1</sup> |           |      | Unique OTUs <sup>2</sup> |      |      | Shared OTUs <sup>3</sup> |           |            |              |
|-----------|---------------------|-------------------------|-----------|------|--------------------------|------|------|--------------------------|-----------|------------|--------------|
|           |                     | N9                      | N11       | N12  | N9                       | N11  | N12  | N9<br>N11                | N9<br>N12 | N11<br>N12 | All<br>Nests |
| <i>Ba</i> | unique <sup>4</sup> | 5441                    | 3161      | 3638 | 5270                     | 3014 | 3474 | 61                       | 78        | 54         | 32           |
|           | 99                  | 2575                    | 1529      | 1339 | 2055                     | 1111 | 893  | 174                      | 202       | 100        | 144          |
|           | 98                  | 2075                    | 1321      | 1105 | 1514                     | 857  | 645  | 191                      | 187       | 90         | 184          |
|           | 97                  | 1738                    | 1130      | 935  | 1192                     | 675  | 503  | 189                      | 166       | 75         | 191          |
|           | 95                  | 1232                    | 853       | 670  | 777                      | 468  | 320  | 156                      | 121       | 51         | 178          |
| <i>Py</i> | unique              | 16<br>351               | 11<br>780 | 9649 | 13<br>064                | 8846 | 7354 | 1475                     | 836       | 483        | 976          |
|           | 99                  | 10<br>689               | 8250      | 6466 | 7350                     | 5325 | 4248 | 1428                     | 719       | 460        | 1033         |
|           | 98                  | 6675                    | 5136      | 3919 | 4304                     | 2909 | 2289 | 1041                     | 444       | 300        | 886          |
|           | 97                  | 4532                    | 3462      | 2533 | 2715                     | 1777 | 1306 | 778                      | 320       | 188        | 719          |
|           | 95                  | 2453                    | 1925      | 1353 | 1308                     | 844  | 579  | 467                      | 160       | 96         | 518          |

<sup>1</sup>Indicates the number of OTUs present in each refuse dump.

<sup>2</sup>Total number of OTUs unique to a specific dump.

<sup>3</sup>Total OTUs shared between dumps from colonies: N9 and N11; N9 and N12; N11 and N12; and across all three nests. Values from the 'All' column only include OTU shared across all three dumps. <sup>4</sup>Less than 0.005% sequence dissimilarity.

**Supplementary Table A7.8.** Taxonomic identity of bacterial OTUs shared across all *A. colombica* three nests from the *Ba* and *Py* datasets at 99% OTU cutoff.

| Phylum                           | Class               | full-length             |                           | pyrotag    |             |
|----------------------------------|---------------------|-------------------------|---------------------------|------------|-------------|
|                                  |                     | Total OTUs <sup>1</sup> | Total clones <sup>2</sup> | Total OTUs | Total reads |
| Acidobacteria                    | Acidobacteria Grp 1 | –                       | –                         | 2          | 83          |
|                                  | Acidobacteria Grp 2 | –                       | –                         | 1          | 72          |
|                                  | Acidobacteria Grp 3 | –                       | –                         | 1          | 107         |
|                                  | Acidobacteria Grp 4 | 1                       | 17                        | 3          | 631         |
|                                  | Acidobacteria Grp 5 | –                       | –                         | 5          | 178         |
|                                  | Acidobacteria Grp 6 | –                       | –                         | 11         | 812         |
|                                  | Unc. Acido          | –                       | –                         | 7          | 930         |
| Actinobacteria                   | Acidimicrobiales    | –                       | –                         | 17         | 603         |
|                                  | Actinomycetales     | 31                      | 569                       | 192        | 30 382      |
|                                  | Solirubrobacterales | –                       | –                         | 15         | 1496        |
|                                  | Unc. Actino         | –                       | –                         | 10         | 327         |
| Bacteroidetes                    | Bacteroidia         | –                       | –                         | 7          | 217         |
|                                  | Flavobacteria       | 11                      | 314                       | 43         | 10 093      |
|                                  | Sphingobacteria     | 15                      | 429                       | 94         | 26 105      |
| Chlamydiae <sup>†</sup>          | Chlamydiae          | –                       | –                         | 7          | 217         |
| Chloroflexi <sup>††</sup>        | Anaerolineae        | –                       | –                         | 2          | 117         |
|                                  | Caldilineae         | 1                       | 27                        | 2          | 181         |
|                                  | Chloroflexi         | –                       | –                         | 1          | 23          |
|                                  | Thermomicrobia      | 1                       | 47                        | 4          | 284         |
|                                  | Unc. Chloroflexi    | –                       | –                         | 8          | 237         |
| Crenarchaeota <sup>†</sup>       | Thermoprotei        | 1                       | 115                       | 2          | 44          |
| Deinococcus–Thermus <sup>†</sup> | Deinococci          | –                       | –                         | 3          | 884         |
| Firmicutes                       | Bacilli             | 2                       | 78                        | 29         | 2425        |
|                                  | Clostridia          | –                       | –                         | 5          | 69          |
|                                  | Unc. Firmicutes     | –                       | –                         | 2          | 43          |
| Gemmatimonadetes <sup>††</sup>   | Gemmatimonadetes    | 5                       | 40                        | 5          | 1294        |
| Nitrospira <sup>†</sup>          | Nitrospira          | –                       | –                         | 1          | 10          |
| Planctomycetes                   | Planctomycetacia    | –                       | –                         | 18         | 599         |
| Proteobacteria                   | α-Proteobacteria    | 31                      | 408                       | 185        | 19 850      |
|                                  | β-Proteobacteria    | 20                      | 395                       | 106        | 24 770      |
|                                  | δ-Proteobacteria    | –                       | –                         | 12         | 1586        |
|                                  | γ-Proteobacteria    | 20                      | 206                       | 136        | 46 983      |
|                                  | Unc. Proteo         | –                       | –                         | 22         | 2250        |
|                                  | Spirochaetes        | –                       | –                         | 13         | 552         |
| Spirochaetes <sup>†</sup>        | Spirochaetes        | –                       | –                         | 13         | 552         |
| TM7 <sup>†</sup>                 | Unc. TM7            | 2                       | 15                        | 1          | 23          |
| Verrucomicrobia <sup>†</sup>     | Opitutae            | –                       | –                         | 4          | 250         |
|                                  | Subdivision 3       | –                       | –                         | 2          | 59          |
|                                  | Verrucomicrobiae    | –                       | –                         | 2          | 162         |
| Unc. Bacteria                    |                     | 3                       | 16                        | 53         | 8820        |
| Total                            |                     | 144                     | 2676                      | 1033       | 183 768     |

<sup>1</sup>Total number of OTUs shared across all three dumps for a given taxa.<sup>2</sup>Number of clones associated with shared OTUs.<sup>†</sup>Denotes taxa present as < 0.5% total clones (pyrotag dataset).<sup>††</sup>Denotes taxa present as < 1% total clones (pyrotag dataset). Dash indicates that OTUs from those taxa that were not shared across dumps and/or detected. We identified a diverse group of microbes found in all three dumps, including many rare representatives.

**Supplementary Table A7.9.** Number (and percentage) of unique OTUs from Supplementary Table A7.8 present as singletons or doubletons for the full-length (*Ba*) and pyrotag (*Py*) 16S datasets at various OTU definitions.

| Dataset   | OTU<br>cutoff | Nest         |            |            |
|-----------|---------------|--------------|------------|------------|
|           |               | N9           | N11        | N12        |
| <i>Ba</i> | unique        | 4777 (90%)   | 2715 (90%) | 3047 (87%) |
|           | 99            | 1747 (85%)   | 1000 (90%) | 750 (84%)  |
|           | 98            | 1241 (82%)   | 771 (90%)  | 516 (80%)  |
|           | 97            | 972 (81%)    | 606 (89%)  | 396 (78%)  |
|           | 95            | 623 (80%)    | 429 (91%)  | 246 (77%)  |
| <i>Py</i> | unique        | 11 545 (88%) | 8080 (91%) | 6584 (89%) |
|           | 99            | 6321 (86%)   | 4739 (89%) | 3653 (86%) |
|           | 98            | 3572 (83%)   | 2553 (87%) | 1854 (81%) |
|           | 97            | 2183 (80%)   | 1520 (85%) | 1028 (78%) |
|           | 95            | 971 (74%)    | 682 (80%)  | 424 (73%)  |

**Supplementary Table A7.10.** Total OTUs shared between all layers (D0–D4) within each dump from the full-length (Ba) and pyrotag (Py) datasets at various OTU definitions.

| Distance | <i>Ba</i>   |                         |             |            |             |            | <i>Py</i>   |            |             |            |             |            |
|----------|-------------|-------------------------|-------------|------------|-------------|------------|-------------|------------|-------------|------------|-------------|------------|
|          | N9          |                         | N11         |            | N12         |            | N9          |            | N11         |            | N12         |            |
|          | No.<br>OTUs | %<br>total <sup>1</sup> | No.<br>OTUs | %<br>total | No.<br>OTUs | %<br>total | No.<br>OTUs | %<br>total | No.<br>OTUs | %<br>total | No.<br>OTUs | %<br>total |
| unique   | 0           | 0.00                    | 0           | 0.00       | 0           | 0.00       | 77          | 0.47       | 71          | 0.60       | 9           | 0.09       |
| 99       | 0           | 0.00                    | 1           | 0.07       | 0           | 0.00       | 90          | 0.84       | 77          | 0.93       | 12          | 0.19       |
| 98       | 1           | 0.05                    | 1           | 0.08       | 0           | 0.00       | 101         | 1.51       | 83          | 1.62       | 19          | 0.48       |
| 97       | 3           | 0.17                    | 4           | 0.35       | 1           | 0.11       | 113         | 2.49       | 76          | 2.20       | 20          | 0.79       |
| 95       | 5           | 0.41                    | 13          | 1.52       | 2           | 0.30       | 119         | 4.85       | 75          | 3.90       | 23          | 1.70       |

<sup>1</sup>Percent total indicates the proportion of total OTUs for a given nest that were shared across all strata within that nest. Few OTUs were shared across all strata from individual refuse dumps. Even at more relaxed OTU definitions, the proportion of shared diversity to total diversity was relatively low.