

DISSERTATION

PUTTING MICROBIAL POLYPHENOL METABOLISM ON THE MAP:  
USING MICROBIOME SCIENCE TO REVISE SOIL CHEMICAL PARADIGMS

Submitted by

Bridget B. McGivern

Department of Soil and Crop Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2022

Doctoral Committee:

Advisor: Kelly C. Wrighton

Ann Hagerman  
Thomas Borch  
Jessica Prenni  
Michael J. Wilkins

Copyright by Bridget B. McGivern 2022

All Rights Reserved

## ABSTRACT

### PUTTING MICROBIAL POLYPHENOL METABOLISM ON THE MAP: USING MICROBIOME SCIENCE TO REVISE SOIL CHEMICAL PARADIGMS

Polyphenols are chemically diverse natural products found in virtually all higher plants. Polyphenols are important modulators of ecosystems, for example in the human gut and in terrestrial systems, with these functionalities governed by the microbiome. However, there are contradictory views regarding microbiome-polyphenol interactions. For example, in the gut, microbiota anoxically degrade or transform dietary polyphenols to release bioactive metabolites that confer health benefits to the host. In soils, while microbial communities can use polyphenols under oxic conditions, anoxia is thought to prevent their biodegradation. The enzyme latch hypothesis describes how accumulated polyphenols preserve soil carbon in anoxic systems via inhibition of a key enzyme, polyphenol oxidase. The overarching aim of this dissertation was to elucidate the microbial metabolism of polyphenols under anoxic conditions.

In Chapter 1, I summarize the prevailing views about polyphenols in anoxic soils. In particular, I focus on the enzyme latch and the existing knowledge gaps surrounding microbial polyphenol metabolism. My review of the literature revealed that the controversial enzyme latch theory is largely supported by low-resolution methods, and that the current paradigms do not fit with what we know about microbial metabolism in gut ecosystems.

In Chapter 2, I tested a basic premise of the Enzyme Latch: that soil microbial communities cannot degrade polyphenols under anoxic conditions. To test this, I conducted a laboratory study where anoxic microcosms were constructed using soil from a wetland. This experiment had three microcosms treatments: (i) live soils amended with a model polyphenol, a condensed tannin (CT), (ii) live soils unamended as a control, and (iii) autoclaved soil amended with CT to account for potential abiotic reactions between the CT and soil matrix. To describe the microbial interactions with the CT, over 20 days these microcosms were tracked with genome-resolved metaproteomes and a variety of metabolomics approaches. The metabolite data provided chemical evidence that the CT was depolymerized, and ultimately degraded to small compounds. Importantly, comparison of the live and autoclaved metabolomes suggested that abiotic processes contributed to CT degradation. Overlaying metaproteome data, it was inferred that 3 bacterial taxa were involved in this degradation. One, a *Kosakonia*, likely depolymerized the CT while the other two, a *Holophaga* and a novel lineage in the Sporomusales, degraded the CT monomer and degradation products. Analysis of the overall microbial community functions showed that CT addition did not restrict carbon or nitrogen cycling as current soil theory assumes, and in fact may have enhanced them. Chapter two of my thesis provided a proof-of-concept that soil microbial communities degrade polyphenols under anoxic conditions, in contrast to the enzyme latch.

Building from Chapter 2, I next investigated what is known about microbial polyphenol metabolism across all systems. Chapter 3 describes a computational tool I designed for annotating polyphenol metabolism in microbial genomes, Curated

Annotations for Microbial Polyphenol Enzymes and Reactions (CAMPER). Chapter 3 provides the rationale for this tool, the installation and usage, and the outputs. This is a publicly available tool, and my hope is that it makes it easier for the broader scientific community to track polyphenol metabolism in their own datasets.

Armed with CAMPER and with the evidence from Chapter 2 that soil microbes can use polyphenols under anoxic conditions, I went on to investigate microbial polyphenol metabolism at the ecosystem level. In Chapter 4, I couple genome-resolved metatranscriptomes, metabolomes, and geochemistry from an arctic peatland to describe polyphenol metabolism across this site. First, I refute the major assumptions of the Enzyme Latch using multiple methods. Then, I apply CAMPER to the microbial genomes and metatranscriptomes to identify the active polyphenol metabolizers in the site, supporting findings with metabolite data. Finally, I showed the metatranscriptome expression of CAMPER modules can predict important geochemistry. Overall, this work suggests polyphenols must be considered as potential carbon sources in terrestrial systems, regardless of the redox conditions in soils.

The final chapter of this dissertation (Chapter 5) summarizes the key findings from the lab and field investigations of microbial polyphenol metabolism and offers perspective on other systems where polyphenol metabolism should be studied. In summary, the aims of this dissertation were to summarize the views of microbial polyphenol metabolism and the enzyme latch (Chapter 1), to investigate these in the lab and in the field (Chapter 2-4), and to summarize how this dissertation has revised our knowledge of microbial polyphenol metabolism (Chapter 5). Ultimately, this dissertation

shows microbial polyphenol metabolism is a diverse, dynamic, and active part of soil carbon cycles.

## ACKNOWLEDGEMENTS

The challenging science in this dissertation was compounded by a university move, a global pandemic, and multiple “unprecedented” headline news events. I have made it through it all entirely because I have excellent mentors, collaborators, friends, and family.

First, thank you to **Dr. Kelly Wrighton** for letting me join your lab, and taking me on your journey West. You have been an amazing mentor and teacher to me, and while I have so many things to thank you for, for now I will say: Thank you for sharing your wisdom, for celebrating the big and small moments, for advocating for me and my science, and for giving me the space pursue my “gut feelings”. I am so glad I have you in my corner.

I would also like to thank **Dr. Ann Hagerman** for serving as a co-advisor, mentor, and committee member to me. Thank you for promoting me and my science to your scientific community, and for giving your time to meet with me. I appreciate your advice and support.

I am grateful to the rest of my committee, **Dr. Jessica Prenni**, **Dr. Mike Wilkins**, and **Dr. Thomas Borch**, for their time and guidance throughout my time at Colorado State University.

The science within this dissertation was wholly enabled by awesome collaborators. I want to specifically thank **Dr. Malak Tfaily** for serving as a great mentor and teammate.

Thank you to my wonderful **W2 lab community**, past and present, for providing a supportive and fun place to do science. A special thank you to **Reb Daly** for training me in the lab, and for all your management and organizational efforts that enabled me to work without interruptions. Thank you to **Dr. Mikayla Borton** for training me, and for always being down for a happy hour or a potluck. Thank you to **Josue Rodriguez-Ramos** for your friendship and comradery over the years. Thank you to **Kaela Amundson** for a being a dear friend and steadfast listener, always available for a rant.

Thank you to **Andrea Westlie** for being a best friend. I can't quite put into words how important you've been to me over the last 4 years, but suffice it to say the best part of moving to Colorado was getting to be near you again. I hope fate puts us in the same place again someday!

Thank you to **Dillon Janda** for filling my life with love over the last few years. You and **Penelope** have shown me that play is just as important as work.

Most of all, thank you to my parents **Barb** and **Gene McGivern**, and my brothers **George** and **Peter**. You are the best cheerleaders and having your support and guidance has sustained me through my PhD program.

Finally, thank you to the staff and experts at **CSU Counseling Services** and the **Kendall Reagan Nutrition Center** for the critical care and support over the years. It takes a whole person to do science and having access to these services was critical in enabling me to thrive.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	vi
Chapter 1: Introduction.....	1
1.1 <i>Polyphenols</i> .....	1
1.2 <i>Microbial interactions with polyphenols</i> .....	2
1.3 <i>The enzyme latch and quadruple lock</i> .....	3
1.4 <i>A microbiome-centric view of the enzyme latch</i> .....	5
Chapter 1 Figures .....	7
Chapter 1 References .....	8
Chapter 2 – Decrypting bacterial polyphenol metabolism in an anoxic wetland soil .....	13
2.1 <i>Summary</i> .....	13
2.2 <i>Introduction</i> .....	13
2.3 <i>Results</i> .....	16
2.3.1 Establishing laboratory microcosms to explore polyphenol fate in anoxic soils .....	16
2.3.2 Metabolomic evidence supports abiotic and biotic polyphenol degradation ..	17
2.3.3 Genome-resolved metaproteomics reveals enrichment of polyphenol responsive microbes .....	24
2.3.4 Polyphenol biodegradation occurs through metabolic exchange in anoxic soil .....	28
2.3.5 Anoxic soil carbon cycling is resistant to polyphenol amendment .....	33
2.4 <i>Discussion</i> .....	36
2.5 <i>Methods</i> .....	38
2.5.1 Soil sample Collection.....	38
2.5.2 Condensed Tannin purification .....	39
2.5.3 Reactor design and set up .....	40
2.5.4 16S rRNA Gene Analyses .....	42
2.5.5 Metagenomic Sequence and Assembly.....	43
2.5.6 Metaproteomic Extraction and Spectral Analysis.....	44
2.5.7 Metaproteomic database creation and analyses .....	46
2.5.8 Integrated metabolomic approaches.....	48
2.5.9 FT-ICRMS analysis.....	50
2.5.10 LC-MS metabolomic analysis .....	51

2.5.11 NMR metabolomic analysis .....	52
Chapter 2 Tables.....	53
Chapter 2 Figures .....	55
Chapter 2 References .....	78
Chapter 3 – CAMPER: Curated Annotations for Microbial Polyphenol Enzymes and Reactions .....	92
3.1 Overview.....	92
3.2 CAMPER Data.....	93
3.3 Using CAMPER with DRAM .....	94
3.4 CAMPER standalone tool: CAMPER_DRAMKit.....	95
3.4.1 Set up with Conda.....	95
3.4.2 Setup with pip .....	96
3.4.3 Installing with DRAM.....	96
3.4.4 Using CAMPER_DRAMKit.....	96
3.4.5 Other Tools and flags.....	98
3.5 I just want to run your BLAST and HMM searches on my own! .....	99
3.6 CAMPER Outputs.....	99
3.6.1 Raw annotations .....	100
3.6.2 Distillate .....	100
Chapter 3 Figures .....	102
Chapter 4 – Auditing the enzyme latch unveils polyphenols as unbudgeted microbial carbon substrates in northern peatlands.....	105
4.1 Summary .....	105
4.2 Introduction.....	105
4.3 Enzyme latch disproved with multiple methods.....	107
4.4 Stordalen Mire polyphenols are diverse and habitat specific .....	108
4.5 Diverse lineages express polyphenol metabolizing gene in Stordalen Mire microbiomes .....	109
4.6 Microbial polyphenol metabolism is active across Stordalen Mire .....	111
4.7 Active polyphenol metabolism by undescribed lineages .....	113
4.8 Polyphenol metabolism predicts porewater carbon dioxide .....	114
4.9 Conclusion.....	116
4.10 Methods.....	117
4.10.1 Field Site and Sampling .....	117

4.10.2 Field Geochemical Analyses.....	117
4.10.3 Metagenome Assembled Genome Taxonomy, Phylogeny, and Annotation .....	118
4.10.4 RNA Extraction .....	118
4.10.5 Metatranscriptome Analysis.....	119
4.10.6 Enzyme Assays .....	120
4.10.7 Folin-Ciocalteu Assay .....	121
4.10.8 Organic Matter metabolite Extraction.....	122
4.10.9 LC-MS.....	122
4.10.10 FT-ICRMS Analysis.....	124
4.10.11 Refinement of Tannin and Lignin Van Krevlen classes .....	125
4.10.12 Nuclear Magnetic Resonance (NMR) Analysis.....	125
4.10.13 CAMPER Construction .....	126
4.10.14 Statistics.....	127
 Chapter 4 Figures .....	 129
Chapter 4 References .....	135
 Chapter 5 – Conclusion.....	 141
5.1 Summary .....	141
5.2 Future Research Directions.....	141
5.2.1 Polyphenol metabolism in agricultural soils .....	141
5.2.2 Polyphenol metabolism in ruminants .....	142
 Chapter 5 References .....	 143
 Appendices .....	 144
<i>Appendix A: Chapter 2 Supplementary Text</i> .....	144
<i>Appendix B: Chapter 2 Supplementary Files</i> .....	155
<i>Appendix C: Chapter 3 Supplementary Files</i> .....	156

## Chapter 1: Introduction

### **1.1 Polyphenols**

Polyphenols are a diverse class of plant natural products that are important across ecosystems. Historically, polyphenols were defined by their biochemical roles as antioxidants, and more recently have been structurally defined as compounds containing more than one phenolic ring<sup>1</sup>. However, in the scientific literature and in practice, this definition has been relaxed to include compounds with a single phenolic ring bearing more than one hydroxyl group. There are more than 10,000 known structures for polyphenols following this broader definition. Here, we will consider polyphenols following phenol-explorer database classifications<sup>2</sup>, spanning 5 families: the flavonoids, the lignans, the stilbenes, the phenolic acids, and the broad “other” family. These structures can be modified by glycosylation or methoxylation, or polymerized into macromolecules like lignin and hydrolysable or condensed tannins (**Figure 1.1**).

While recent studies have shown that fungi<sup>3,4</sup> and bacteria<sup>5,6</sup> can produce polyphenols, plants are likely the predominant source of these compounds. Plants produce polyphenols for a variety of purposes, for example in the case of lignin, they can serve structural roles<sup>7</sup>. Some plants produce and exude flavonoids to induce nitrogen-fixing soil microorganisms to form root nodules<sup>8</sup>. They are also suggested to protect plants against pathogens<sup>9</sup> and herbivores<sup>10</sup>. Therefore, ecosystems affected by plants and plant biomass are likely impacted by polyphenols.

## **1.2 Microbial interactions with polyphenols**

Polyphenols represent a large and diverse carbon pool. Estimates from food data suggest humans consume over 1g polyphenols daily<sup>11</sup>, while in soils concentrations can be found up to 100mg/g soil<sup>12</sup>. Owing to their abundance, the ways that microbiomes in these systems interact with polyphenols, and how these interactions influence the overall system, is of interest. These interactions can be classified as either microbial inhibiting or benefitting.

Some polyphenols are antimicrobials towards specific microorganisms, with several potential modes of action. One commonly inferred mechanism is cell membrane disruption, wrecking cell structure<sup>13</sup>. Besides being directly toxic, polyphenols are also thought to indirectly impede microbial metabolism. One example of this is chelation of metals needed as cofactors<sup>14</sup>, while it is also suggested that polyphenols bind extracellular organic matter like proteins or carbohydrates, preventing access to these substrates<sup>15,16</sup>. By binding proteins, they can also bind and inactivate extracellular enzymes<sup>17,18</sup>. It's important to note that these claims of antimicrobial effects largely originate from *in vitro* experiments on clinically relevant microorganisms<sup>14</sup>. Further, it has been noted that polyphenols exhibit heterogeneous and unpredictable inhibition trends across compounds and organisms<sup>19</sup>. Therefore, it is not accurate to assume polyphenols will always be inhibitory to microbial communities.

Conversely, polyphenols are also thought to serve as microbial carbon and or energy sources. Substantial attention has been paid to microbial lignin degradation in the context of biofuel development<sup>20,21</sup>. However, most knowledge of microbial polyphenol metabolism is derived from gut microbial isolates. This research targets the

health benefits associated with human gut microbiota processing of dietary polyphenols<sup>22</sup>. For instance, soy products contain the polyphenolic isoflavone daidzin. Roughly 30% and 60% of Western and Asian populations, respectively, possess gut microbes capable of converting daidzin to equol, a metabolite proposed to be protective against hormone-based diseases<sup>23</sup>. Illustrating the importance of understanding microbial polyphenol metabolism, the increased relative abundance of daidzin-converting gut microbes in Asian populations relative to Western populations mirrors the relative decreased incidence of hormone-based diseases<sup>23</sup>. It should be noted that while examples like daidzin metabolism are appreciated, given the chemical diversity of polyphenols there is still much to learn about microbial polyphenol metabolism across systems.

Polyphenols enter soil systems from decaying leaves and bark, from nuts and fruits, and via rain throughfall in the canopy. In oxic soils, polyphenols are regarded as microbial carbon and energy sources, however under anoxic conditions, these same compounds are thought to be “recalcitrant” to degradation<sup>24</sup>. This idea directly contradicts knowledge from gut systems that include examples of microbes metabolizing polyphenols under anoxic conditions. Despite knowledge in gut ecosystems, this idea has been propagated in the soil literature for decades<sup>24</sup> by a theory called the enzyme latch.

### ***1.3 The enzyme latch and quadruple lock***

The enzyme latch is a biogeochemical theory that originated to explain why peatlands, and specifically bogs, store substantial amounts of carbon. In this theory, it was assumed that anaerobic conditions in peatlands limit the activity of the enzyme

polyphenol oxidase (PPO), leading to accumulated polyphenols binding hydrolase enzymes, ultimately limiting decomposition<sup>24,25</sup>. These ideas have been expanded beyond bogs and wetlands into systems like mangrove forests<sup>26</sup>, and even into seagrass rhizospheres<sup>27</sup>.

While several studies have supported the enzyme latch model, others have challenged the findings. Specifically, conflicting support was found in some studies for relationships between polyphenols and respiration rates<sup>28</sup>, between oxygen availability and enzyme activity<sup>29</sup>, and between polyphenols and enzyme activity<sup>30,31</sup>. In response, the enzyme latch was refined in the 'quadruple lock' model to expand the impacts of polyphenols beyond enzymes to also include direct microbial toxicity, iron limitation, and substrate deprivation<sup>32</sup>.

Critically, the development of and support for the enzyme latch and quadruple lock hypotheses has largely relied on low-resolution geochemical assays, masking complexity in polyphenol content and microbial activity. For example, polyphenol content has traditionally been inferred through the Folin-Ciocalteu assay, which is not specific for polyphenols but instead measures the total oxidizable substrates<sup>33,34</sup>. When considering the chemical diversity of polyphenols, and the diversity of other oxidizable substrates in soils, the need for methods that specifically examine polyphenol identity and content is clear. Beyond this, soil enzyme activity assays are noisy and hard to interpret<sup>35</sup>. In addition, from a microbial perspective, soil enzyme assays are not representative of all carbon metabolic strategies. Furthermore, the negative correlation of polyphenols with carbon dioxide (CO<sub>2</sub>) emissions in recent studies<sup>32</sup> may not be due to causality, as there are many other biotic and abiotic interactions in soils that

contribute to CO<sub>2</sub> production. Thus, the existing body of literature calls into question whether the enzyme latch is supported at all, and if it is, then mechanisms underpinning the latch and their ecosystem consequences warrant further clarification.

Finally, the assumption that polyphenol oxidase is the only enzyme capable of degrading polyphenols in soils is not compatible with knowledge of polyphenol metabolism in other systems. Many examples from the human gut microbiome demonstrate metabolic routes for polyphenol turnover in anoxic systems, with potential for these metabolic functions to include diverse ecosystems including soils<sup>36</sup>. Collectively, the methods and lack of microbial knowledge behind the enzyme latch have created blind spots for how microbial communities are perceived to act with polyphenols in terrestrial systems. This is not inconsequential, as the authors behind the enzyme latch are now suggesting polyphenol supplementation in climatically-vulnerable systems to promote carbon sequestration<sup>37</sup>. Therefore, understanding the microbial interactions with polyphenols in anoxic soil systems is critical to predict if these compounds indeed lock carbon in soils, or are active participants in nutrient cycling in soils.

#### ***1.4 A microbiome-centric view of the enzyme latch***

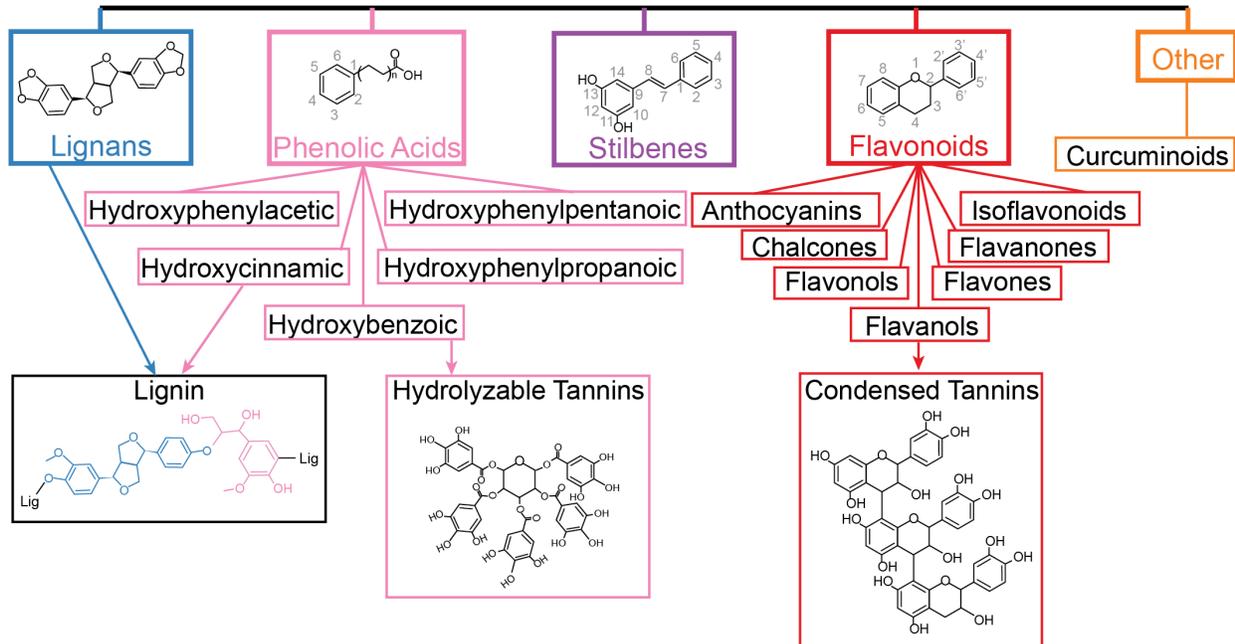
I used multi-omic methods across laboratory and field systems to interrogate the microbial assumptions of the enzyme latch. At the heart of this work, I address the microbial metabolism of polyphenols under anoxic conditions. Specifically, the three objectives of my dissertation research were to:

1. Test the assumption that soil microbial communities cannot degrade polyphenols under anoxic conditions (Chapter 2)

2. Identify microbial polyphenol metabolic enzymes and develop a tool to annotate these genes in microbial genomic data (Chapter 3)
3. Examine the enzyme latch and describe microbial polyphenol metabolism in an arctic peatland (Chapter 4)

Together, this work reframes the paradigm that polyphenols are recalcitrant to soil microbial communities and provides groundwork for future research to identify the biogeochemical impacts of microbial polyphenol metabolism.

## Chapter 1 Figures



**Figure 1.1** The structural diversity of polyphenols. The five major families of polyphenols are shown, followed by their subfamilies. Grey numbers indicate ring positions. These structures can be polymerized to form lignin, hydrolyzable tannins, or condensed tannins, which are shown in boxes.

## Chapter 1 References

1. Quideau, S., Deffieux, D., Douat-Casassus, C. & Pouységu, L. Plant polyphenols: Chemical properties, biological activities, and synthesis. *Angew. Chemie - Int. Ed.* **50**, 586–621 (2011).
2. Rothwell, J. A. *et al.* Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database* **2013**, bat070–bat070 (2013).
3. Zhang, W. *et al.* Priming of rhizobial nodulation signaling in the mycosphere accelerates nodulation of legume hosts. *New Phytol.* (2022).  
doi:10.1111/nph.18192
4. Juvvadi, P. R., Seshime, Y. & Kitamoto, K. Genomics reveals traces of fungal phenylpropanoid-flavonoid metabolic pathway in the filamentous fungus *Aspergillus oryzae*. *J. Microbiol.* **43**, 475–86 (2005).
5. Wang, R.-J. *et al.* Three New Isoflavonoid Glycosides from the Mangrove-Derived Actinomycete *Micromonospora aurantiaca* 110B. *Mar. Drugs* **17**, 294 (2019).
6. Kang, H. R. *et al.* Termisoflavones A–C, Isoflavonoid Glycosides from Termite-Associated *Streptomyces* sp. RB1. *J. Nat. Prod.* **79**, 3072–3078 (2016).
7. Vanholme, R., Demedts, B., Morreel, K., Ralph, J. & Boerjan, W. Lignin Biosynthesis and Structure. *Plant Physiol.* **153**, 895–905 (2010).
8. Peters, N. K., Frost, J. W. & Long, S. R. A Plant Flavone, Luteolin, Induces Expression of *Rhizobium meliloti* Nodulation Genes. *Science (80-. )*. **233**, 977–980 (1986).

9. Wallis, C. M. & Galarneau, E. R.-A. Phenolic Compound Induction in Plant-Microbe and Plant-Insect Interactions: A Meta-Analysis. *Front. Plant Sci.* **11**, (2020).
10. Coley, P. D., Bryant, J. P. & Chapin, F. S. Resource Availability and Plant Antiherbivore Defense. *Science (80-. )*. **230**, 895–899 (1985).
11. Huang, Q., Braffett, B. H., Simmens, S. J., Young, H. A. & Ogden, C. L. Dietary Polyphenol Intake in US Adults and 10-Year Trends: 2007-2016. *J. Acad. Nutr. Diet.* **120**, 1821–1833 (2020).
12. Fierer, N., Schimel, J. P., Cates, R. G. & Zou, J. Influence of balsam poplar tannin fractions on carbon and nitrogen dynamics in Alaskan taiga floodplain soils. *Soil Biol. Biochem.* **33**, 1827–1839 (2001).
13. Cushnie, T. P. T., Hamilton, V. E. S., Chapman, D. G., Taylor, P. W. & Lamb, A. J. Aggregation of *Staphylococcus aureus* following treatment with the antibacterial flavonol galangin. *J. Appl. Microbiol.* **103**, 1562–1567 (2007).
14. Daglia, M. Polyphenols as antimicrobial agents. *Curr. Opin. Biotechnol.* **23**, 174–181 (2012).
15. Adamczyk, B., Sietiö, O. M., Biasi, C. & Heinonsalo, J. Interaction between tannins and fungal necromass stabilizes fungal residues in boreal forest soils. *New Phytol.* **223**, 16–21 (2019).
16. Jakobek, L. Interactions of polyphenols with carbohydrates, lipids and proteins. *Food Chem.* **175**, 556–567 (2015).
17. Haslam, E. Polyphenol–protein interactions. *Biochem. J.* **139**, 285–288 (1974).
18. Mcdougall, G. J. & Stewart, D. The inhibitory effects of berry polyphenols on

- digestive enzymes. *BioFactors* **23**, 189–195 (2005).
19. Bouarab-Chibane, L. *et al.* Antibacterial Properties of Polyphenols: Characterization and QSAR (Quantitative Structure–Activity Relationship) Models. *Front. Microbiol.* **10**, (2019).
  20. Bugg, T. D. H., Ahmad, M., Hardiman, E. M. & Rahmanpour, R. Pathways for degradation of lignin in bacteria and fungi. *Nat. Prod. Rep.* **28**, 1883–1896 (2011).
  21. Janusz, G. *et al.* Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol. Rev.* **049**, 941–962 (2017).
  22. Goris, T., Cuadrat, R. R. C. & Braune, A. Flavonoid-Modifying Capabilities of the Human Gut Microbiome—An In Silico Study. *Nutrients* **13**, 2688 (2021).
  23. Rafii, F. The Role of Colonic Bacteria in the Metabolism of the Natural Isoflavone Daidzin to Equol. *Metabolites* **5**, 56–73 (2015).
  24. Freeman, C., Ostle, N. & Kang, H. An enzymic “latch” on a global carbon store: A shortage of oxygen locks up carbon in peatlands by restraining a single enzymes. *Nature* **409**, 149 (2001).
  25. Fenner, N. & Freeman, C. Drought-induced carbon loss in peatlands. *Nat. Geosci.* **4**, 895–900 (2011).
  26. Kim, J. *et al.* Microbial decomposition of soil organic matter determined by edaphic characteristics of mangrove forests in East Asia. *Sci. Total Environ.* **763**, 142972 (2021).
  27. Sogin, E. M. *et al.* Sugars dominate the seagrass rhizosphere. *Nat. Ecol. Evol.* (2022). doi:10.1038/s41559-022-01740-z
  28. Brouns, K., Keuskamp, J. A., Potkamp, G., Verhoeven, J. T. A. & Hefting, M. M.

- Peat origin and land use effects on microbial activity, respiration dynamics and exo-enzyme activities in drained peat soils in the Netherlands. *Soil Biol. Biochem.* **95**, 144–155 (2016).
29. Hall, S. J., Treffkorn, J. & Silver, W. L. Breaking the enzymatic latch: Impacts of reducing conditions on hydrolytic enzyme activity in tropical forest soils. *Ecology* **95**, 2936–2945 (2014).
  30. Romanowicz, K. J. *et al.* Understanding drivers of peatland extracellular enzyme activity in the PEATcosm experiment: mixed evidence for enzymic latch hypothesis. *Plant Soil* **397**, 371–386 (2015).
  31. Urbanová, Z. & Hájek, T. Revisiting the concept of ‘enzymic latch’ on carbon in peatlands. *Sci. Total Environ.* **779**, 146384 (2021).
  32. Fenner, N. & Freeman, C. Woody litter protects peat carbon stocks during drought. *Nat. Clim. Chang.* **10**, 363–369 (2020).
  33. Everette, J. D. *et al.* Thorough study of reactivity of various compound classes toward the folin-Ciocalteu reagent. *J. Agric. Food Chem.* **58**, 8139–8144 (2010).
  34. Sánchez-Rangel, J. C., Benavides, J., Heredia, J. B., Cisneros-Zevallos, L. & Jacobo-Velázquez, D. A. The Folin-Ciocalteu assay revisited: Improvement of its specificity for total phenolic content determination. *Anal. Methods* **5**, 5990–5999 (2013).
  35. Fierer, N., Wood, S. A. & Bueno de Mesquita, C. P. How microbes can, and cannot, be used to assess soil health. *Soil Biol. Biochem.* **153**, 108111 (2021).
  36. Borton, M. A. *et al.* Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. *Proc. Natl.*

*Acad. Sci. U. S. A.* **115**, E6585–E6594 (2018).

37. Alshehri, A. *et al.* A Potential Approach for Enhancing Carbon Sequestration During Peatland Restoration Using Low-Cost, Phenolic-Rich Biomass Supplements. *Front. Environ. Sci.* **8**, 48 (2020).

## **2.1 Summary**

Microorganisms play vital roles in modulating organic matter decomposition and nutrient cycling in soil ecosystems. The enzyme latch paradigm posits microbial degradation of polyphenols is hindered in anoxic peat leading to polyphenol accumulation, and consequently diminished microbial activity. This model assumes that polyphenols are microbially unavailable under anoxia, a supposition that has not been thoroughly investigated in any soil type. Here, we use anoxic soil reactors amended with and without a chemically defined polyphenol to test this hypothesis, employing metabolomics and genome-resolved metaproteomics to interrogate soil microbial polyphenol metabolism. Challenging the idea that polyphenols are not bioavailable under anoxia, we provide metabolite evidence that polyphenols are depolymerized, resulting in monomer accumulation, followed by the generation of small phenolic degradation products. Further, we show that soil microbiome function is maintained, and possibly enhanced, with polyphenol addition. In summary, this study provides chemical and enzymatic evidence that some soil microbiota can degrade polyphenols under anoxia and subvert the assumed polyphenol lock on soil microbial metabolism.

## **2.2 Introduction**

Polyphenols are one of the most abundant types of plant secondary metabolites. This prevalent chemical group is heterogeneous, consisting of over 10,000 structurally

---

<sup>1</sup> This chapter was reproduced verbatim from “McGivern, et al. Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nature Communications* (2021)”. The text benefitted from writing and editing contributions from contributing authors and reviewers selected by the publisher. The ordering of the materials in this dissertation are consistent with the content available online but have been renumbered to reflect incorporation into this dissertation.

divergent compounds<sup>1</sup>. These compounds are abundant in differing habitats: they enter the soil systems through litter decay or leaching<sup>2</sup>, while in gut systems these plant-derived metabolites are consumed in high concentrations from polyphenol-rich foods like berries and cocoa<sup>3</sup>. In the human gut, it is recognized that the gut microbiome plays an integral role in the anaerobic processing of dietary polyphenols to enable host absorption<sup>3</sup>. Similarly, in ruminants, microbial interactions with polyphenols in feed have ramifications for animal nutrition and husbandry<sup>4</sup>. Despite the prevalence and recognized importance of polyphenol compounds, the mechanisms underlying anaerobic microbial polyphenol metabolism are just being unveiled in gut systems<sup>5,6</sup>, and remain largely enigmatic in soil systems.

Despite knowledge from gut systems, in soils, and especially in polyphenol-rich peatlands<sup>7</sup>, it is widely assumed that microbial polyphenol degradation is an obligately aerobic metabolism, and thus cannot occur under anoxia. Consequently, the “enzyme latch”<sup>8,9</sup> hypothesis states that polyphenols accumulate under anoxic soil conditions and further control soil microbial carbon cycling as these compounds (1) are toxic to microorganisms, (2) inactivate microbial extracellular enzymes, and/or (3) bind substrates, thus depriving microorganisms of nutrients and limiting microbial activity<sup>10</sup> (**Figure 2.1A**). According to this model, polyphenols serve as a “lock”<sup>10</sup> to stabilize soil carbon in anoxic soils (**Figure 2.1B**). Based on these assumptions, it has been proposed that polyphenol amendment can be a tool for slowing rates of soil organic matter decomposition to mitigate carbon loss from peatlands<sup>7-9,11</sup>. However, the studies supporting these assertions in peat, or in any soil system, have not directly interrogated microbial metabolism in anoxic soils, instead inferring microbial community activity from

bulk level properties like respiration rate<sup>12–14</sup>, enzyme assays<sup>7–9</sup>, inferred biomass<sup>12,15</sup>, or cellular morphology<sup>10</sup>. These poorly defined interactions between soil microbiota and polyphenols must be elucidated to resolve the role of these compounds in soil carbon sequestration, especially in the face of changing climate.

Recent developments in genome sequencing technology coupled to improved computational methods make historically complex soil communities more tractable with multiple ‘omics approaches<sup>16–18</sup>. These methodological advances afford a renewed opportunity to discover the biochemistry underpinning microbial-polyphenol responses in soils<sup>16,17</sup>. Here our research goals include (i) investigating the possibility of microbially-mediated polyphenol transformations in anoxic soils, and (ii) determining the impact of polyphenols on overall microbial community function. To resolve these fates of polyphenols in anoxic soils, we use a structurally defined, model polyphenol substrate—a condensed tannin—as an amendment to our controlled, anoxic soil reactors. Periodically over 20 days, we probe this model soil microbiome with a variety of metabolomic methods and genome-resolved metaproteomics to discern the biotic and abiotic responses to the model polyphenol under anoxia. Collectively, our findings provide multi-omics evidence for polyphenol degradation and maintenance of overall microbial community function. These results represent a critical step in describing microbial polyphenol metabolism in an anoxic soil, refining the presumed metabolic roles of soil microbiota in long-held soil biogeochemical paradigms.

## 2.3 Results

### 2.3.1 Establishing laboratory microcosms to explore polyphenol fate in anoxic soils

To date many studies of the effects of polyphenols on soil microbiota have focused on boreal peat soils<sup>7,12</sup>. Yet in these soils, temperature (<20°C) was suggested to be a possible kinetic controller on microbial growth and enzyme activity, thus limiting polyphenol metabolism<sup>19</sup>. To extend these prior studies, we selected plant-covered, mineral soils from a microbially well-studied temperate, freshwater wetland<sup>18,20</sup>, thereby eliminating kinetic constraints and expanding our search for these metabolisms across a broader range of soil types. These wetland surface soils contained polyphenols (**Figure 2.2**) and have been shown to be tractable using multi-omics methods<sup>18,20</sup>, and thus were used as a model soil for evaluation of anaerobic polyphenol metabolism.

Using these surface soils as the inoculum, we amended anoxic laboratory microcosms with and without a model polyphenol. Owing to the known chemical heterogeneity among polyphenols, a structurally-characterized condensed tannin<sup>21</sup> (CT, **Figure 2.1B**) was selected as a model polyphenol substrate. CT are generally recognized as recalcitrant in diverse soils<sup>22,23</sup>, and were recently described as a significant inhibitor of microbial activity in a riparian peatland<sup>7</sup>. The CT polymer is comprised of oligomers of epicatechin with a terminal catechin unit, all of which are connected by interflavan bonds (**Figure 2.1B**). The average degree of polymerization is 16, yielding an average molecular weight of 4600 Da<sup>21</sup>. Reactors were amended with a CT loading of 375 mg CT/g soil, which is on par with reports of polyphenols in soils (up to 100 mg/g soil<sup>12</sup>), and consumption in the human diet (500 mg/day<sup>3</sup>). Importantly, our selected concentration exceeded the sorption limit for mineral soils (5-10 mg

polyphenol/g soil<sup>24</sup>), ensuring bioavailability for our microbially focused studies. From the triplicate, anoxic soil reactors, 16S rRNA genes, metabolites, and genome-resolved metaproteomes were sampled on days 1, 3, 7, 10, 14, and 20 (**Figure 2.1C**).

Our experimental design included two control treatments to (i) discern polyphenol-stimulated responses from native, background soil microbial activity and (ii) differentiate microbially-mediated CT degradation from abiotic CT degradation resulting from reactions with the soil matrix. First, to separate the impacts of polyphenols from background soil microbial processes, we performed parallel, temporal analyses on CT-amended and unamended control soil reactors (unamended control, -CT). Second, given that CT is known to abiotically react with components in soils<sup>22</sup>, we also amended autoclaved soil with CT (autoclaved soil, +CT). This latter control did not contain amplifiable DNA over the course of the experiment, supporting microbial-inactivation during the time course monitored in this treatment (**Appendix A**). While we recognize the potential for autoclaving to alter soil chemistry<sup>25</sup>, we show at inoculation there was little difference in the soil chemical landscape between autoclaved and unautoclaved CT-amended soil microcosms (**Figure 2.3**). Together these findings support the utility of autoclaved soils as a comparative metabolite control to identify microbial and soil abiotic transformations of the CT polymer. Collectively, this experimental design, analyzed with integrated high-resolution techniques, offered a new platform to resolve soil microbiota responses to polyphenols under anoxic conditions.

### *2.3.2 Metabolomic evidence supports abiotic and biotic polyphenol degradation*

Our primary goal was to monitor chemical transformations of a model polyphenol between active and inactive soil communities to discover evidence for microbiological

degradation products under anoxic conditions. Prior to this research, low-resolution chemical assays (e.g. Folin–Ciocalteu for polyphenols or acid butanol for CTs) were commonly used to assess polyphenol content in soils<sup>12</sup>. However, the Folin-Ciocalteu method suffers from a lack of specificity because it measures oxidizable substrates, such as polyphenols, but also including a variety of organic and inorganic constituents of biological systems including soils<sup>26,27</sup>. Further, the acid butanol method is highly specific for CT, but responds poorly to CT in soils or other complex matrices because of interfering interactions between CT and protein or particulates<sup>28,29</sup>. It is also not possible to detect structural changes to the CT polymer using the acid butanol assay, as for example it does not respond differentially based on degree of polymerization<sup>30</sup>. Beyond analytical methods, most earlier studies amended soils with crude mixtures of polyphenols (e.g. leaf extracts<sup>31</sup>) with these mixtures likely obscuring identification of polyphenol degradation products, while other studies lacked microbially-inactivated controls that likely prohibited clear assignment of degradation products to microbial processes. Here, we used a structurally-defined CT polymer (**Figure 2.1B**) and employed multiple control treatments, while using various high-resolution metabolomic techniques to track CT depolymerization and degradation products over time.

We first wanted to observe changes at the CT-oligomer level over time, with the temporal increase in smaller oligomers indicating depolymerization (i.e. interflavan bond cleavage, **Figure 2.4A**) of the larger CT polymer. Our Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry (FTICR-MS) analysis captured nearly 90,000 peaks across all samples that corresponded to compounds in a specific relatively high molecular weight mass range. Within these peaks, we developed a workflow that

identified peaks corresponding to CT oligomers and transformation products (**Figure 2.5**). We carried out subsequent Kendrick Mass Defect (KMD) analysis on these CT peaks using (epi)catechin as the base unit (described in **Figure 2.6**). This KMD analysis resolved CT oligomers and derived compounds without assigning chemical formulas (**Figure 2.4B**). Within a single KMD plot, the distribution of epi(catechin) oligomers and derived compounds ranging from monomers to hexamers were visualized (**Figure 2.4B**). Clouds of points were separated along a horizontal axis by oligomer size (mass). Importantly, Kendrick plots for multiply-charged polymers separate along a vertical axis based on an “isotopic split”<sup>32</sup>. In this phenomenon, Kendrick plots of polymers at charge state  $z$  exhibit  $z$  clear horizontal lines separated by  $1/z$  KMD<sup>32</sup> (**Figure 2.6**). Therefore, in our Kendrick plots, the two horizontal lines separated by  $\sim 0.5$  KMD indicated that we had two subpopulations of polymer oligomer species in our spectra with  $-1$  and  $-2$  charge. The primary horizontal line (KMD  $\sim 0$ ) corresponds to the singly- or doubly-charged  $^{12}\text{C}$  monoisotopic species, the singly-charged  $^{13}\text{C}$  species, and the doubly-charged  $^{13}\text{C}_{2,4,\dots}$  species. The separated peaks at  $\sim 0.5$  KMD represent the doubly-charged  $^{13}\text{C}_{1,3,5,\dots}$ -containing species (**Figure 2.6**). Synthesizing mass data and inferring  $-1$  or  $-2$  charge, we identified peaks on the Kendrick plot in oligomer size regions where points corresponding to oligomers and their transformation products (ex. degradation intermediates) can be found (**Figure 2.4B**, blue, purple, and pink rectangles).

The Kendrick plots for live and autoclaved CT-amended reactors show the presence of CT oligomers and transformation products ranging in size from monomer-hexamer at all timepoints. We confirmed the CT polymer did not contain detectable CT monomers or other flavonoids (e.g. quercetin) in its pure form prior to amendment

(**Figure 2.7**). Therefore, the appearance of monomer peaks on day zero in both live and autoclaved soil microcosms suggests either low levels of monomers were present in the soil sample, or that abiotic processes originating in the soil microcosm generated some monomers soon after adding the polymer to the soil. In looking at Kendrick plots generated for the live, unamended control soils (**Figure 2.8B**), we recovered negligible peaks that could be attributed to CT oligomers or monomers, thus supporting the likelihood that abiotic reactions between CT and the soil matrix, and not background soil CT concentrations, were sources for these compounds at day 0 in the CT-amended samples.

The Kendrick plots for biotic soils revealed a marked increase in the number of peaks corresponding to CT oligomers and transformation products (**Figure 2.8D**) at days 10 and 14. Although the autoclaved soil control reaches the same level of richness by day 20, the rate of CT transformation is enhanced in biotic microcosms. Further supporting this, in biotic soils there is an increase in smaller CT oligomers (e.g. CT tetramers, trimers, dimers, and monomers) over time and particularly at day 10, a trend not observed in the autoclaved soils until day 20 (**Figure 2.4B**, purple rectangles). Further supporting the accumulation of smaller oligomers (< 6-mer) in microbially-active soils, the peaks detected in the 0.5 KMD region increased in the biotic relative to autoclaved soils (**Figure 2.4B**). Peaks in this region likely derive from naturally-occurring  $^{13}\text{C}$ -containing compounds that are only detected when the parent  $^{12}\text{C}$  peaks from equivalent compounds are highly abundant<sup>33</sup> (**Figure 2.4B**, **Figure 2.6**). This latter finding further supported that biotic microcosms contained more CT depolymerization and transformation products (from monomers to hexamers) than the autoclaved control

(**Figure 2.4B**) at later time points, signifying that microbiota in the soils contributed to CT depolymerization. These microbially-enabled depolymerizations were in addition to abiotic transformations of the CT polymer that were observed in the autoclaved samples within a twenty-day period.

Given FTICR-MS indicated CT depolymerization to smaller oligomers and monomers over time in microbially-active soils, we tracked the production of CT phenolic monomers (e.g. epicatechin or catechin) and subsequent degradation of these monomers using liquid chromatography-mass spectrometry (LC-MS). The LC-MS data supported the FTICR-MS data, providing additional evidence for CT depolymerization, as the monomers were detected in the biotic and autoclaved soils at all timepoints (**Figure 2.4C**). Pairwise comparisons of the biotic and abiotic data indicated that C<sub>15</sub> monomers epicatechin and catechin were significantly enriched at day 10 in the biotic incubations (**Figure 2.4C**). In parallel, we also detected the C<sub>15</sub> flavonoid quercetin<sup>34</sup> in both treatments, but like the CT monomers it was also only significantly enriched in microbially active reactors at day 10, supporting the biotic production of this compound from (epi)catechin as others have postulated<sup>34</sup>. Thus, consistent with our FTICR-MS findings, we see an enrichment of C<sub>15</sub> monomers and close derivatives occurring in the microbially-active soils midway through the experiment (day 10). Together, our FTICR and LC mass spectrometry approaches contributed to a model where the interflavan bonds in the CT polymer were broken from a contribution of biotic and abiotic processes, yielding shorter CT oligomers and CT monomers catechin and epicatechin. This data contradicts the long-standing dogma in soils that the interflavan bonds linking monomers in the CT polymer are stable under anoxic conditions<sup>5,7,35</sup>.

Importantly, the C<sub>15</sub> flavonoids (epicatechin, catechin, or quercetin) decreased in abundance after day 10 only in microbially-active soils, suggestive of further biodegradation (**Figure 2.4E**). Based on our metabolite identifications in the CT-amended live and autoclaved soils, it is likely these flavonoids underwent heterocyclic C-ring fission (position 1 and 4 orange ring, **Figure 2.4C-D**) to generate a C<sub>6</sub> compound (phloroglucinol, blue ring) from the A-ring and C<sub>6</sub>-C<sub>3</sub> acid from the B-ring and C-ring carbon atoms (**Figure 2.4E**). The C<sub>6</sub>-C<sub>3</sub> acid can be envisioned as the parent (e.g. by loss of CO<sub>2</sub>) of phenylacetate derivatives (C<sub>6</sub>-C<sub>2</sub>, **Figure 2.4F**) and several putatively identified benzoic acids (C<sub>6</sub>-C<sub>1</sub>) and simple phenols (**Appendix B**)<sup>36</sup>.

With decreased abundance of C<sub>15</sub> flavonoids from microbial degradation, we observed a concomitant increase in many downstream phenolic metabolites in the microbially-active CT treated soils. Specifically, (i) four phenolic metabolites (C<sub>6</sub>-C<sub>3</sub> and C<sub>6</sub>-C<sub>2</sub>) were significantly enriched at multiple timepoints (**Figure 2.4E-F**), (ii) two C<sub>6</sub>-C<sub>1</sub> and C<sub>6</sub> metabolites were significantly enriched at day 20 (possibly 4-methylcatechol, hydroquinone; **Appendix B**), and (iii) another 3 phenolic metabolites were uniquely detected via NMR (C<sub>6</sub>-C<sub>3</sub> phenylpropionic acid, C<sub>6</sub>-C<sub>2</sub> 3,4-dihydroxyphenylacetic acid, C<sub>6</sub>-C<sub>2</sub> 3-hydroxyphenylacetic acid; **Appendix B**). The flavonoids and phenolic compounds identified by LC-MS and NMR had differing dynamics between our biotic and abiotic controls, indicative of unique production from microbial activity, and they were present in relatively negligible amounts in the unamended controls (**Figure 2.8**), further indicating that these products derived from the added CT. Therefore, accounting for differences between the biotic samples and both control reactors, we concluded that

the smaller phenolic compounds derived from microbial biodegradation of the added CT (**Figure 2.4A**).

Detection of some phenolic compounds could not be statistically resolved between microbially-active and autoclaved treatments, although they were generally less abundant in the autoclaved controls (**Figure 2.4E**). This result pointed to abiotic processes as additional transformers of CT monomers in anoxic soils. As the C<sub>15</sub> monomer was transformed, the LC-MS data suggested the 3,4-dihydroxylation pattern of the parent flavonoid compound B-ring was retained across biotic and autoclaved soils (**Figure 2.4E-F**). However, in microbially-active microcosms we also detected compounds with altered hydroxylation patterns, suggestive of distinctly biotic transformations: dehydroxylation yielding 3-hydroxy derivatives (**Figure 2.4E**), or rearrangement to yield a 2-hydroxy derivative<sup>37</sup> (**Figure 2.4F**). Also in these microbially-active soils, we detected a phenolic amine, 3,4-dihydroxyphenylalanine (DOPA), that was enriched significantly at later timepoints (phase 3) (**Figure 2.4E**). Collectively, this variety of phenolic metabolites detected in the later phases reinforced our hypothesis that while abiotic transformations of CT occurred in our anoxic soils, there were clear signals of microbial CT and monomer biodegradation that occurred on different time scales and yielded unique products.

Broadly, the fate of the CT polymer in microbially-active anoxic soils paralleled some polyphenolic transformations reported in mammalian fecal metabolomes<sup>38</sup>. We observed increased caffeic acid (**Figure 2.4E**) and putative dihydroxybenzoic acids (e.g. vanillic acid, **Appendix B**), which are suggested metabolite biomarkers<sup>39</sup> for anoxic polyphenol degradation in feces. Yet these proposed biomarkers were also detected in

our autoclaved CT-amended soils, further reinforcing the need to partition abiotic and biotic processes when working in chemically complex matrices like feces or soil. A more detailed analysis of the shared and unique features of polyphenol degradation in soils compared to the human gut is provided (**Figure 2.9, Appendix A**).

In summary, we provided chemical evidence for polyphenol degradation in soils under anoxic conditions. With support from multiple analytical methods, we concluded that CT likely underwent abiotic transformations, yet distinct increases in CT oligomers, monomers, and putative biodegradation products over time were detected only in microbially-active soils. This highly-resolved metabolite data provides a chemical framework for microbial polyphenol degradation in anoxic soils, a scaffolding that can be leveraged in future, more targeted, research using varied polyphenol substrates, as well as across a wider range of soil types and conditions.

### *2.3.3 Genome-resolved metaproteomics reveals enrichment of polyphenol responsive microbes*

In light of our metabolite data indicating active microbial polyphenol degradation, we next explored the impact of polyphenols on the soil microbiome. To uncover the key microbial players and functions underlying anoxic polyphenol responses in our soil reactors, we constructed a genome database composed of metagenome-assembled genomes (MAGs) from CT-amended and unamended samples at various timepoints. Specifically, metagenomic sequencing from the microcosms at days five, ten, and twenty were obtained (**Figure 2.1C**), totaling 500 Gbps sequencing (**Appendix B**). This sequencing depth represents 9-fold more sequencing per sample compared to published field wetland metagenome studies to date, thereby increasing the sensitivity

for detecting the breadth of microbial functions encoded in these soils<sup>40</sup> (**Figure 2.10**). From this sequencing, we assembled and reconstructed 294 MAGs, which were dereplicated at 99% average nucleotide identity into 155 MAGs (**Figure 2.11, Appendix B**), of which 87% were medium- and high-quality genomes<sup>41</sup> (**Figure 2.11, Appendix B**). Based on read mapping to this soil-derived MAG database, the majority (65%, n=101) of genomes were present across treatments. Despite the extensive depth of sequencing, 17% of genomes (n=26) were only recovered in non-CT reactors, while 18% (n=28) of genomes were only recovered from CT reactors (**Figure 2.12**). Importantly, of these CT-amendment specific MAGs, just 29% (n=8) were recovered at every sampled timepoint, highlighting the need for time-resolved metagenomes to capture community MAG composition in soil microcosms. The dereplicated MAG database evenly recruited metagenome reads across the samples, indicating there was little bias in assembly and binning due to treatment type (**Figure 2.13**). This dataset illustrated the value of targeted amendments, temporal sampling, and deep sequencing for bringing to light conditionally rare taxa that may have ecosystem-relevant metabolic capabilities<sup>42</sup>.

The dereplicated MAG database (n=155) contained genomes from 19 phyla, many of which represent the most abundant and cosmopolitan lineages in soils<sup>40</sup> (**Figure 2.11**). However, using the Genome Taxonomy Database toolkit (GTDB-tk)<sup>43</sup>, we found that a subset of our genomes represented newly sampled lineages (5 orders across 3 phyla), and a large proportion of our MAGs belonged to lineages defined only by alphanumeric identifiers in the GTDB at the class (17%), order (6%), or family (21%) levels (**Appendix B**). Further stressing the phylogenetic novelty in these soils, less than

1% of our soil microcosm 16S rRNA amplicon sequencing variants (ASVs) had similarity (>97%) to 16S rRNA genes represented in RefSoil<sup>44</sup> (a database of soil isolate genomes) (**Table 2.1**). The discrepancy between genomes uncovered in these soil microcosms and those included in public soil genome databases underscores the need for establishing study- or site-specific genome databases for uncovering cryptic biochemistry in soils.

To maximize the recovery of functions in our metaproteome analysis, we combined genes from all metagenomic assemblies, including binned genes from our MAG database and unbinned genes from metagenomic assemblies (**Figure 2.14**), to build a representative dereplicated (100% amino acid identity) gene database. Importantly, we verified changes in observed peptide recruitment derived from changes at the peptide level rather than a database effect (**Figure 2.13, Appendix A**). After mapping the metaproteomes obtained from CT-amended and unamended microcosms at six timepoints to our dereplicated gene database (n=36), we recovered 11,942 peptides that mapped to 50,446 potential proteins (**Appendix B**). From here, proteins were categorized into three groups based on if the peptides were unique to specific genomes (**Figure 2.14E**, categories detailed in **Methods**). Nearly 60% of the recovered peptides were uniquely recruited to 119 of 155 dereplicated MAGs (known as “binned uniques”, see Methods), enabling identification of active community members in our genome database over time (**Figure 2.15, Figure 2.16**). Notably, 47 MAGs recruited peptides exclusively in CT-microcosms, while just 3 MAGs were inferred to be active exclusively in unamended control soils. Alternatively, the remaining 69 MAGs recruited

peptides in both CT and unamended microcosms, hinting at the metabolic plasticity harbored in soils.

Mirroring trends in microbial 16S rRNA gene composition and exometabolite changes over time (**Figure 2.17**), metaproteomes of CT and unamended control microcosms diverged temporally (**Figure 2.15A, Appendix B**). The gene expression of members in the unamended control were relatively stable across the experimental period (i.e. no temporal clustering, **Figure 2.15A-B**). MAGs belonging to members of Chromatiaceae, *Contendobacter*, *Methanotherix*, MBNT15, and Methylomirabilota recruited 50% of binned unique peptides in the unamended control reactors. Collectively these MAGs accounted for less than 5% of binned unique peptides in the CT treatment, indicating the capacity for the polyphenol to shift active populations in soils under our study conditions. While this represents one of the first reports on the impacts of polyphenols on soil microbial community gene expression, similar temporal shifts in microbial community 16S rRNA gene membership have been observed with complex and pure polyphenols in soils and guts<sup>15,45,46</sup>.

In contrast to the unamended controls, CT-amended soils displayed a multi-phase gene expression response (**Figure 2.15A-C**). In phase 1, metaproteomes from CT treatments at day 1 could not be differentiated from unamended controls (**Figure 2.15A**). In phase 2 (days 3-10), a MAG from the Proteobacterial genus *Kosakonia* (CTSoil\_132, dark purple) accounted for 80% of the binned unique peptides from the CT-treated samples, with peak gene expression observed on day 3 (**Figure 2.15A&C**). In phase 3 (day 10-20), while *Kosakonia* expression was still detected, the CT-amended reactor metaproteome replicates displayed heterogenous responses (**Figure 2.15A&C**,

Phases 3a and 3b), dominated by either a novel member of the Acidobacterial genus *Holophaga* (CTSoil\_7, teal) or three novel MAGs in the Sporomusales undescribed family UBA7701 (CTSoil\_80, CTSoil\_81, & CTSoil\_82, dark blue). By genome-wide average amino acid identity and ribosomal protein similarity, these three Sporomusales MAGs likely represent three different genera (**Figure 2.18**). Of these three MAGs, CTSoil\_81 was dominant across the metaproteome data, recruiting four-times more peptides than the other two Sporomusales. While we did detect peptides from these three dominant MAGs (*Kosakonia*, *Holophaga*, or Sporomusales CTSoil\_81) in non-CT amended controls, these were annotated as primarily housekeeping (e.g. RNA polymerase) or hypothetical proteins (**Figure 2.15B**). Together, this suggested that while these microorganisms may have subsisted from metabolisms independent of polyphenols, they demonstrated different functionality under polyphenol exposure. Based on these findings, we concluded members of these 3 taxa were stimulated by polyphenols in anoxic soils. As such, we sought to link the metaproteome functions of these taxa and the broader microbial communities to our polyphenol degradation metabolite scheme.

#### 2.3.4 Polyphenol biodegradation occurs through metabolic exchange in anoxic soil

Metabolite evidence indicated soil microbiota depolymerized CT in the first 10-days (**Figure 2.11**), consistent with when *Kosakonia* was most active via metaproteomics (**Figure 2.15**). Given the size of the CT polymer, we expected any microbial depolymerization to be extracellular and thus we were particularly interested in the expression of two putatively-secreted enzymes from *Kosakonia* during this phase. One of these enzymes, a peroxidase (AA2) has been biochemically demonstrated to

aerobically degrade phenolic-rich lignin polymers<sup>47</sup>, while the other, a 1-4,benzoquinone reductase (AA6), is known to be indirectly involved in lignin degradation<sup>47</sup> (**Figure 2.19A**). The peroxidase, a predicted katG-type, uses H<sub>2</sub>O<sub>2</sub>-derived radicals to carry out 1-electron oxidations of a chemical mediator—potentially a phenolic compound or Mn<sup>2+</sup><sup>48</sup>. This extracellular low molecular weight mediator can diffuse to react with substrates outside the enzyme's spatial range. The benzoquinone reductase can participate in Fenton cycling to support H<sub>2</sub>O<sub>2</sub> pools<sup>47</sup> (**Figure 2.19A**). Analogous to what is proposed for aerobic lignin degradation, these oxidations could generate radical sites within the CT that promote depolymerization through cascades of bond scissions<sup>47</sup>. Further supporting this proposed role in anoxic CT depolymerization, these two enzymes (AA2, AA6) were recently implicated in pure-culture, anaerobic lignin degrading experiments by a close relative of *Kosakonia*<sup>49,50</sup>. As lignin is also a complex polyphenolic polymer, it is reasonable to extend the roles for these lignin associated enzymes to include CT depolymerization. Given these tantalizing shared findings at both the soil microcosm and isolate levels, biochemical characterization of these enzymes is warranted to expand roles for these canonically aerobic enzymes into anaerobic polyphenol degradation.

Analysis of the *Kosakonia* genome failed to detect known phenolic compound biodegradation pathways, suggesting *Kosakonia* enrichment is not fueled by phenolic catabolism. In support of this, during phase 1 and phase 2 we detected simultaneous expression of genes for sugar transport (e.g. maltose, fructose sugar phosphotransferase systems), central carbon metabolism, and acetate production (**Figure 2.20**). Thus, it is possible that *Kosakonia* performed CT transformation for

chemical detoxification, not energy-generation, while co-metabolizing sugars fermentatively<sup>51</sup>. In support of this *Kosakonia*-mediated detoxification, expression of genes for two previously observed mechanisms of CT tolerance were detected: RND-type transporters, to remove toxic phenolics from the cell<sup>52</sup>, and Spy proteins, thought to maintain cell membrane integrity in response to CT-induced environmental stress<sup>53</sup>. Collectively, this time-series expression data paired to high-resolution metabolite products during phases 1 and 2 (days 1-10) signified *Kosakonia* detoxified CT while fermenting sugars, ultimately serving as the most likely candidate for CT depolymerization in the live soil microcosms.

*Kosakonia* is also the most likely candidate for DOPA production, a C<sub>6</sub>-C<sub>3</sub> phenolic amine metabolite that was significantly produced in the microbially-active soils at later time points (**Figure 2.11E** and **Figure 2.19C**, purple arrow). We suggest *Kosakonia* produced DOPA via an aromatic amino acid aminotransferase, that was exclusively produced by *Kosakonia* in phases 2 and 3 when DOPA was produced (**Figure 2.21**). Plant root exudation and litter decay are commonly considered the primary source of soil DOPA, where this compound has broadly antagonistic allelochemical properties<sup>54</sup>. As an alternative source of DOPA in soils, our plant-free microcosms highlight that microbes could produce this compound from polyphenol-derived phenolics (**Figure 2.11E**). Beyond soils, this result may have cross-ecosystem ramifications. If similar microbial biochemistry occurs in the gut, DOPA could be microbially produced from dietary polyphenols, which could cross the blood-brain barrier and be converted to dopamine by host enzymes<sup>55,56</sup>, providing a plausible rationale for the positive gut-brain connection with polyphenol-rich foods (e.g. wine, chocolate)<sup>57</sup>.

Next, we investigated metabolic roles of microorganisms in the latter half of our experiment that could support the proposed biodegradation scheme where C<sub>15</sub> flavonoids (epicatechin, catechin, quercetin) were converted to smaller phenolic acids (**Figure 2.11C-F**). During this time, metaproteomic data implicated increased activity of a MAG affiliated with *Holophaga* and three MAGs (CTSoil\_80-82) within the Sporomusales family UBA7701 (**Figure 2.15C**). In comparing to known flavonoid degrading enzymes, proteome profiles from *Holophaga* and the Sporomusales UBA7701 MAGs showed these MAGs likely carried out the transformations observed in our metabolite data.

The first enzyme in this proposed flavonoid monomer degradation pathway was a chalcone isomerase (CHI), which could generate a chalcone<sup>58</sup> from opening the C-ring (position 1) of quercetin, a C<sub>15</sub> flavonoid detected only in our microbially active soils, likely from (epi)catechin monomers (**Figure 2.19B**). This chalcone could be reduced to phloretin by a second enzyme, a NADH-dependent flavanone- and flavanonol-cleaving reductase<sup>36</sup> (FCR, **Figure 2.19B**). C-ring cleavage is then completed with release of C<sub>6</sub> phloroglucinol and C<sub>6</sub>-C<sub>3</sub> acids by a third enzyme, phloretin hydrolase<sup>59,60</sup> (PHY, **Figure 2.19B**). While we recovered CHI and FCR from both *Holophaga* and two Sporomusales MAGs (CTSoil\_80 & 81), peptides for the last enzyme PHY were only confidently detected from *Holophaga* (**Figure 2.19B**), however the two Sporomusales MAGs encode this gene in their genome (CTSoil\_80 & 81, **Figure 2.19B** dotted line). This microbially-produced suite of enzymes likely catalyzed the degradation of the CT-oligomer derived flavonoids to other phenolic compounds observed after day 10 (**Figure 2.11A**).

While we note these enzymes (CHI, FCR, PHY) were first uncovered and described in flavonoid-degrading gut microbial isolates<sup>6</sup>, they remain poorly annotated in KEGG (and other databases) remaining as “hypothetical”, or non-specific classes like “oxidoreductases”. As such, we used non-homology-based annotation approaches, including coordinated gene expression-metabolite data combined with structural protein modeling, to inform these gene annotations (**Figure 2.22, Appendix B**). To the best of our knowledge, this is the first report of these enzymes in soil-derived microorganisms, collectively illustrating the ways that currently cryptic processes in soil can be informed by cross-ecosystem analyses from more tractable microbiomes.

Together these multi-omics data provided evidence for the biodegradation of CT monomers and their derivatives to phenolic acids (specifically phloroglucinol and C<sub>6</sub>-C<sub>3</sub> acids) by *Holophaga* and members of the Sporomusales. We note, the C<sub>6</sub> metabolite phloroglucinol was not detected in our exometabolites, but this was consistent with its typical rapid entry into primary metabolism<sup>5</sup>. Moreover, *Holophaga* and Sporomusales MAGs expressed putative phloroglucinol reductases (PGR), the key enzyme for phloroglucinol degradation via an energy-generating pathway producing acetate and butyrate<sup>61</sup> (**Figure 2.19C**).

In addition to monomer (C<sub>15</sub>) degradation, both *Holophaga* and the Sporomusales UBA7701 expressed several enzymes that carry out other phenolic transformations (**Figure 2.19C**). For example, from days 10-20, both *Holophaga* and UBA7701 MAGs expressed indole-pyruvate oxidoreductase, which could reduce C<sub>6</sub>-C<sub>3</sub> phenylpyruvates to observed C<sub>6</sub>-C<sub>2</sub> phenylacetates<sup>62</sup> (IPOR, **Figure 2.19C**). Further, they also produced phenylacetate CoA-ligase (PaaK, **Figure 2.19C**), the key enzyme

for degrading phenylacetate via an anaerobic pathway that feeds to central metabolism<sup>63</sup>. Lastly, and in support of a specialized form of anaerobic respiration, both *Holophaga* and *Sporomusales* MAGs expressed genes for the Car-system which could allow caffeic acid<sup>64</sup> reduction (a C<sub>6</sub>-C<sub>3</sub> phenolic metabolite detected in our CT reactors, **Figure 2.11F**). Consistent with prior reports and supported by the metaproteome data, we propose these taxa couple sugar and phenolic oxidation (and maybe CO<sub>2</sub> fixation, **Figure 2.20**), to the reduction of the abiotically-generated CT metabolite caffeate as an electron acceptor, generating 3,4-dihydroxypropionate (**Figure 2.19B**).

Our metaproteome results illustrated the vast levels of functional redundancy that reside in soils, where members of two different phyla (Acidobacteria and Firmicutes) expressed nearly identical metabolic pathways for C<sub>15</sub> flavonoid biodegradation and phenolic metabolism. Taken together, these late phase dominant members (*Kosakonia*, *Holophaga*, *Sporomusales*) expressed enzymes to metabolize a range of CT oligomers and their derived metabolites, demonstrating that this model polyphenol was accessible to soil microbiota under anoxic conditions. Ultimately, these findings illustrate the latent metabolic versatility awaiting discovery within microbiomes across soils.

### 2.3.5 Anoxic soil carbon cycling is resistant to polyphenol amendment

Our metabolite and metaproteome data illustrated that members of the soil microbiome can degrade polyphenols under anoxia. Beyond supposed limited polyphenol degradation, the enzyme latch paradigm suggests that polyphenols suppress microbial activity under anoxic conditions by binding extracellular hydrolase enzymes (e.g. CAZymes, peptidases) and substrates (e.g. polysaccharides, proteins)<sup>10</sup>. Our metaproteome data indicated diverse microbial taxa were active under CT-

amendment (**Figure 2.15**), and we next wanted to explore the impacts of CT on general microbial metabolic activities.

Additional analyses of the FTICR-MS data revealed polysaccharide-like compounds decreased over time in microbially active CT reactors (**Figure 2.23**), findings that would not be expected if microbial activity was halted as expected by the polyphenol enzyme latch. However, we note polysaccharide-like compounds were higher initially in CT-amended microcosms and thus perhaps more available. Reasons for this could include CT amendment priming the liberation of sugars in soils<sup>65</sup>, or low level contamination of CT, yet the latter is not supported by an in depth molecular characterization of the pure CT<sup>21</sup> (**Figure 2.7**). Regardless of the origin, our metaproteomic data supported increased degradation of polysaccharide-like compounds observed in the CT-amendment. We detected expression of 15 different carbohydrate-active enzymes (CAZymes) in CT reactors over phases 2 and 3, yet we did not recover peptides for CAZymes in the unamended control soils (**Figure 2.20**). Furthermore, we observed a corresponding decrease in LC-MS identified disaccharides over time in microbially-active CT reactors that was identical to unamended controls (**Figure 2.23**), suggesting active carbon substrate utilization was unimpeded by CT-amendment under these anoxic conditions. Taken together, our enzyme and metabolite data did not support the enzyme latch model where polysaccharides are inaccessible to the anaerobic soil microbial community in the presence of polyphenols (**Figure 2.1A**)<sup>10</sup>.

Consistent with unhindered anaerobic carbohydrate metabolism under polyphenol exposure, we observed CT-exclusive expression of sugar phosphotransferase systems (PTS, proteins used for transporting sugars into the cell),

and unchanged expression of glycolytic enzymes at all timepoints (**Figure 2.20**). The most striking difference between CT and unamended control metaproteomes was expression of microbial fermentation pathways only under CT treatment, particularly during phase 3 when CT has been depolymerized to fermentable phenolics<sup>66</sup> (**Figure 2.23**). The CT responsive MAGs (*Kosakonia*, *Holophaga*, and *Sporomusales\_UBA7701*) were inferred to be the biggest contributors to fermentative enzymes, accounting for half of the unique peptides assigned, but other members of the Firmicutes and Acidobacteria phyla also expressed these pathways (**Figure 2.20**). This metaproteome data was reinforced at the metabolite level, where the CT-exclusive production of formate, butyrate, and acetate was observed over time (**Figure 2.23**). In summary, our genome-resolved metaproteomics enabled a new view of anaerobic soil microbial carbon catabolism, where polyphenol amendment did not restrict basal microbiome function.

Based on a handful of studies in the rumen, it was historically assumed that methanogens were directly inhibited by polyphenols<sup>67</sup>, yet recent studies have suggested the opposite may be true, as methanogen 16S rRNA genes were enriched in rice paddy field soils amended with lignin-derived phenols<sup>68</sup>. Here, we demonstrated that methanogenic gene expression was not impacted by CT treatment relative to unamended controls in our anoxic reactors (**Figure 2.20**). Methane was below 12.5 ppm in all samples (CT-amended and unamended) after 20-days. However, there were metaproteomic hits for the key methanogenesis gene *mcrA* maintained across treatments, from the acetoclastic *Methanotherix* that was implicated as the dominant methanogen in these soils under field conditions<sup>18</sup>, and another from a

Thermoplasmatota methanogen, which was described as a key contributor to methylotrophic (C<sub>1</sub>-methyl) methanogenesis in these soils<sup>20</sup> (**Figure 2.20**). As mentioned above, the methanogenic substrate acetate increased in CT-amended reactors over time, while methanol was detected in biotic and autoclaved CT-amended reactors across time, but not in unamended controls (**Appendix B**). Taken together, our multi-omic data failed to provide evidence that CT was toxic to these soil methanogens, and instead uncovered how abiotic and biotic CT transformations may contribute to cross-feeding these climatically relevant microorganisms in anoxic soils.

## **2.4 Discussion**

This study provided evidence that the anoxic soil microbiome is capable of polyphenol metabolism that includes depolymerization of a condensed tannin polymer and subsequent monomer degradation. We offered a new multi-omics enabled view of the soil microbiome's response to a high molecular weight polyphenol under anoxia. Together our data support a model in which polyphenols in soils are not as microbially inert as previously claimed.

Importantly, our findings provide a new scaffolding that others can leverage. We expanded the definition of soil polyphenol degrading enzymes from solely (poly)phenol oxidase<sup>9,10</sup>, to include at least nine other enzymes (**Figure 2.19**). Additionally, we highlighted canonically aerobic enzymes (i.e. peroxidase) that may play unrecognized roles in anoxic transformations of polymeric carbon, as has been recently suggested for other historically regarded aerobic enzymes under anoxia<sup>49,50,69</sup>. Our metaproteomic data unveiled the metabolic handoffs and redundancies between three anaerobic, polyphenol-responsive taxa in the soil microbial community (*Kosakonia*, *Holophaga*, and

Sporomusales UBA7701). Finally, we showed that the underlying capacity for anaerobic carbon cycling by the soil microbiome was largely unchanged by polyphenol amendment.

We acknowledge our approach used laboratory soil reactors separated from environmental factors like fluctuating temperature, continual organic matter inputs, and interactions with micro- and macrofauna. Thus, more detailed and field-oriented studies are needed to uncover the occurrence and consequences of anoxic polyphenolic degradation under native conditions, across a range of soil types, and with different polyphenol substrates. However, here we provide initial metabolite and enzyme signatures for this process that can now be explored in greater detail in future studies.

Our study contributes to a growing body of recent research dispelling long-held notions of soil microbiomes as being intractable due to their chemical and biological heterogeneity<sup>17,70</sup>. By employing multiple metabolite approaches, we tracked the transformations of a defined polyphenol along a molecular weight gradient and biochemical hierarchy: from FTICR-MS-identified oligomers, to LC-MS-identified flavonoids and phenolic acids, to NMR-identified fermentation products. Moreover, our metabolite findings echoed one another across methods (i.e. mutual detection of monomers at day 10 between FTICR-MS and LC-MS), allowing for seamless tracking of metabolites in soils. On top of this resolved view of carbon chemistry, we overlaid microbial community-wide proteome data, linking transformations of structurally-defined metabolites to enzymes that were uniquely assigned to specific genomes. We highlight the potential for the tools used here, along with a suite of other emerging

technologies<sup>71-74</sup>, to illuminate soil microbiological and chemical processes historically confined to the “black box” of soil biochemistry.

Beyond the boundaries of these laboratory reactors, polyphenols have long been thought to act as controllers of global soil carbon storage<sup>75</sup>. In fact, several recent studies have suggested polyphenol-supplementation as a strategy to prevent carbon loss in peatlands<sup>7,10,11</sup>. However, in light of the genome-resolved metaproteome and metabolite evidence from this study, the extent that polyphenols sequester soil carbon warrants further investigation. While our study demonstrated that under anoxic conditions the soil microbiome in a freshwater wetland can degrade polyphenols and bypass proposed polyphenol locks on carbon cycling, translating this finding to climate mitigation strategies, especially relevant to peat systems, requires: quantifying the kinetics and environmental constraints of these transformations on the overall carbon budget, expanding research to other relevant polyphenol substrates, and investigating the effects of abiotic and biotic polyphenol transformations associated with diverse soil types. Our findings pave a way for these research avenues, providing metabolite and enzyme framework for mining these processes from complex systems. Collectively, our results highlight the promise of modern soil microbiome technologies for uncovering the ecological and biochemical mechanisms underlying long-held soil biogeochemical paradigms.

## **2.5 Methods**

### *2.5.1 Soil sample Collection*

We used a soil sample collected from a plant-covered mudflat (August 2015) in Old Woman Creek National Estuarine Research Reserve<sup>18</sup> (OWC) (41°22'N 82°30'W).

The soil sample was stored at  $-20^{\circ}\text{C}$  until use. While we recognize that thawing these frozen soils for use in the laboratory may have impacted soil carbon availability, these soils routinely experience freeze thaw throughout the winter months and thus are exposed to fluctuating temperatures.

### *2.5.2 Condensed Tannin purification*

The broad class of plant secondary metabolites known as polyphenols includes three types of high molecular weight compounds, the lignins, the hydrolysable tannins and the condensed tannins<sup>76</sup>. Lignins are highly methoxylated derivatives of the  $\text{C}_6\text{-C}_3$  phenylpropanoids, and their fate and effects in soils have been extensively examined<sup>47</sup>. The unmodified phenolic moieties of tannins make these compounds more highly reactive than lignin, including their ability to serve as antioxidants, as metal binding agents, and their quintessential property of protein binding/precipitation<sup>77</sup>. Of the two classes of tannins, the hydrolysable tannins are highly susceptible to chemical and enzymatic decomposition via hydrolysis of ester linkages, and their metabolic fate in gut and soil microbiomes is well-established<sup>5</sup>. The condensed tannins, or proanthocyanidins, comprise flavan-3-ol subunits connected by chemically stable interflavan bonds that are degraded most conveniently with strong acid under oxidizing conditions<sup>78</sup>. Because condensed tannin (CT) appears to be more recalcitrant to degradation under biological conditions, it is an excellent substrate for this proof-of-concept study. Sorghum grain is a unique source for easily purifying hundreds of mg of CT as a chemically homogeneous preparation with a simple structure suitable for detailed metabolomic tracing.

Mature grain from high tannin *Sorghum bicolor* (L.) Moench grain (Hi-Tannin Sumac NM03-9905, Scott Bean, USDA Manhattan Kansas) was stored at 4°C. Tannin was extracted from ground grain with methanol containing ascorbic acid and purified by ethyl acetate extraction to remove small phenolics, followed by Sephadex LH20 chromatography to isolate the high molecular weight fraction<sup>21,79</sup>. The freeze-dried powder was stored at -20°C. The tannin was characterized by thiolysis to establish that the average degree of polymerization was 16, with a catechin terminal unit and epicatechin extenders (**Figure 2.1B**). The material's purity was assessed with NMR and HPLC (**Figure 2.7, Appendix A**)<sup>21</sup>.

### 2.5.3 Reactor design and set up

To establish microcosms, frozen soil was thawed at room temperature for 1 hour. 5g of soil and the headspace was degassed in a Wheaton serum bottle for 30 minutes with 5 psi of N<sub>2</sub> gas. A slurry was prepared by anoxically-transferring 125mL anoxic sterile water via N<sub>2</sub>-degassed, sterile syringe to the degassed soil-containing serum bottle sealed with a butyl rubber stopper and an aluminum crimp. After inoculating the biologically active reactors with soil slurry (both CT amended and unamended), as discussed below, the remaining soil slurry was autoclaved three-times for 30min each, and then inoculated into reactors as in the live controls. We confirmed we could not recover DNA or amplify DNA from the reactors inoculated with autoclaved soil slurry at each time point the biologically active samples were taken (**Appendix A**), supporting their microbially inactive status.

Anoxic reactors were established and sampled using prior methods that were demonstrated to support the growth of obligatory anaerobic metabolisms in soils and

subsurface samples<sup>20,80–82</sup>. The medium was basal bicarbonate-buffered<sup>20</sup>, consisting of (per liter): 0.25 g ammonium chloride, 0.60 g sodium phosphate, 0.10 potassium chloride, 2.5 g sodium bicarbonate, 10 ml DL-vitamin mixture (**Table 2.2**), and 10 ml DL-mineral mixture<sup>83</sup> (**Table 2.2**), and was brought to a pH of 7.0 using 1 mM NaOH. The biotic and autoclaved reactors were prepared with 90mL and 45mL, respectively, of media in 200mL serum bottles with a N<sub>2</sub>-CO<sub>2</sub> (80:20) headspace using standard anaerobic microbiology practices<sup>80,84</sup>. The anoxic soil slurry (autoclaved or biotic) was added to the reactors in a 1:10 dilution. CT-amended reactors (autoclaved and biotic) were established by adding anoxic, sterile CT stock solution in DI water (15 mg/mL), to achieve a final dosing of 1.5 mg/mL reactor. Reactors were flushed with N<sub>2</sub>-CO<sub>2</sub> (80:20) gas in media-soil slurry and serum bottle head space for 40 minutes to ensure removal of trace oxygen before incubation.

Reactors were incubated in the dark and at 25°C, consistent with field soil temperatures<sup>20</sup>. Here we selected field-relevant temperate operation (25°C)<sup>20</sup> to remove kinetic constraints on polyphenolic microbial growth and enzyme activity that were previously indicated in low temperature studies from boreal peatland soils<sup>7,12,19</sup>, as we consider it possible this temperature stress may have confounded interpretations of microbial polyphenol metabolism. Subsamples were collected over 20-days for 16S rRNA gene, metagenomic, metaproteomic, and various metabolomic and geochemical analyses (**Figure 1B**). All subsamples were collected with care for maintaining anoxic conditions according to standard anaerobic microbiology protocols<sup>80,82,84</sup>, briefly, sampling was performed using sterile syringes that were degassed completely with N<sub>2</sub>-CO<sub>2</sub> (80:20, vol/vol) to ensure no oxygen transfer. Subsamples were immediately

dispensed into their respective storage tubes, flash frozen, and stored at  $-80^{\circ}\text{C}$  until processing/analysis.

Methane production was measured after 20-days as in Narro et al<sup>20</sup>. Briefly, we used a Shimadzu (GC-2014) gas chromatograph (GC) equipped with a thermal conductivity detector and using helium as a carrier gas at  $100^{\circ}\text{C}$  to quantify methane from triplicate CT-amended and unamended control microcosms reactors at day 0 and at day 20.

#### 2.5.4 16S rRNA Gene Analyses

Total nucleic acids were extracted from the microcosms at days 0, 1, 3, 5, 7, 10, 14, and 20 using the Qiagen DNeasy PowerSoil Kit, and were stored at  $-20^{\circ}\text{C}$  until sequencing. Sequencing of the V4 region of the 16S rRNA gene was performed at Argonne National Laboratory's Next Generation Sequencing Facility on the Illumina MiSeq using 251-bp paired-end reads and the Earth Microbiome Project primers<sup>85</sup>. Reads were demultiplexed and analyzed within QIIME2 (2017.10) using DADA2<sup>86</sup> to produce an amplicon sequence variant (ASV) by sample table (**Appendix B**), with taxonomy assigned using SILVA classifier (silva132.250). We filtered the feature table to contain only ASV's observed in at least 3 samples. To survey ASV in reference databases, we BLASTed ASVs against RefSoil cultivated isolate genomes (of which 96% (n=882) encode a 16S rRNA gene)<sup>44</sup>. ASV sequences were considered positive hits if they matched a sequence at greater than 97% identity over at least 74bp (**Table 1.1**).

### 2.5.5 Metagenomic Sequence and Assembly

For days 5, 10, and 20, we obtained a CT- and control microcosm metagenome from pooled triplicate samples (n=6 metagenomes). For this, genomic DNA was prepared for metagenomic sequencing using the Nextera XT Low Input-Illumina library creation kit, and was sequenced at the Department of Energy Joint Genome Institute on the Illumina NovaSeq 6000. Fastq files were trimmed using Sickle (v 1.33)<sup>87</sup>, and trimmed reads were assembled using IDBA-UD<sup>88</sup> using k-mers (40, 60, 80, and 100). To maximize assembly, we performed (1) subtractive assemblies, iteratively assembling reads that did not map to assembled scaffolds  $\geq 3$  kb at 97% identity on all metagenomes, and (2) subassemblies using 25% of the combined CT-amended metagenome trimmed reads. Information for metagenome statistics, including assembly information, are found in **Appendix B**. For each assembly, scaffolds  $\geq 2.5$  kb were binned using MetaBAT2<sup>89</sup> (v2.12.1), and MAG completion was assessed using AMPHORA2<sup>90</sup> and checkM<sup>91</sup> (v1.1.2). MAGs were kept in the database if they were  $>50\%$  complete and  $<10\%$  contaminated by either of these tools, or if it was  $>35\%$  complete with  $<1\%$  contamination in the event they recruited peptides in metaproteomes. MAGs were dereplicated at 99% identity using dRep<sup>92</sup> (v2.6.2). MAG taxonomy was assigned using GTDB-tk (v1.3.0) R05-RS95<sup>43</sup>. See **Appendix B** for MAG quality and taxonomy information.

MAGs and assemblies were annotated using DRAM<sup>93</sup>. CAZymes were inferred from the DRAM hits. Enzymes in **Figure 2.19C** (except PGR) were mined from DRAM *raw* outputs. To mine C<sub>15</sub> flavonoid enzymes (**Figure 2.19B** and PGR), we constructed a custom database using published, characterized proteins<sup>6,36,59,60</sup> (**Appendix B**). Using

BLASTp, we searched for these enzymes in the metaproteome and in MAGs and putative hits were identified using a bit score cutoff greater than 150. Blast hits that met this criterion were further structurally modelled using PHYRE2<sup>94</sup> web server to support putative roles. See **Figure 2.22** and **Appendix B** for structural modelling and BLASTp information, and sequences.

To quantify MAG relative abundance in each temporal sample and condition, trimmed metagenomic reads were mapped to the dereplicated MAG set using bbmap<sup>95</sup> (v38.70) at `minid=95`, and output as sam files which were converted to sorted bam files using samtools<sup>96</sup> (v1.9). We had two requirements for a MAG to be found in a sample: first we required reads to map to at least 75% of a MAG in a given sample, and second the MAG had to have at least 3X coverage in that sample. To determine MAGs that had reads mapped to at least 75% of the MAG, we used CoverM<sup>97</sup> (v0.3.2) in genome mode to output MAGs that passed this threshold (`--min-covered-fraction 75`). To obtain MAG coverage, we used CoverM<sup>97</sup> (v0.3.2) in genome mode to output `reads_per_base` (reads mapped/genome length), and from this calculated MAG coverage as `reads_per_base x 151bp`. A bin was “present” in CT or in control if it was found with at least 3X average coverage across the MAG and had reads mapped to at least 75% of the MAG in any of the timepoints, or was “present” in both treatments if these two criteria were met in both CT and control metagenomes (ex. Present at day 5 in CT and at day 5 in Unamended). This information is given in **Appendix B**.

### *2.5.6 Metaproteomic Extraction and Spectral Analysis*

Liquid culture (5 mL) from each microcosm sample was collected anaerobically, centrifuged for 15 min at 10,000 ×g, separated from the supernatant that was used for

metabolite characterization and stored at  $-80^{\circ}\text{C}$  until shipment to Pacific Northwest National Laboratory. Proteins in the pellet were precipitated and washed twice with acetone. Then the pellet was lightly dried under nitrogen. 200 $\mu\text{L}$  of an 8M urea solution was added to the protein pellet, vortexed into solution. A bicinchoninic acid (BCA) assay (Thermo Scientific, Waltham, MA USA) was performed to determine protein concentration. Following the assay, 10mM dithiothreitol (DTT) was added to the samples and incubated at  $60^{\circ}\text{C}$  for 30 minutes with constant shaking at 2,552 xg. Samples were then diluted 8-fold for preparation for digestion with 100 mM  $\text{NH}_4\text{HCO}_3$ , 1 mM  $\text{CaCl}_2$  and sequencing-grade modified porcine trypsin (Promega, Madison, WI) was added to all protein samples at a 1:50 (w/w) trypsin-to-protein ratio for 3 h at  $37^{\circ}\text{C}$ . Digested samples were desalted using a 4-probe positive pressure Gilson GX-274 ASPEC™ system (Gilson Inc., Middleton, WI) with Discovery C18 100 mg/1 mL solid phase extraction tubes (Supelco, St.Louis, MO), using the following protocol: 3 mL of methanol was added for conditioning followed by 2 mL of 0.1% TFA in  $\text{H}_2\text{O}$ . The samples were then loaded onto each column followed by 4mL of 95:5:  $\text{H}_2\text{O}$ :ACN, 0.1% TFA. Samples were eluted with 1mL 80:20 ACN: $\text{H}_2\text{O}$ , 0.1% TFA. The samples were concentrated down to  $\sim 30\mu\text{L}$  using a Speed Vac and a final BCA was performed to determine the peptide concentration and samples were diluted to 0.1  $\mu\text{g}/\mu\text{L}$  with nanopure water for MS analysis.

All mass-spectrometric data were acquired using an Orbitrap Lumos (Thermo Scientific) connected to a nanoACQUITY UPLC M-Class liquid chromatography system (Waters) via in-house 30-CM x 75- $\mu\text{M}$  column packed using Repronil-pur 1.9- $\mu\text{m}$  C18 particles (Dr. Maisch HPLC GmbH, Germany) and in-house built electrospray

apparatus. MS/MS spectra were compared with the custom metagenome and MAG database using the search tool MS-GF<sup>98</sup>. Contaminant proteins typically observed in proteomics experiments were also included in the protein collections searched. The searches were performed using  $\pm 15$ -ppm parent mass tolerance, parent signal isotope correction, partially tryptic enzymatic cleavage rules, and variable oxidation of methionine. In addition, a decoy sequence approach was employed to assess false-discovery rates. Data were collated using an in-house program, imported into a SQL server database, filtered to  $\sim 1\%$  false-discovery rate (peptide to spectrum level), and combined at the protein level to provide (i) unique peptide count (per protein) and (ii) observation count (spectral count) data. We required at least two unique peptides per protein for identification, and for analyses used spectral counts from these identified proteins to calculate normalized spectral abundance factor (see below). See **Appendix B**.

### *2.5.7 Metaproteomic database creation and analyses*

The database for our metaproteome analysis was constructed from a dereplicated (100% amino acid identity) set of genes that were identified on binned and unbinned metagenomic scaffolds (i.e. all scaffolds >2.5 kb) (Fig 4B). The inclusion of unbinned genes was done to allow us to account for assembled, expressed genes that were not assigned to genomic bins. We verified this Dereplicated Gene Database equally recruited metagenome reads from CT amended and CT unamended reactors, and thus was not biased by treatment (**Figure 2.13, Appendix A**). The CT-amended and unamended metaproteomes were mapped to this same Dereplicated Gene Database (**Figure 2.14**).

When reporting proteins identified in our metaproteome data, we assigned protein hits from our Dereplicated Gene Database to three categories (**Figure 2.14**). The first status was reported as “Non-Unique” if peptides identified from the mass spectra were assigned to *in silico* peptides that mapped to multiple genes in our Dereplicated Gene Database. The second status was reported as “unbinned unique” if peptides identified from the mass spectra were assigned to *in silico* peptides that mapped to a single gene, but this gene was not assigned to one of the reconstructed MAGs and was only assigned to an unbinned assembled scaffold. The third status was “binned unique”, where peptides identified from the mass spectra were assigned to *in silico* peptides that mapped to a single gene that was contained within a binned genome from our MAG database.

The three-classification system used in this metaproteomic analysis was designed to maximize the reporting of any expressed genes in a complex microbial community like soils, while also conservatively assigning gene expression to a specific genome where appropriate. The non-unique classification accounted for strain heterogeneity in soils with (i.e. several near identical genes in our database come from very closely related organisms and equally recruit peptides) and for proteins that have highly conserved sequences (i.e. ATP synthase). The expression patterns of these genes would have been excluded from downstream analyses if we relied only on unique peptide recovery. The unbinned-unique classification accounted for the fact some of the genes in our Dereplicated Gene Database were from assembled scaffolds that could not be assigned to a MAG through the genome binning process. The analyses reported in the manuscript used the binned-unique data (unless noted), with all reported

proteome classification data shown in the supplementary analyses (**Figure 2.16**, **Appendix B**).

We took an untargeted, discovery-based approach to our metaproteomes and used label-free quantitation, consistent with many metaproteomic studies in environmental microbiomes to date<sup>99–102</sup>. Specifically, we used spectral counts where the number of unique spectra recovered for peptides are assumed to scale with their abundance. However, spectral counts are imperfect as they are biased by protein size and by sample-to-sample variation<sup>103</sup> (**Figure 2.14**). Therefore, we converted spectral counts to normalized spectral abundance factor (NSAF), which includes normalizations that account for spectral count bias, making it a preferred method of quantitation from untargeted metaproteomes<sup>103–106</sup>. To calculate NSAF, the spectral count of a protein is divided by the protein length to give protein spectral abundance. This value is then divided by the sum of all protein spectral abundances to give the normalized spectral abundance<sup>105</sup>. This enabled comparison of a protein's relative abundance within and across samples.

### *2.5.8 Integrated metabolomic approaches*

Historically, microbial transformations of polyphenols were inferred using low-resolution assays for total polyphenol content (i.e. the Folin–Ciocalteu assay) or CT-specific assays (ie. the acid butanol assay)<sup>12</sup>. Results from these assays have been the basis for theories like the “enzyme latch”, enabling the persistent idea that polyphenols are not susceptible to degradation under anoxic conditions<sup>9</sup>. However, these assays are not suitable for quantifying polyphenol content broadly in soils and especially for detailing the effects of microbial degradation of polyphenols in soils. For example, the

widely-used Folin–Ciocalteu assay has limited quantitative application<sup>26,27</sup> as it is nonspecific for quantifying polyphenols in complex matrices like soils, as the reagents react with a wide variety of compounds (e.g. thiols, vitamins, proteins, and inorganics<sup>26</sup>) contained within the soil matrix, thereby giving error prone concentrations of bulk polyphenols. Furthermore, polyphenols are structurally diverse, and “total polyphenol” content gives little information on structural changes. Additionally, the acid butanol assay for determining CT concentrations was shown to be non-specific for differentiating oligomer sizes of polyphenols<sup>30</sup>, meaning it would not resolve microbial depolymerization of the parent polyphenol into oligomers, a process which is a key indicator of degradation of condensed tannins<sup>5</sup>. Further complicating the scenario, CT is highly reactive with protein biomass and soil matrix<sup>22</sup>, thus it is difficult to differentiate removal of CT by sorption and loss of CT due to biotransformation by microbes<sup>28,29</sup>. Therefore, we used high resolution instrumental approaches instead of chemical assays to identify metabolites indicative of (i) increased polymer depolymerization (breakdown into smaller oligomers and monomers) over time and (ii) production of further phenolic degradation metabolites.

To determine depolymerization of CT over time and the chemical degradation produced from microbial processes, we integrated metabolite data from several analytical techniques. Using this data, we specifically looked for metabolite evidence of the following fates for the added CT: depolymerization, here defined as breakage of the interflavan bond (**Figure 2.4**), biodegradation, here defined as signals that were unique to biologically-active soils relative to autoclaved soil, and transformation, here defined as signals that were temporally-distinct but could not be differentiated between

biologically-active and autoclaved soils. Furthermore, we used this metabolite data to support other metabolisms happening in the reactors.

### 2.5.9 FT-ICRMS analysis

We had two goals with our FTICR-MS analysis: (i) monitor changes in the CT polymer over time and (**Figure 2.4, Figure 2.5**) (ii) monitor changes in biochemical classes over time (**Figure 2.5, Figure 2.23**). Fourier Transform Ion Cyclotron resonance mass spectrometry (FTICR-MS) was used to collect high resolution mass spectra of the supernatant samples from reactors (microcosms) by direct injection in negative ion mode (**Appendix A**). For peaks that could be attributed to the CT polymer (**Figure 2.5, Appendix A**), Kendrick mass defect (KMD) analysis<sup>107</sup> was then used to compare the fate of (epi)catechin CT oligomers over time in both biologically active and inactive (autoclaved soil) reactors. We used a modified version of KMD commonly used for polymer ions, proposed by Sato et al<sup>32,108</sup>, calculated using equations 1-3.

$$KM(\text{ion})=m/z(\text{ion})*(290/290.079038) \quad (1)$$

$$NM\_CAT(\text{ion})=\text{roundup}(KM(\text{ion})) \quad (2)$$

$$KMD(\text{ion})=NM\_CAT(\text{ion})-KM(\text{ion}) \quad (3).$$

To track changes in biochemical classes over time, putative chemical formulas of all peaks were assigned using Formularity (v1.0.0) software<sup>109</sup> (**Figure 2.5**). Biochemical compound classes were reported as relative abundance values based on counts of C, H, and O for the following H:C and O:C ranges as in Tfaily *et al*<sup>110</sup>. For more detailed information on FTICR-MS methodology and analyses, see **Appendix A**. Processed data is provided in **Appendix B**, and raw data provided in archive (doi:10.5281/zenodo.4552584.).

### 2.5.10 LC-MS metabolomic analysis

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to identify exometabolites across samples over time. Metabolites were extracted into ethyl acetate from filtered supernatant samples after acidification with HCl. Both the aqueous and organic phases were dried down, redissolved, and analyzed by LC-MS/MS (**Appendix A**) using an Agilent 1290 UHPLC system connected to a Thermo Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer equipped with a Heated Electrospray Ionization (HESI-II) source probe. Separation, ionization, fragmentation and data acquisition parameters are specified in **Appendix B**. Briefly, metabolites were separated by gradient elution followed by MS1 and data dependent (top 2 most abundant MS1 ions not previously fragmented in last 7 seconds) MS2 collection; targeted data analysis was performed by comparison of sample peaks to a library of analytical standards analyzed under the same conditions. Three parameters were compared: matching m/z, retention time and fragmentation spectra using Metabolite Atlas (<https://github.com/biorack/metatlas>)<sup>111,112</sup>. Additional methodological details, including LC-MS parameters and MS resolution, are provided in **Appendix B**. Identification and standard reference comparison details are provided in **Appendix B**. For more information on LC-MS analyses, see **Appendix A**. To determine significantly discriminating LC-MS exometabolites, we applied a linear model to the log<sub>2</sub>-transformed peak area data using limma<sup>113</sup> (v3.42.2) in R on log<sub>2</sub>-transformed data to compare metabolites in live and autoclaved treatments at each timepoint. Limma statistics are given in **Appendix B**.

### 2.5.11 NMR metabolomic analysis

To follow important organic acids, we used NMR on supernatant samples. Supernatant samples (180  $\mu$ L) were combined with 2,2-dimethyl-2-silapentane-5-sulfonate- $d_6$  (DSS- $d_6$ ) in  $D_2O$  (20  $\mu$ L, 5 mM) and thoroughly mixed prior to transfer to 3 mm NMR tubes. NMR spectra were acquired on a Varian 600 MHz VNMRS spectrometer equipped with a 5 mm triple-resonance (HCN) cold probe at a regulated temperature of 298 K. The  $90^\circ$   $^1H$  pulse was calibrated prior to the measurement of each sample. The one-dimensional  $^1H$  spectra were acquired using a nuclear Overhauser effect spectroscopy (NOESY) pulse sequence with a spectral width of 12 ppm and 512 transients. The NOESY mixing time was 100ms and the acquisition time was 4s followed by a relaxation delay of 1.5 s during which presaturation of the water signal was applied. Time domain free induction decays (57472 total points) were zero filled to 131072 total points prior to Fourier transform. Chemical shifts were referenced to the  $^1H$  methyl signal in DSS- $d_6$  at 0 ppm. The 1D  $^1H$  spectra were manually processed, assigned metabolite identifications and quantified using Chenomx NMR Suite 8.3. Metabolite identification was based on matching the chemical shift, J-coupling and intensity of experimental signals to compound signals in the Chenomx and custom in-house databases. Quantification was based on fitted metabolite signals relative to the internal standard (DSS- $d_6$ ). Signal to noise ratios (S/N) were measured using MestReNova 14 with the limit of quantification equal to a S/N of 10 and the limit of detection equal to a S/N of 3. Processed data is available in **Appendix B**, and raw data provided in archive (doi:10.5281/zenodo.4552584).

Chapter 2 Tables

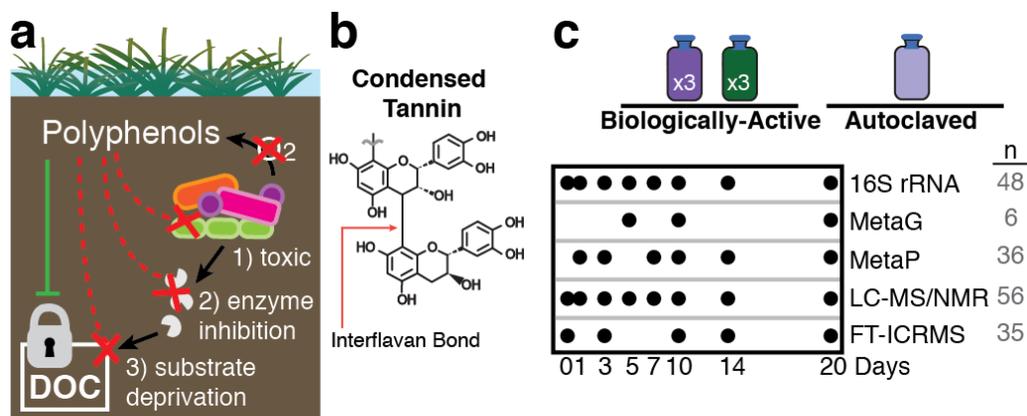
**Table 2.1. Positive hits from of ASVs BLAST to RefSoil database.**

<b>AS V</b>	<b>RefSoil Subject</b>	<b>%id</b>	<b>length</b>	<b>e-value</b>	<b>Lowest SILVA ASV taxonomy level</b>
1	SVEN_RS3716 5	100	136	9.18E-65	Streptomyces
2	SVEN_RS3716 5	100	136	9.18E-65	Streptomyces
3	SVEN_RS3716 5	100	136	9.18E-65	Streptomyces
4	SVEN_RS3716 5	100	136	9.18E-65	Streptomyces
5	SVEN_RS3716 5	99.26 5	136	4.27E-63	Bacillus
6	SVEN_RS3716 5	97.79 4	136	9.24E-60	Streptomyces
7	SVEN_RS3716 5	97.05 9	136	4.30E-58	Streptomyces
8	SVEN_RS3716 5	97.05 9	136	4.30E-58	Bacillus
9	BSN5_RS2123 0	100	74	2.68E-30	Oceanobacillus;uncultured bacterium
10	BSN5_RS2123 0	97.29 7	74	5.80E-27	Streptomyces
11	BSN5_RS2123 0	97.29 7	74	5.80E-27	Streptomyces

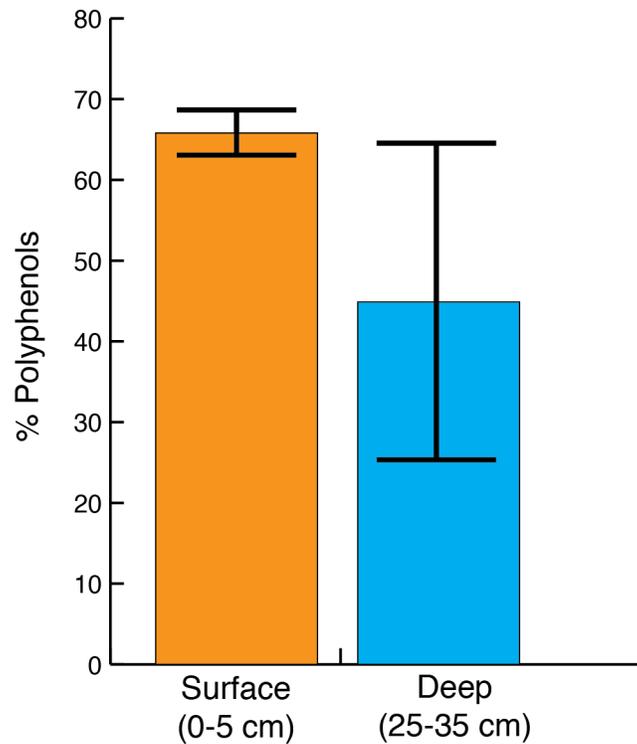
**Table 2.2. DL-Vitamins and Minerals in Media according to Lovley *et al.*<sup>83</sup>**

	<b>Components</b>	<b>Mg/L in ddiH<sub>2</sub>O</b>
Vitamin Mix	d-Biotin	2
	Folic Acid	2
	Pyridoxine HCl	10
	Riboflavin	5
	Thiamine	5
	Nicotinic Acid	5
	Pantothenic Acid	5
	Vitamin B12	0.1
	p-Amino benzoic Acid	5
	D,L-6,8-thiotic Acid	5
Mineral Mix	NTA Disodium Salt	1.5
	MgSO <sub>4</sub> ·7H <sub>2</sub> O	3
	MnSO <sub>4</sub> ·H <sub>2</sub> O	0.5
	NaCl	1.0
	FeSO <sub>4</sub> ·7H <sub>2</sub> O	0.1
	CaCl <sub>2</sub> ·2H <sub>2</sub> O	0.1
	CoCl <sub>2</sub> ·6H <sub>2</sub> O	0.1
	ZnCl	0.13
	CuSO <sub>4</sub> ·5H <sub>2</sub> O	0.01
	AlK(SO <sub>4</sub> ) <sub>2</sub> ·12H <sub>2</sub> O	0.01
	Boric Acid	0.01
	Na <sub>2</sub> MoO <sub>4</sub> ·2H <sub>2</sub> O	0.025
	NiCl <sub>2</sub> ·6H <sub>2</sub> O	0.024
	Na <sub>2</sub> WO <sub>4</sub> ·2H <sub>2</sub> O	0.025
	Na <sub>2</sub> SeO <sub>4</sub>	0.02

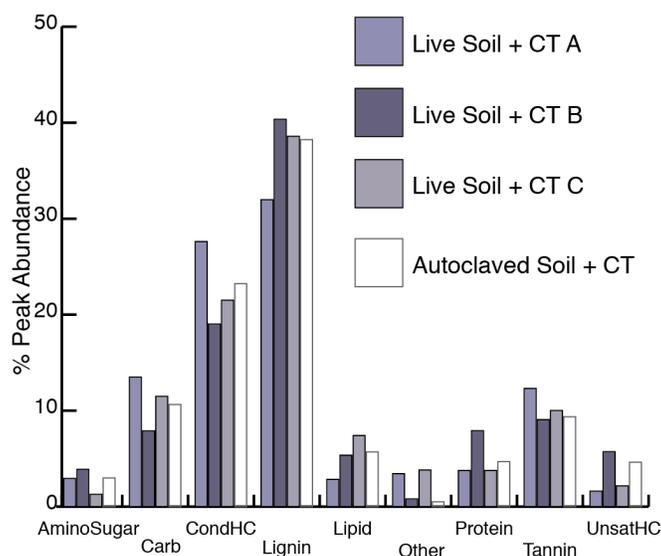
## Chapter 2 Figures



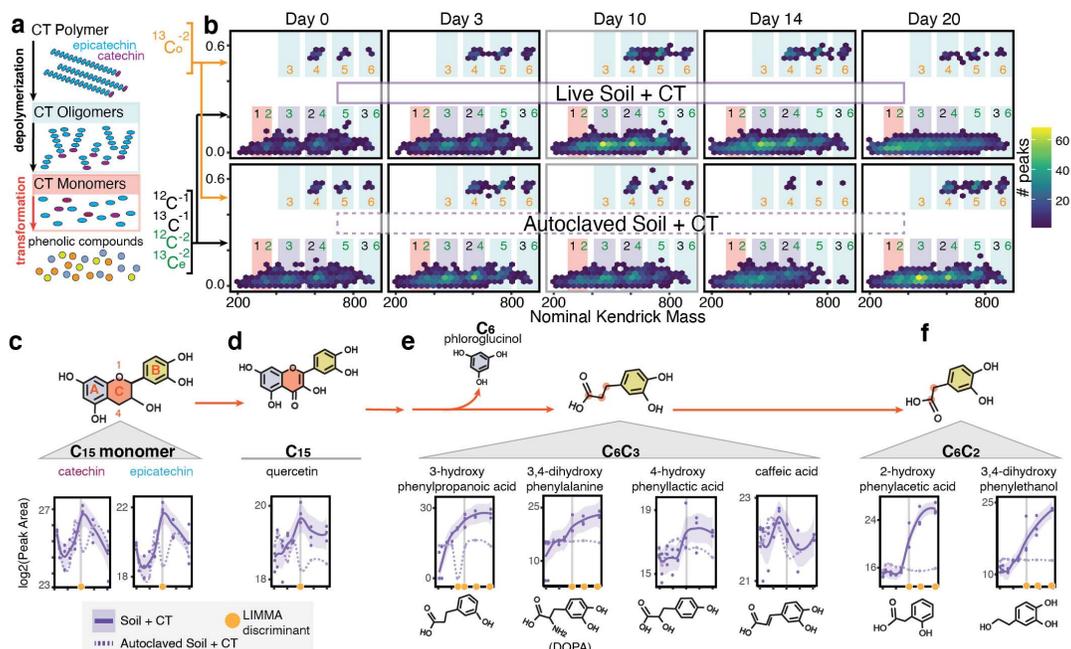
**Figure 2.1.** (a) Schematic summarizing the polyphenol lock paradigm<sup>24,32</sup>, demonstrating the ways these compounds may control microbial carbon transformations in anoxic soils. The green solid line indicates that in anoxic soils, polyphenols promote the lock on dissolved organic carbon (DOC). The dotted-red lines show the three commonly proposed mechanisms by which polyphenols restrict the activity of soil microorganisms to lock soil carbon, including (1) toxicity to microorganisms, (2) inhibiting microbial extracellular enzymes, and (3) binding and depriving microorganisms of nutrients. (b) A purified condensed tannin (CT) was selected as the model polyphenol in this study due to its inferred lack of microbial degradation in anoxic soils<sup>39</sup>. This model compound is well characterized chemically<sup>40</sup> and has an average degree of polymerization of 16, where interflavan bonds (red arrow) connect monomers of epicatechin with a single catechin monomer cap. (c) The experimental design included soil reactors from three treatments (i) Biologically-active CT amended (dark purple), (ii) Biologically-active unamended control (green), and (iii) CT-amended autoclaved control (light purple). Autoclaved soils only included metabolite analyses, while microbially-active soils were analyzed with the suite of multi-omics approaches. The timepoints of each type of analysis are shown, with the total number (n) of samples across treatments denoted on the right in grey.



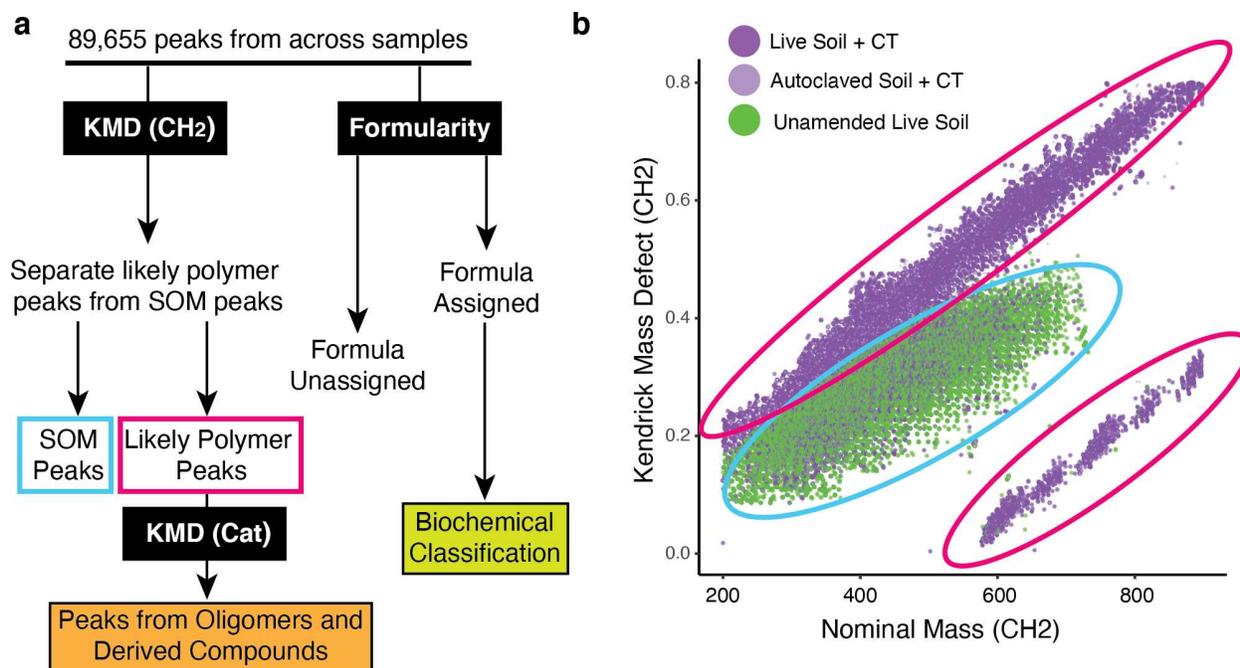
**Figure 2.2.** Abundance of polyphenol-like compounds in field surface and deep wetland soils. Polyphenol-like compound abundance in Fourier-transform ion cyclotron resonance mass spectrometry (FTICR-MS) data from the wetland, which was used as the soil microbial community inoculum in our microcosms. Here we used the surface soil as our inoculum. Barplots of FTICR-MS identified lignin- and tannin-like peak abundance in surface (orange) and deep (blue) soils. Error bars show one standard deviation (n=3 each).



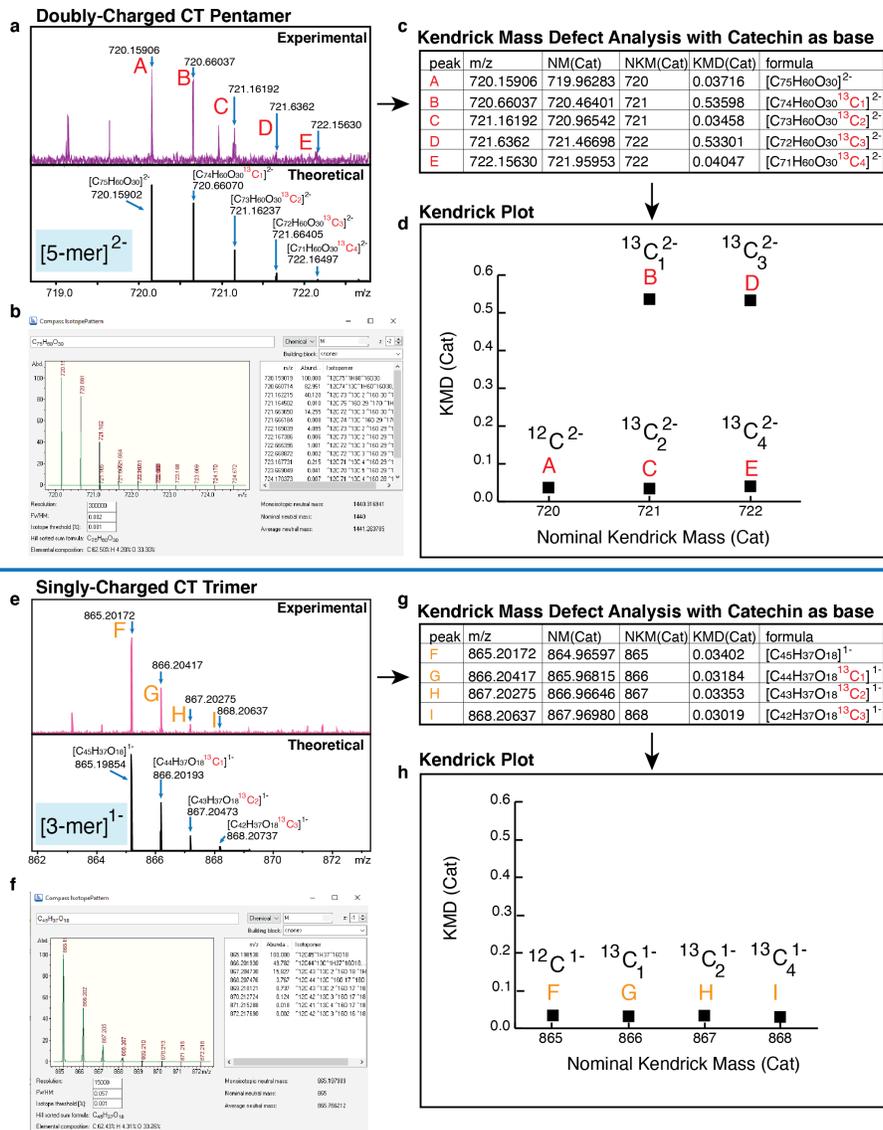
**Figure 2.3.** Abundance of FTICR-MS identified biochemical classes between live and autoclaved soil samples. Peak abundance of biochemical classes identified in FTICR-MS across biotic CT replicate and autoclaved CT soil microcosms at day 0. Peaks were classified as described in **Materials and Methods**, and the relative abundance is shown across each biochemical class for each sample. Raw data is provided in **Appendix B**.



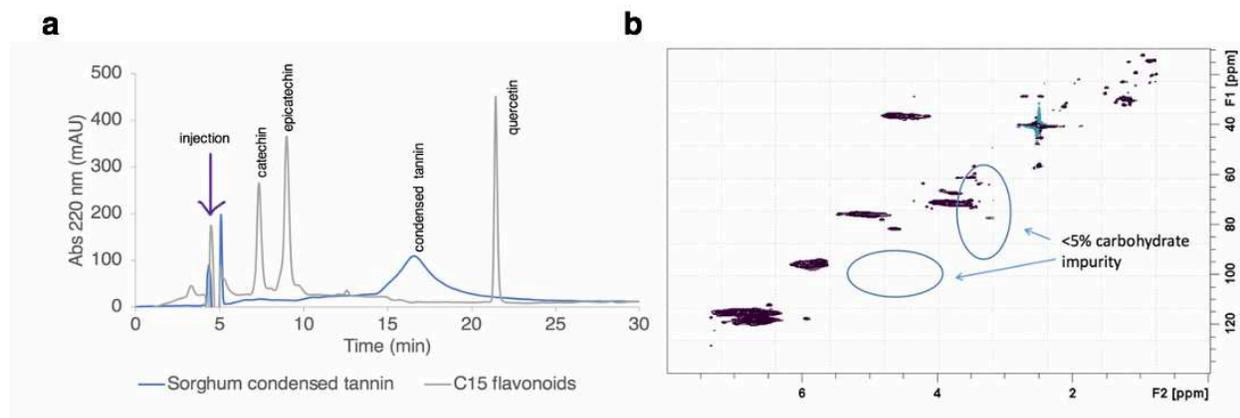
**Figure 2.4.** (a) Model CT polymers have an average degree of polymerization of 16, with repeating epicatechin (blue) units capped with a terminal catechin (magenta). Depolymerization breaks the interflavan bonds of the polymer backbone, generating smaller sized oligomers and monomers. These can be further transformed, by biotic or abiotic processes, to phenolic compounds. (b) Kendrick mass defect (KMD) hex plots for the peaks detected in replicate C of biologically-active (top) and autoclaved (bottom) CT-amended microcosms. KMD is given relative to (epi)catechin. Hex plots divide plot area into equal size hexagons, and hexagons are colored according to the number of data points that fall in that area. At left, peak information that enabled oligomer assignments is shown by colors where  $^{13}\text{C}_o$  (orange) and  $^{13}\text{C}_e$  (green) denote doubly-charged compounds containing odd and even numbers of  $^{13}\text{C}$ , respectively (see **Figure 2.6** for detailed examples). Colored rectangles are shown around regions where CT oligomers (blue) and monomers/dimers (pink) and their derived compounds are expected to occur, with key regions highlighted in purple. Corresponding colored numbers indicate oligomer sizes: monomer (1), dimer (2), trimer (3), tetramer (4), pentamer (5), and hexamer (6) peaks. Kendrick plots for all replicates at all timepoints are found in **Figure 2.8A-C**. (c-f) Metabolites detected via LC-MS can be organized into (c) CT C<sub>15</sub> monomers (epicatechin, catechin), (d) other C<sub>15</sub> flavonoids (quercetin), (e) C<sub>6</sub>-C<sub>3</sub> phenolic compounds, and (f) C<sub>6</sub>-C<sub>2</sub> phenolic compounds. Metabolite dynamics are shown with lines indicating average peak area (n=3) for CT (purple) microcosms, and shaded areas the 95% confidence interval with individual data points plotted. Dotted lines show signal from autoclaved CT-amended soil control. Orange circles indicate timepoints at which active soil signal significantly differed from autoclaved soil signals (LIMMA, p<0.05, log<sub>2</sub>FC>1.5). Vertical grey lines mark day 10. In the illustration of the monomer structure in (c), red letters label flavonoid rings, and red numbers correspond to C-ring position.



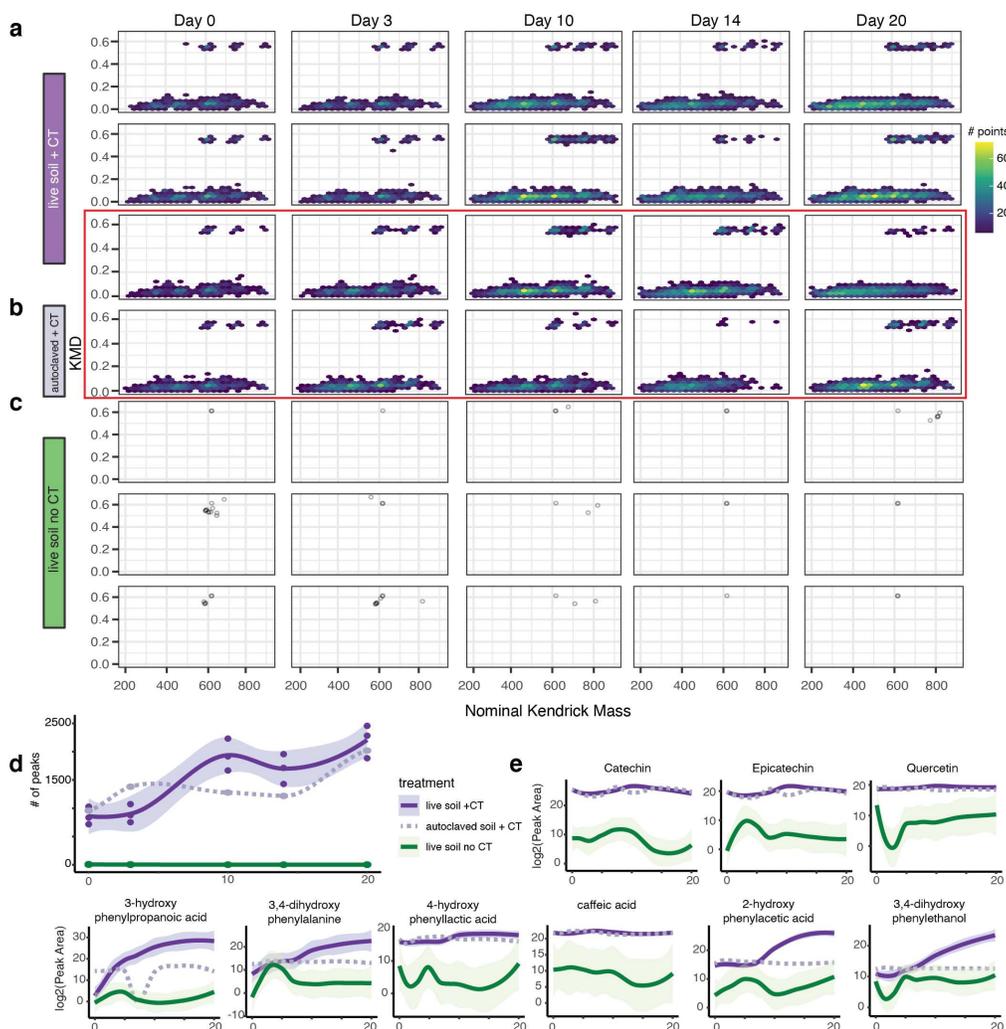
**Figure 2.5.** FTICR-MS workflow and identification of CT-derived peaks across samples. (a) We had two goals for analyzing the FT-ICRMS data: (i) identify peaks that derived from CT polymer oligomers and derived compounds (orange box), and (ii) assign identified peaks to biochemical classes (green box). This workflow is described in **Appendix A**. (b) CT polymer peaks were identified across all identified peaks by Kendrick Mass defect analysis using CH<sub>2</sub> as the base unit (**Appendix A**), and comparison across CT-amended samples (purple) and unamended samples (green). Peaks that derived predominantly from CT-amended samples (pink ovals) were assumed to correspond to CT and CT-derived compounds while majority shared peaks (blue oval) were assumed to be soil organic matter (SOM) derived.



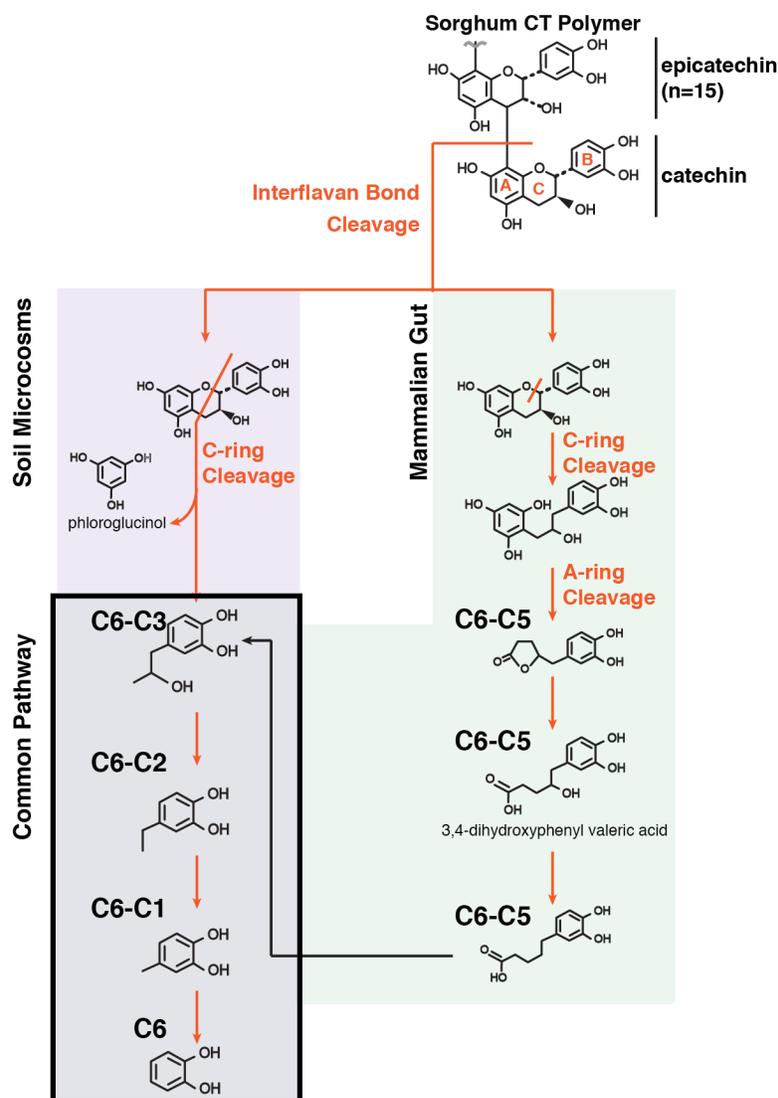
**Figure 2.6.** Identification and example Kendrick Mass Defect analysis of variably-charged CT oligomers. (a) FTICR mass spectrum of a doubly-charged CT pentamer in Live Soil +CT replicate A at day 10 (top) compared to the expected mass spectrum for a doubly-charged CT pentamer (bottom). Peak masses are indicated, and the formulas for peaks in the theoretical spectra are given. (b) The isotope abundances match expected abundances (**Appendix A**). (c) The five experimentally identified peaks (A-E) were used in KMD analysis using (epi)catechin as base, and (d) plotted in a Kendrick plot. (e-h) This same analysis was repeated with peaks corresponding to a singly charged CT trimer identified in Live Soil +CT replicate A at day 10. Comparison of the Kendrick plots for the doubly-charged (d) and singly-charged (h) isotopomers reveals isotopic splitting where KMD separates isotopes at 1/2 as in Fouquet *et al*<sup>41</sup>. Abbreviations are as follows: Nominal Mass (NM), Nominal Kendrick Mass (NKM), and Kendrick Mass Defect (KMD).



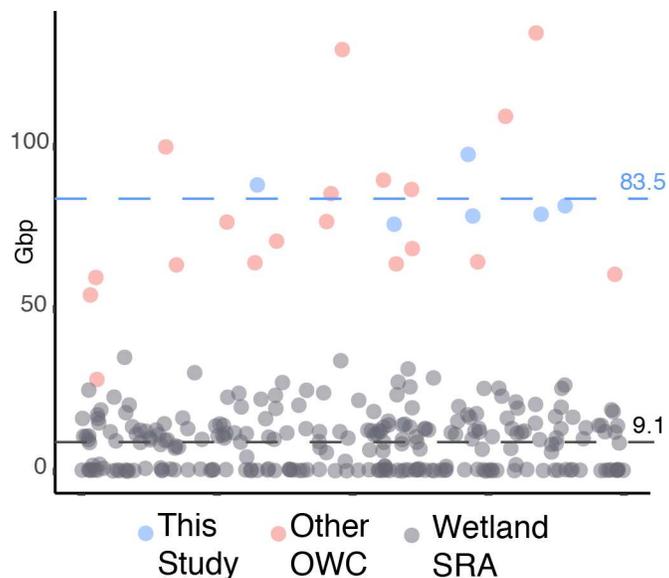
**Figure 2.7.** Characterization of the CT substrate used in this study showed minimal contamination from phenolics or carbohydrates. (a) Overlaid chromatograms for a Sorghum condensed tannin purified sample (blue) and a mixture of commercial catechin, epicatechin and quercetin (grey). The Sorghum condensed tannin elutes as a broad peak with no resolution of individual polymers or oligomers. Here, we observed the CT contained no monomers. (b) A representative  $^1\text{H}$ - $^{13}\text{C}$  HSQC NMR spectrum of purified condensed tannin from Sorghum. The blue ovals represent the areas where carbohydrate signals are found. The signals in those regions are all assigned to condensed tannin structural features as detailed in Reeves *et al.*<sup>40</sup>. For chromatography and NMR methods, see **Appendix A**.



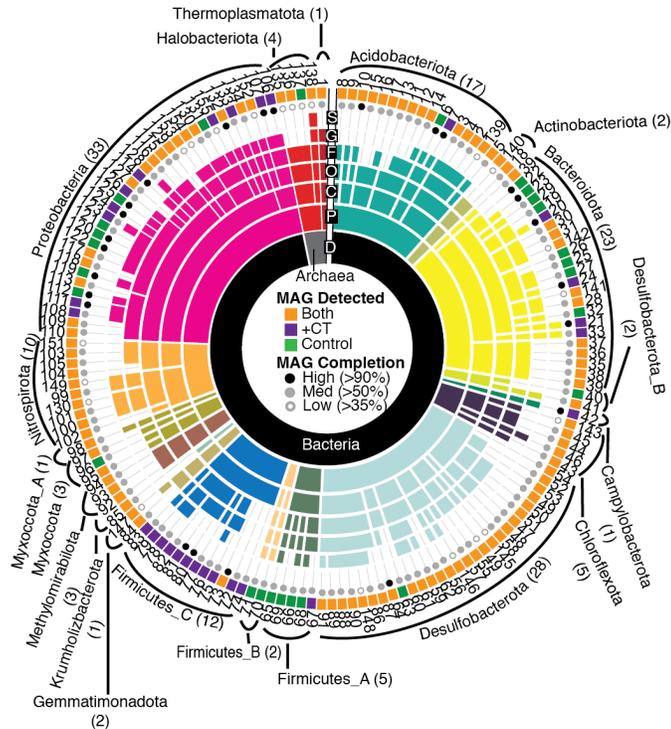
**Figure 2.8.** Unamended control soils do not contain appreciable CT oligomers or putative CT-degradation products. Kendrick mass defect plots based on (epi)catechin as the repeat unit for (a) live soil +CT replicates, (b) autoclaved soil + CT control, and (c) unamended control soil. A red box is shown around the two plots shown in Figure 2. Hex plots are used in a and b because of the high number of data points, while individual points are plotted in c. Hex plots divide plot area into equal size hexagons, and hexagons are colored according to the number of data points that fall in that area. (d) The number of CT-associated peaks are shown with lines indicating average number of peaks ( $n=3$ ) for CT (purple) and ( $n=3$ ) unamended control (green) microcosms, and shaded areas the 95% confidence interval. Dotted lines show peaks in autoclaved CT-amended soil control. (e) LC-MS identified compounds highlighted in **Figure 2.4** for all samples. Metabolite dynamics are shown with lines indicating average peak area ( $n=3$ ) for CT (purple) and ( $n=3$ ) unamended control (green) microcosms, and shaded areas the 95% confidence interval. Dotted lines show signal from autoclaved CT-amended soil controls. Raw Data is provided in **Appendix B**.



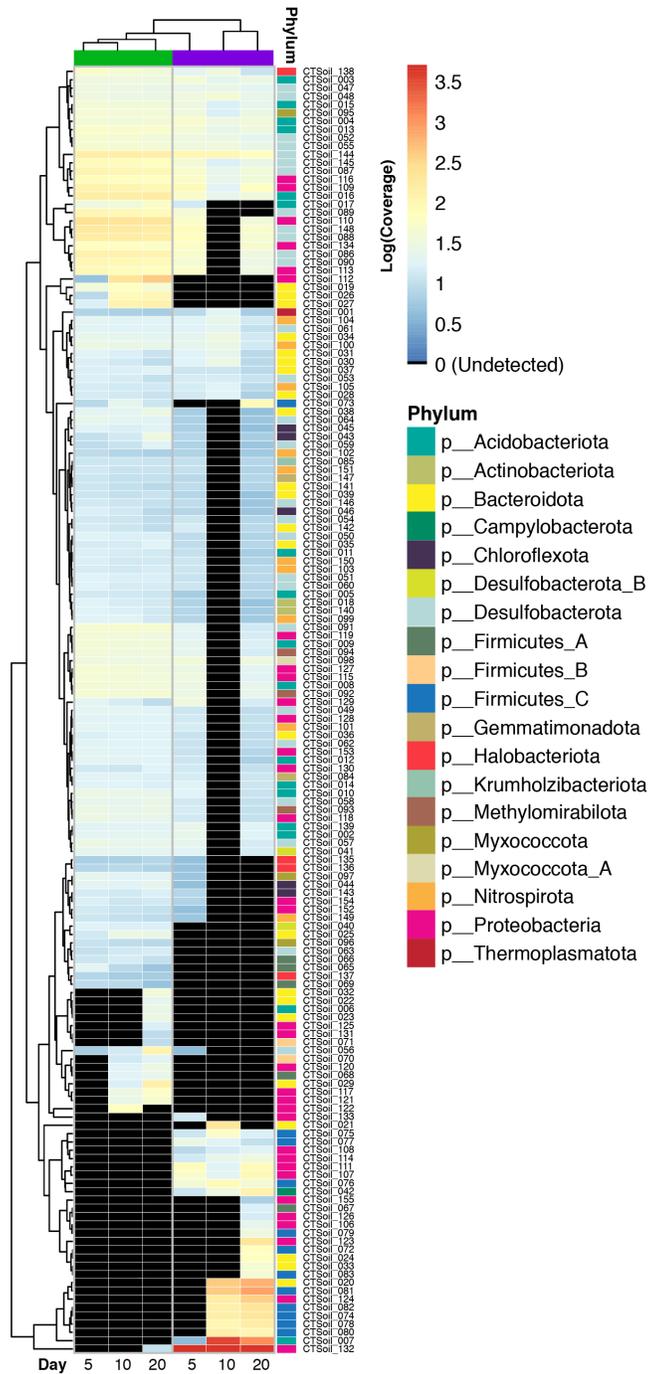
**Figure 2.9.** Comparison of proposed CT degradation pathways between mammalian guts and soil from this study. Metabolomic evidence suggests differing strategies for initial steps in CT degradation between mammalian gut and soil systems. In our soil system, depolymerization is followed by a C-ring opening to yield the C<sub>6</sub>-C<sub>3</sub> products characteristic of our reactors, while in the gut the products point to cleavage of the C-ring followed by opening of the A-ring to yield C<sub>6</sub>-C<sub>5</sub> products. These C<sub>6</sub>-C<sub>5</sub> products are suggested in gut studies to converge at C<sub>6</sub>-C<sub>3</sub> products, suggesting some commonality in degradation across systems.



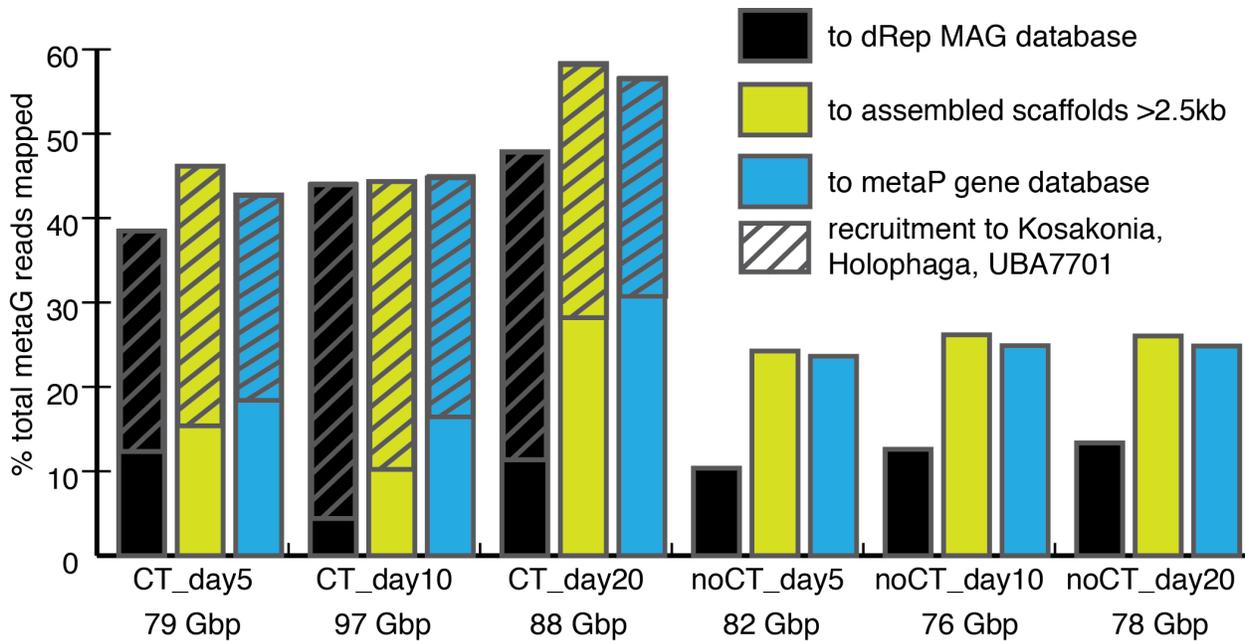
**Figure 2.10.** The depth of metagenomic sequencing in this study exceeds standard sequencing depth for wetland soil metagenomes. Distribution in sequencing (as Gigabases) for Sequence Read Archive (SRA) runs identified as “WGS” with “Wetland” in the metadata. Samples from this study are shown in blue, while other studies from our lab conducted on Old Woman Creek (OWC) samples are shown in red. All other identified SRA runs are shown in grey. The blue and grey dotted lines and numbers at left show the average sequencing per sample for this study and other wetland SRA studies, respectively.



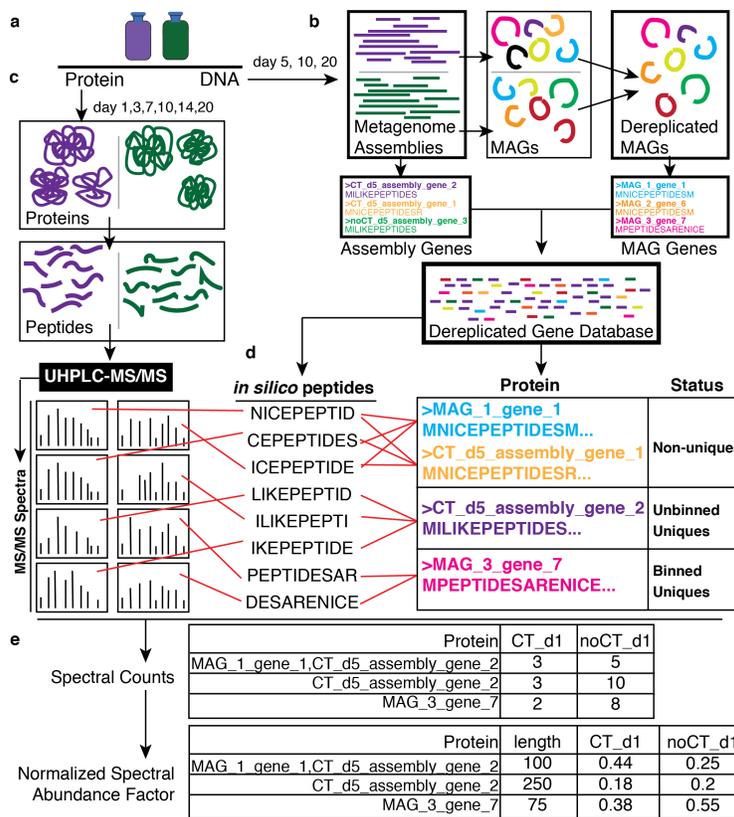
**Figure 2.11.** Taxonomy and detection of the 155 dereplicated metagenome-assembled genomes (MAGs) in our soil reactor genome database. To visualize the taxonomy of these MAGs, sequential colored rings indicated the most resolved taxonomic level that could be assigned by GTDB-tk. Taxonomic level (Domain, Phylum, Class, Order, Family, Genus, Species), is denoted in black with a single letter abbreviation. Ring color corresponded to phylum assignment, with the phylum listed on the outside with the number of dereplicated MAGs in parentheses. Circles at the sunburst edge summarized genome completion, while the listed number is the MAG ID (see **Appendix B**). Colored rectangles at the sunburst edge indicate MAG distribution across treatments, with MAGs detected (see **Materials and Methods** for thresholds) only in CT (purple), or only in control (green), or from both conditions (orange) denoted.



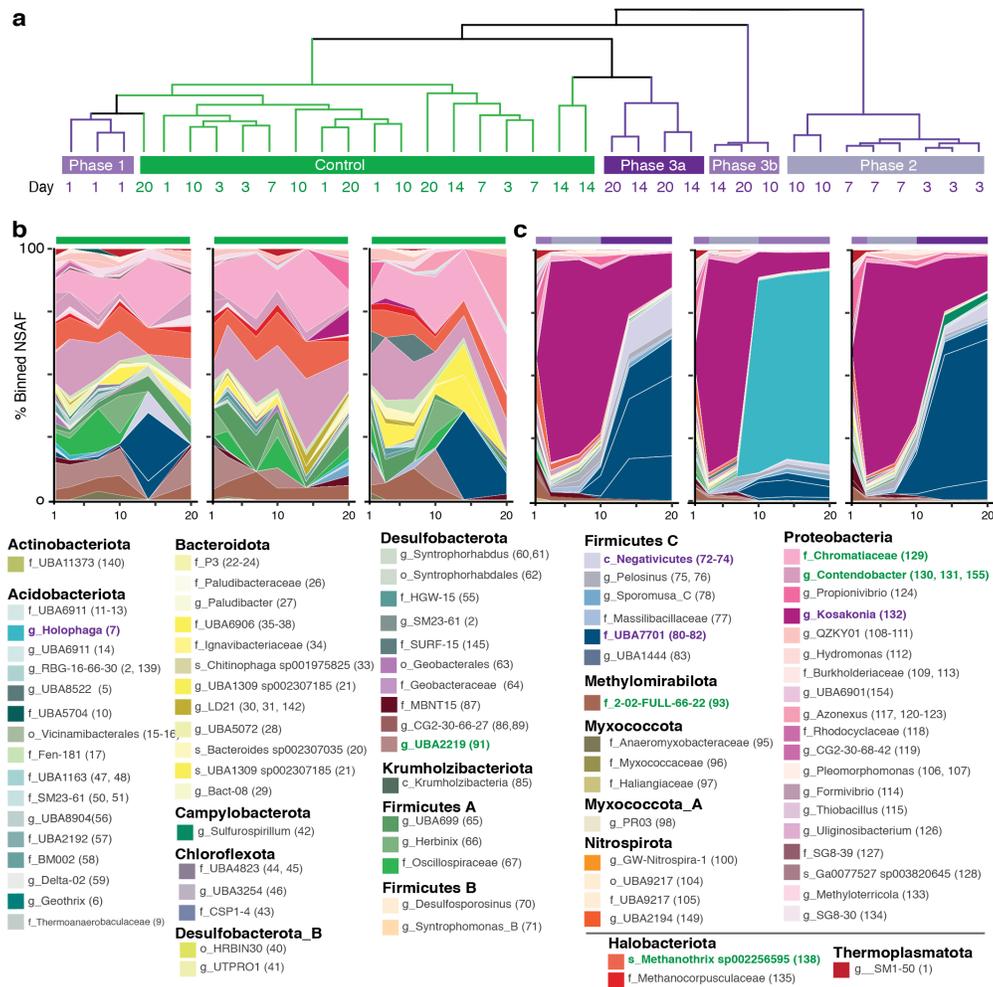
**Figure 2.12.** Metagenome-assembled genome (MAG) detection across the six metagenomes revealed core and treatment specific MAGs. The coverage of the 155 dereplicated MAGs is shown across metagenomes. Bins highlighted in black did not meet “presence” requirement of 75% scaffolds mapped at 95% identity and greater than 3X coverage. Adjacent colored boxes indicate bin Phylum, with label showing bin number. The columns highlighted with green (left) are unamended controls, while the columns highlighted with purple (right) are from CT-amended samples. The same data is reported as presence and absence in **Figure 2.11**.



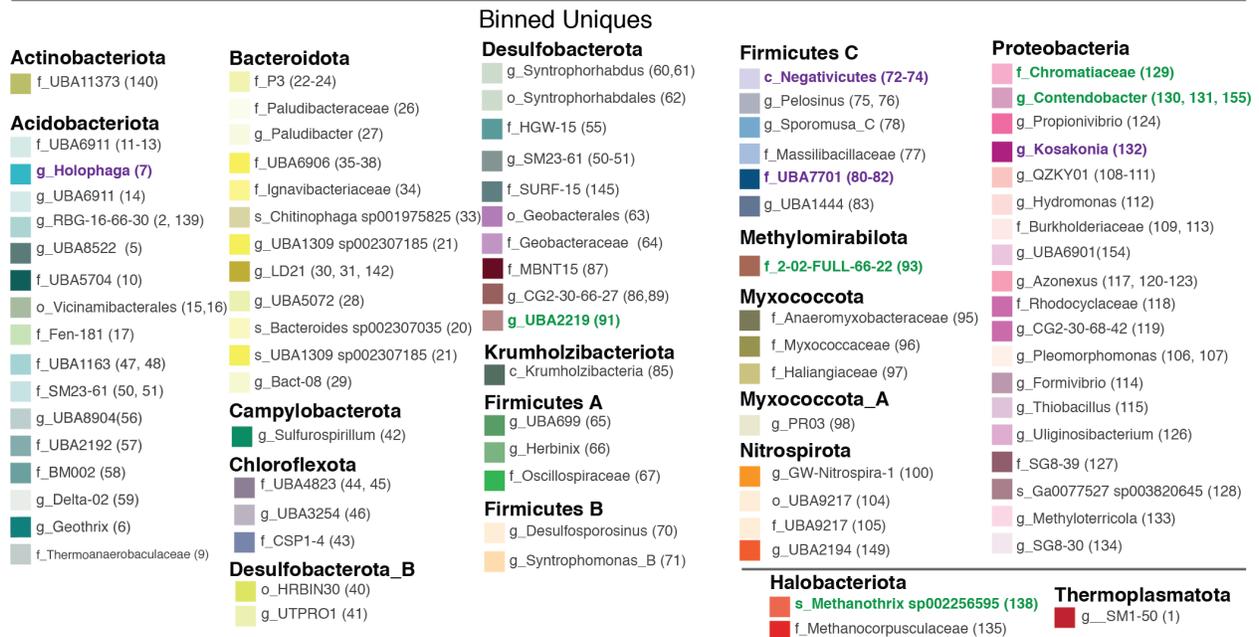
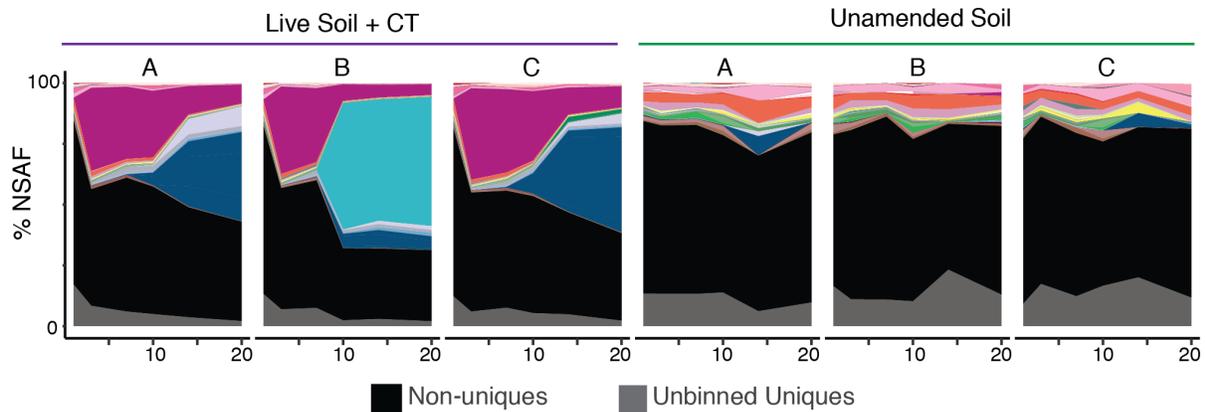
**Figure 2.13.** Metagenomic assembly and binning equivalently captures similar proportions of the microbial communities between treatments when accounting for CT-dominant members. Relative metagenomic read recruitment to the 155 MAGs (black), to all assembled scaffolds greater than 2500 bp (green), and to the metaproteome dereplicated gene database (blue) for each metagenome. Reads were mapped at 95% identity. The fractions of reads recruited to the CT-dominant and exclusive MAGs *Kosakonia*, *Holophaga*, and the 3 *Sporomusales* UBA7701 are shown in grey stripes. The size of each metagenome is listed in gigabase pairs below sample names. These findings show that after removing the dominant lineages (grey stripes), read mapping was nearly equivalent and did not appear biased by treatment (CT versus no CT).



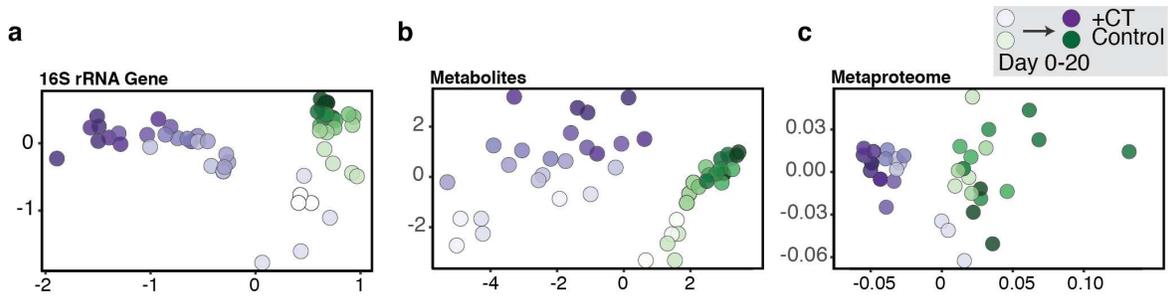
**Figure 2.14.** Schematic showing the workflow for paired metagenome sequencing and metaproteome analyses. (a) DNA and proteins were sampled from triplicate live soil + CT (purple) and unamended soil (green) reactors at days 5, 10, 20 (DNA) and days 1, 3, 7, 10, 14, 20 (protein). (b) Metagenomes at each timepoint were obtained for both CT (purple) and unamended (green) treatments. Metagenomes were assembled and binned to obtain metagenome-assembled genomes (MAGs) across all samples. This set of MAGs was dereplicated at 99% ANI to obtain a MAG database of 155 dereplicated MAGs (**Figure 2.11**). Using amino acid translations of genes derived from this dereplicated MAG database and remaining genes from metagenomic assemblies (on unbinned scaffolds >2500 bp), we compiled a Dereplicated Gene Database (all unique gene sequences) that served as our reference database for our metaproteomes. (c) Metaproteomes at each timepoint were obtained as described in **Materials and Methods**. (d) Spectral matching was carried out using obtained spectra and *in silico* spectra derived from the gene database. From this, proteins were classified as “non-unique” if the recruited peptides could be derived from other proteins in the database. Proteins were classified as “unbinned uniques” if they had peptides that could only be matched to the amino acid sequence derived from a metagenomic unbinned scaffold in our assembly. Proteins were identified as “binned uniques” if they had peptides that could only be matched to that amino acid sequence, and were derived from a single genome in our MAG database. (e) All identified proteins were quantified with label-free spectral counts. This was then corrected for protein length and sample-to-sample variation by conversion to normalized spectral abundance factor.



**Figure 2.15.** (a) Hierarchical clustering of MAG-contributions (binned uniques) to metaproteome samples. Unamended control metaproteomes are shown in green, while the multi-phase response of CT-amended microcosms are highlighted with varying shades of purple. (b-c) Genome-resolved metaproteomic dynamics in CT-amended microcosms. The relative contribution of MAGs to the binned unique peptide pool is shown for the three unamended control (b) and CT (c) microcosm replicates over 20-days. The most refined GTDB-tk assigned taxonomy is listed by phylum, with our MAG ID number in parentheses (**Appendix B**). The names of the top 5 peptide-recruiting MAGs are colored for CT (purple) and control (green) microcosms. **Figure 2.16** shows this data with the total metaproteome data, including unbinned uniques, binned uniques, and non-uniques.



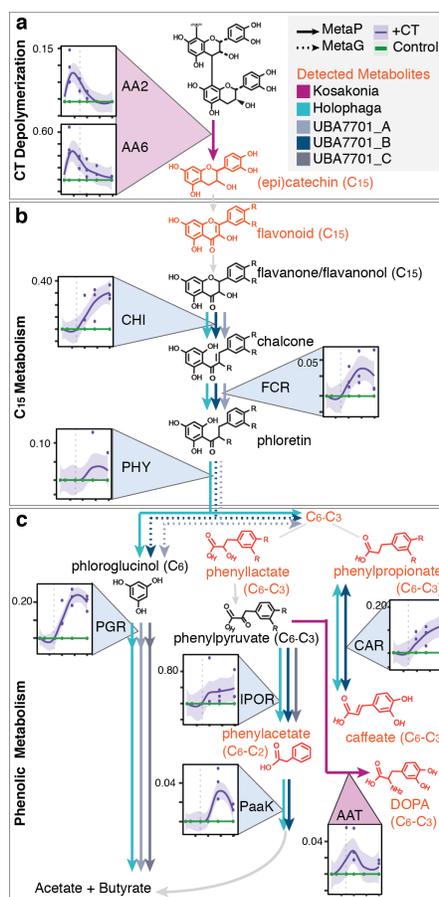
**Figure 2.16.** Relative abundance of all classes of metaproteome hits across CT-amended and unamended samples. Metaproteomic dynamics in soil microcosms shown with the relative peptide recruitment to non-unique (black), unbinned unique (grey) and binned uniques (colored) proteins. Data are graphed for CT amended and unamended control over 20-days with replicates denoted A-C. For contributions from binned uniques alone see **Figure 2.15**. The color for the binned uniques (as described in **Figure 2.15**) is denoted by the most refined GTDB-tk assigned taxonomy, with our MAG ID number in parentheses (**Appendix B**). The names of the top 5 peptide-recruiting MAGs are colored for CT (purple) and control (green) microcosms.



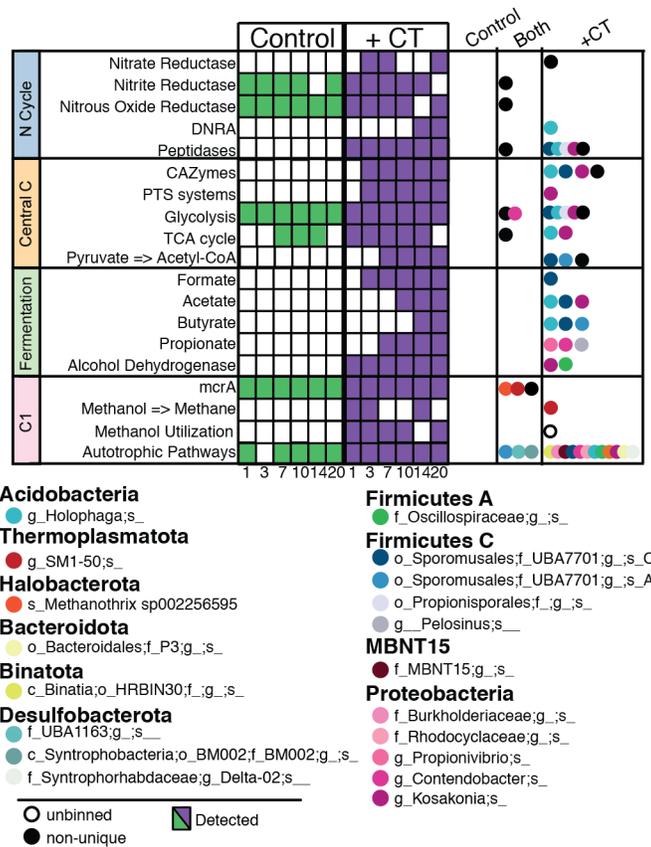
**Figure 2.17.** CT-amended and unamended microbial communities, metabolites, and metaproteomes diverge temporally. Non-metric multidimensional scaling (NMDS) of Bray-Curtis similarity metric of (a) 16S rRNA gene amplicon sequencing variants (stress = 0.09), (b) combined NMR and LC-MS exometabolites (stress=0.08), and (c) metaproteome (stress=0.11). Regardless of the data type, all show statistically significant ( $M_{rpp}$ ,  $p = 0.001$  for all 3) separation by treatment and a time trajectory. Day 0 samples are white, and proceed to darker purple (CT-amended) or green (unamended).

	CTSoil_80	CTSoil_81	CTSoil_82
CTSoil_80			
CTSoil_81	72.7%, 93.01%		
CTSoil_82	71.09%, 91.27%	72.17%, 93.45%	

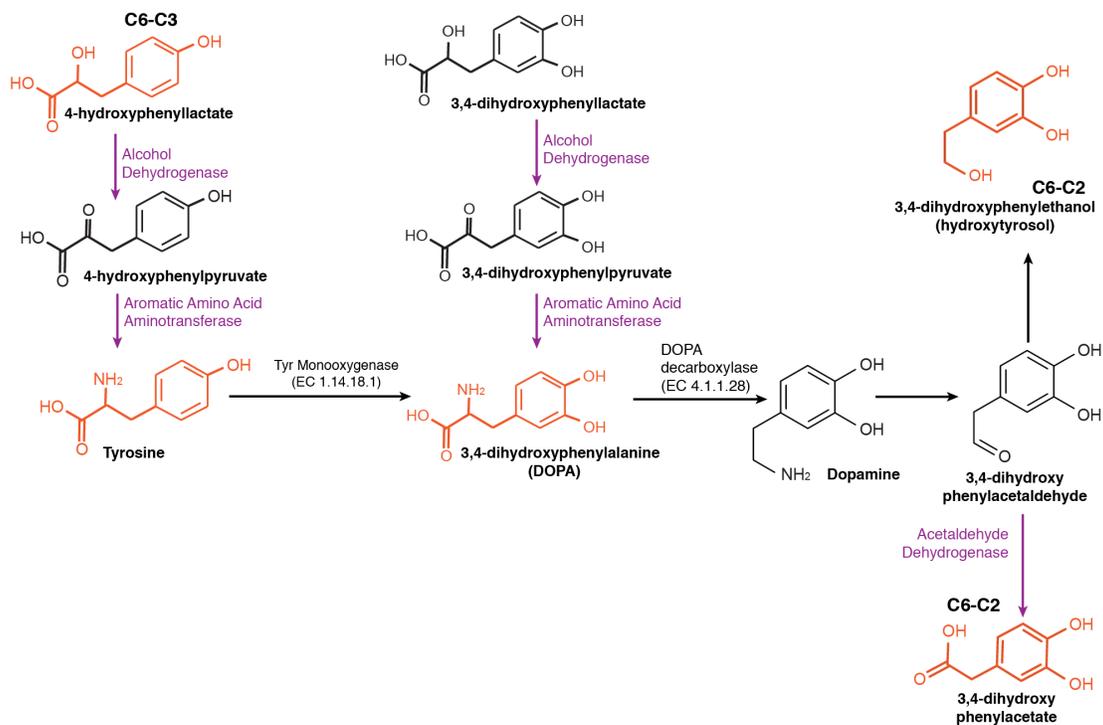
**Figure 2.18.** The three Sporomusales UBA7701 MAGs are likely three different genera. Matrix of MAG Average Amino Acid identity (AAI) and ribosomal protein S3 similarity between the three Sporomusales UBA7701 MAGs. The first number in the cell is AAI, while the second value is ribosomal protein S3 amino acid identity. MAG AAI was calculated using <http://enve-omics.ce.gatech.edu/aai/>.



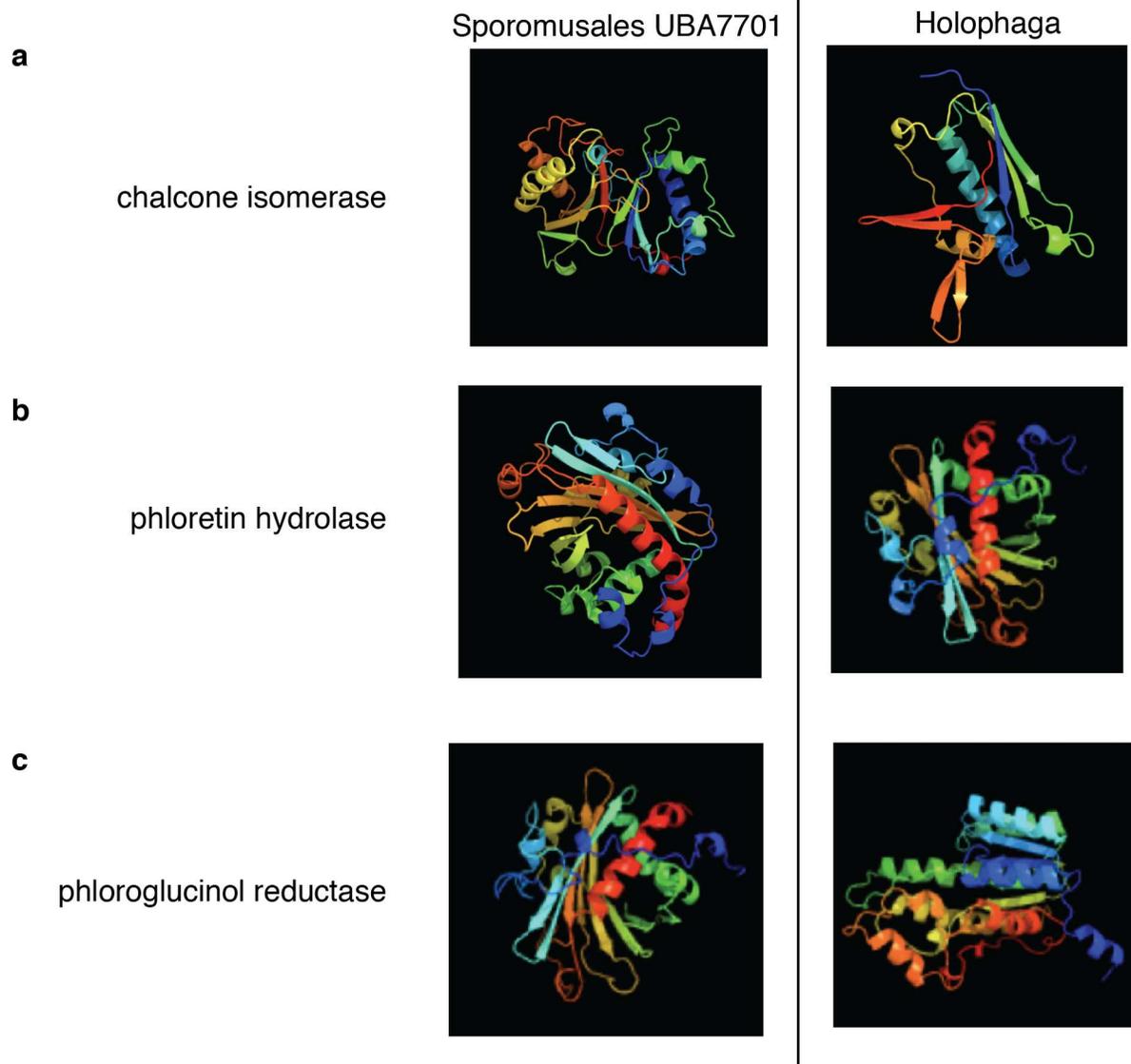
**Figure 2.19.** MAG-resolved, unique peptide evidence for polyphenol degradation. Line graphs indicate average % normalized spectral abundance factor (NSAF) for CT (purple) and unamended (green) soil microcosms, with shaded areas denoting the 95% confidence interval of triplicate measures and individual data points plotted. (a) CT-depolymerization may be mediated by peroxidase activity from AA2 and indirect activity from AA6 expression from *Kosakonia*, (b) C<sub>15</sub> biodegradation may be performed by the coordinated activity of three enzymes (CHI, FCR, PHY), these gene sets are expressed both by *Holophaga* and *Sporomusales* UBA7701 MAGs, and (c) multiple phenolic-active enzyme dynamics expressed by MAGs from these two taxa likely yield energy and produce acetate and butyrate. Enzyme abbreviations are as follows: peroxidase (AA2, EC 1.11.1.21); 1,4-benzoquinone reductase (AA6, EC 1.6.5.6), chalcone isomerase (CHI, EC 5.5.1.6), flavanone-cleaving reductase (FCR), phloretin hydrolase (PHY, EC 3.7.1.4), phloroglucinol reductase (PGR, EC 1.3.1.57), caffeoyl-CoA reductase (CAR, EC 1.3.1.108), indole-pyruvate oxidoreductase (IPOR, EC 1.2.7.8), phenylacetate-CoA ligase (PaaK, EC 6.2.1.30), and aromatic amino acid aminotransferase (AAT, EC 2.6.1.57). Dotted vertical lines are shown to mark day 7 across plots, demarcating phase 1 and 2 from phase 3. Phenolic compounds in orange are detected in metabolomics, with arrow color corresponding to MAGs expressing detected enzymes. Dotted arrows represent metagenome-encoded enzymes.



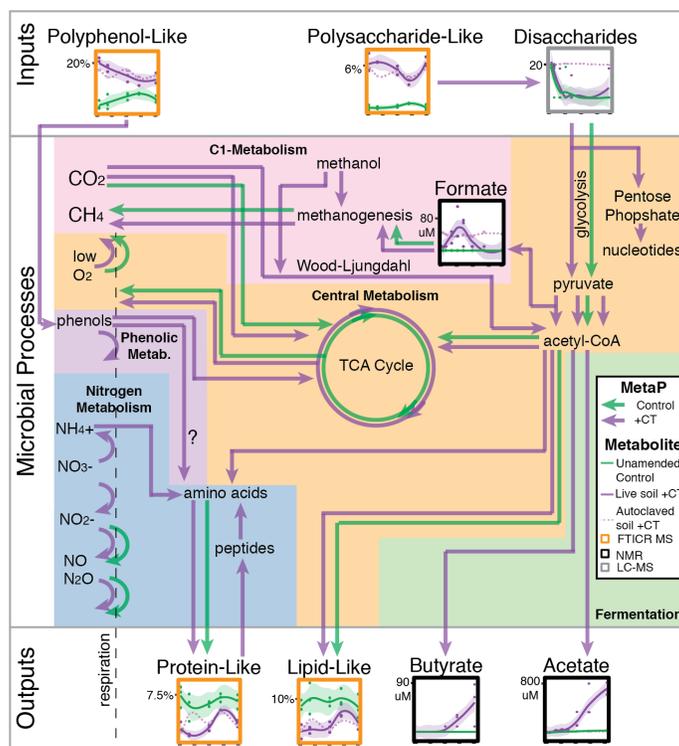
**Figure 2.20.** Genome-resolved metaproteomic dynamics across biogeochemical processes. Processes detected in at least two of three CT (purple) or unamended control (green) microcosms at each timepoint are indicated by colored squares. The genomic origins of identified functions are indicated by circles in columns corresponding to activity in either control or CT microcosms, or activity recovered in both. Proteins that could be uniquely assigned to a MAG are colored by MAG, while hits that were unique but unable to be traced to a MAG are unfilled (unbinned uniques), while proteins that were not unique are black circles (non-uniques).



**Figure 2.21.** Reconstructed metabolite pathway for polyphenol degradation via 3,4-dihydroxyphenylalanine (DOPA). Metabolites detected via LC-MS and NMR analyses are highlighted in orange, with putative enzymes detected via metaproteomics highlighted in purple. Known enzymes for tyrosine conversion to DOPA, and DOPA conversion to dopamine, are listed with EC numbers. In our metaproteomes, we recovered peptides corresponding to *Kosakonia* aromatic amino acid decarboxylase and alcohol dehydrogenase.



**Figure 2.22.** Phyre2 structural models for putative flavonoid active enzymes. Structural models are shown for (a) chalcone isomerase, (b) phloretin hydrolase, and (c) phloroglucinol reductase. Sporomusales UBA7701 enzymes are shown on the left (chalcone isomerase and phloretin hydrolase = CTSoil\_81, phloroglucinol reductase = CTSoil\_82), while *Holophaga* models are shown on the right. N-terminal residues are red, and C-terminal residues are blue. Phyre2 structural model information and gene sequences are found in **Appendix B**.



**Figure 2.23** Schematic of CT impacts to soil microbiome function provided by metaproteome-informed metabolic networks with overlaid metabolite inputs and outputs. Arrows indicate metaproteome gene expression data, with green representing unamended and purple indicating CT-amended pathways. Metabolite dynamics are shown in boxed graphs, with lines indicating average signal for live CT (n=3, solid purple), and unamended control (n=3, green) microcosms with shaded areas including the 95% confidence interval and individual data points plotted, and autoclaved CT soil (dotted light purple). The methodology used to detect the metabolite is highlighted by box color (noted in the graphical legend), with FTICR-MS data given as percent of identified peaks, NMR as  $\mu\text{mol}$ , and LC-MS as  $\log_2(\text{peak areas})$ . Nitrogen metabolism (blue box) is discussed in **Appendix A**.

## Chapter 2 References

1. Li, A. N. et al. Resources and biological activities of natural polyphenols. *Nutrients* 6, 6020–6047 (2014).
2. Hättenschwiler, S. & Vitousek, P. M. The role of polyphenols in terrestrial ecosystem nutrient cycling. *Trends Ecol. Evol.* 15, 238–243 (2000).
3. Williamson, G. The role of polyphenols in modern nutrition. *Nutrition Bulletin* 42, 226–235 (2017).
4. Correddu, F. et al. Can Agro-Industrial By-Products Rich in Polyphenols be Advantageously Used in the Feeding and Nutrition of Dairy Small Ruminants? *Animals* 10, 131 (2020).
5. Bhat, T. K., Singh, B. & Sharma, O. P. Microbial degradation of tannins - A current perspective. *Biodegradation* 9, 343–357 (1998).
6. Braune, A. & Blaut, M. Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* 7, 216–234 (2016).
7. Zak, D. et al. Unraveling the Importance of Polyphenols for Microbial Carbon Mineralization in Rewetted Riparian Peatlands. *Front. Environ. Sci.* 7, (2019).
8. Fenner, N. & Freeman, C. Drought-induced carbon loss in peatlands. *Nat. Geosci.* 4, 895–900 (2011).
9. Freeman, C., Ostle, N. & Kang, H. An enzymic “latch” on a global carbon store: A shortage of oxygen locks up carbon in peatlands by restraining a single enzymes. *Nature* 409, 149 (2001).

10. Fenner, N. & Freeman, C. Woody litter protects peat carbon stocks during drought. *Nat. Clim. Chang.* 10, 363–369 (2020).
11. Alshehri, A. et al. A Potential Approach for Enhancing Carbon Sequestration During Peatland Restoration Using Low-Cost, Phenolic-Rich Biomass Supplements. *Front. Environ. Sci.* 8, 48 (2020).
12. Fierer, N., Schimel, J. P., Cates, R. G. & Zou, J. Influence of balsam poplar tannin fractions on carbon and nitrogen dynamics in Alaskan taiga floodplain soils. *Soil Biol. Biochem.* 33, 1827–1839 (2001).
13. Brouns, K., Keuskamp, J. A., Potkamp, G., Verhoeven, J. T. A. & Heffting, M. M. Peat origin and land use effects on microbial activity, respiration dynamics and exo-enzyme activities in drained peat soils in the Netherlands. *Soil Biol. Biochem.* 95, 144–155 (2016).
14. Bonnett, S. A. F., Maltby, E. & Freeman, C. Hydrological legacy determines the type of enzyme inhibition in a peatlands chronosequence. *Sci. Rep.* 7, (2017).
15. Schmidt, M. A. et al. Soil microbial communities respond differently to three chemically defined polyphenols. *Plant Physiol. Biochem.* 72, 190–197 (2013).
16. Swenson, T. L., Karaoz, U., Swenson, J. M., Bowen, B. P. & Northen, T. R. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nat. Commun.* 9, (2018).
17. Woodcroft, B. J. et al. Genome-centric view of carbon processing in thawing permafrost. *Nature* 560, 49–54 (2018).
18. Angle, J. C. et al. Methanogenesis in oxygenated soils is a substantial fraction of wetland methane emissions. *Nat. Commun.* 8, 1–9 (2017).

19. Pinsonneault, A. J., Moore, T. R. & Roulet, N. T. Temperature the dominant control on the enzyme-latch across a range of temperate peatland types. *Soil Biol. Biochem.* 97, 121–130 (2016).
20. Narrowe, A. B. et al. Uncovering the Diversity and Activity of Methylophilic Methanogens in Freshwater Wetland Soils. *mSystems* 4, (2019).
21. Reeves, S. G. et al. Proanthocyanidin Structural Details Revealed by Ultrahigh Resolution FT-ICR MALDI-Mass Spectrometry,  $^1\text{H}$ -  $^{13}\text{C}$  HSQC NMR, and Thiolysis-HPLC–DAD. *J. Agric. Food Chem.* (2020).  
doi:10.1021/acs.jafc.0c04877
22. Kraus, T. E. C., Dahlgren, R. A. & Zasoski, R. J. Tannins in nutrient dynamics of forest ecosystems - A review. *Plant Soil* 256, 41–66 (2003).
23. Triebwasser, D. J., Tharayil, N., Preston, C. M. & Gerard, P. D. The susceptibility of soil enzymes to inhibition by leaf litter tannins is dependent on the tannin chemistry, enzyme class and vegetation history. *New Phytol.* 196, 1122–1132 (2012).
24. Halvorson, J. J., Gonzalez, J. M. & Hagerman, A. E. Repeated applications of tannins and related phenolic compounds are retained by soil and affect cation exchange capacity. *Soil Biol. Biochem.* 43, 1139–1147 (2011).
25. Buessecker, S. et al. Effects of sterilization techniques on chemodenitrification and  $\text{N}_2\text{O}$  production in tropical peat soil microcosms. *Biogeosciences* 16, 4601–4612 (2019).
26. Everette, J. D. et al. Thorough study of reactivity of various compound classes toward the folin-Ciocalteu reagent. *J. Agric. Food Chem.* 58, 8139–8144 (2010).

27. Sánchez-Rangel, J. C., Benavides, J., Heredia, J. B., Cisneros-Zevallos, L. & Jacobo-Velázquez, D. A. The Folin-Ciocalteu assay revisited: Improvement of its specificity for total phenolic content determination. *Anal. Methods* 5, 5990–5999 (2013).
28. Schofield, J. A., Hagerman, A. E. & Harold, A. Loss of tannins and other phenolics from: Willow leaf litter. *J. Chem. Ecol.* 24, 1409–1421 (1998).
29. Adamczyk, B., Kiikkilä, O., Kitunen, V. & Smolander, A. Can we measure condensed tannins from tannin-protein complexes? - A case study with acid-butanol assay in boreal forest soil organic layer. *Eur. J. Soil Biol.* 64, 40–45 (2014).
30. Li, C., Trombley, J. D., Schmidt, M. A. & Hagerman, A. E. Preparation of an acid butanol standard from fresh apples. *J. Chem. Ecol.* 36, 453–460 (2010).
31. Talbot, J. M. & Finzi, A. C. Differential effects of sugar maple, red oak, and hemlock tannins on carbon and nitrogen cycling in temperate forest soils. *Oecologia* 155, 583–592 (2008).
32. Fouquet, T. N. J. et al. On the Kendrick Mass Defect Plots of Multiply Charged Polymer Ions: Splits, Misalignments, and How to Correct Them. *J. Am. Soc. Mass Spectrom.* 29, 1611–1626 (2018).
33. Fouquet, T. N. J. The Kendrick analysis for polymer mass spectrometry. *J. Mass Spectrom.* 54, 933–947 (2019).
34. Wang, C. M., Li, T. C., Jhan, Y. L., Weng, J. H. & Chou, C. H. The Impact of microbial biotransformation of catechin in enhancing the allelopathic effects of *Rhododendron formosanum*. *PLoS One* 8, 1–14 (2013).

35. Smeriglio, A., Barreca, D., Bellocco, E. & Trombetta, D. Proanthocyanidins and hydrolysable tannins: occurrence, dietary intake and pharmacological effects. *Br. J. Pharmacol.* 174, 1244–1262 (2017).
36. Braune, A., Gütschow, M. & Blauta, M. An NADH-dependent reductase from *Eubacterium ramulus* catalyzes the stereospecific heteroring cleavage of flavanones and flavanonols. *Appl. Environ. Microbiol.* 85, 1233–1252 (2019).
37. Zhao, H., Xu, Y., Lin, S., Spain, J. C. & Zhou, N.-Y. The molecular basis for the intramolecular migration (NIH shift) of the carboxyl group during para - hydroxybenzoate catabolism. *Mol. Microbiol.* 110, 411–424 (2018).
38. Williamson, G., Kay, C. D. & Crozier, A. The Bioavailability, Transport, and Bioactivity of Dietary Flavonoids: A Review from a Historical Perspective. *Comprehensive Reviews in Food Science and Food Safety* 17, 1054–1112 (2018).
39. Monagas, M. et al. Insights into the metabolism and microbial biotransformation of dietary flavan-3-ols and the bioactivity of their metabolites. *Food Funct.* 1, 233–253 (2010).
40. Fierer, N. Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* 15, 579–590 (2017).
41. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* 35, 725–731 (2017).
42. Shade, A. et al. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* 5, (2014).

43. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927 (2020).
44. Choi, J. et al. Strategies to improve reference databases for soil microbiomes. *ISME J.* 11, 829–834 (2017).
45. Arafat, Y. et al. Soil Sickness in Aged Tea Plantation Is Associated With a Shift in Microbial Communities as a Result of Plant Polyphenol Accumulation in the Tea Gardens. *Front. Plant Sci.* 11, 601 (2020).
46. Roopchand, D. E. et al. Dietary polyphenols promote growth of the gut bacterium *Akkermansia muciniphila* and attenuate high-fat diet-induced metabolic syndrome. *Diabetes* 64, 2847–2858 (2015).
47. Janusz, G. et al. Lignin degradation: microorganisms, enzymes involved, genomes analysis and evolution. *FEMS Microbiol. Rev.* 049, 941–962 (2017).
48. Kamimura, N., Sakamoto, S., Mitsuda, N., Masai, E. & Kajita, S. Advances in microbial lignin degradation and its applications. *Current Opinion in Biotechnology* 56, 179–186 (2019).
49. Orellana, R. et al. Multi-time series RNA-seq analysis of *Enterobacter lignolyticus* SCF1 during growth in lignin-amended medium. *PLoS One* 12, e0186440 (2017).
50. DeAngelis, K. M. et al. Evidence supporting dissimilatory and assimilatory lignin degradation in *Enterobacter lignolyticus* SCF1. *Front. Microbiol.* 4, (2013).
51. Chen, Y., Huang, Z., Li, J., Su, G. & Feng, B. Complete Genome Sequence of *Kosakonia radicincitans* GXGL-4A, a Nitrogen-Fixing Bacterium with Capability to Degrade TEX. *Curr. Microbiol.* 77, 1848–1857 (2020).

52. Pletzer, D. & Weingart, H. Characterization and regulation of the Resistance-Nodulation-Cell Division-type multidrug efflux pumps MdtABC and MdtUVW from the fire blight pathogen *Erwinia amylovora*. *BMC Microbiol.* 14, 185 (2014).
53. Zoetendal, E. G., Smith, A. H., Sundset, M. A. & Mackie, R. I. The BaeSR two-component regulatory system mediates resistance to condensed tannins in *Escherichia coli*. *Appl. Environ. Microbiol.* 74, 535–539 (2008).
54. Soares, A. R. et al. The role of L-DOPA in plants. *Plant Signal. Behav.* 9, e28275 (2014).
55. Rekdal, V. M., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J. & Balskus, E. P. Discovery and inhibition of an interspecies gut bacterial pathway for Levodopa metabolism. *Science* (80-. ). 364, (2019).
56. Aura, A. M. et al. Characterization of microbial metabolism of Syrah grape products in an in vitro colon model using targeted and non-targeted analytical approaches. *Eur. J. Nutr.* 52, 833–846 (2013).
57. Westfall, S. & Pasinetti, G. M. The Gut Microbiota Links Dietary Polyphenols With Management of Psychiatric Mood Disorders. *Frontiers in Neuroscience* 13, (2019).
58. Braune, A. et al. Chalcone isomerase from *Eubacterium ramulus* catalyzes the ring contraction of flavanonols. *J. Bacteriol.* 198, 2965–2974 (2016).
59. Schoefer, L., Braune, A. & Blaut, M. Cloning and expression of a phloretin hydrolase gene from *Eubacterium ramulus* and characterization of the recombinant enzyme. *Appl. Environ. Microbiol.* 70, 6131–6137 (2004).

60. Han, J. et al. Discovery and structural analysis of a phloretin hydrolase from the opportunistic human pathogen *Mycobacterium abscessus*. *FEBS J.* 286, 1959–1971 (2019).
61. Haddock, J. D. & Ferry, J. G. Purification and properties of phloroglucinol reductase from *Eubacterium oxidoreducens* G-41. *J. Biol. Chem.* 264, 4423–4427 (1989).
62. Ozawa, Y. et al. Indolepyruvate ferredoxin oxidoreductase: An oxygen-sensitive iron-sulfur enzyme from the hyperthermophilic archaeon *Thermococcus profundus*. *J. Biosci. Bioeng.* 114, 23–27 (2012).
63. Carmona, M. et al. Anaerobic Catabolism of Aromatic Compounds: a Genetic and Genomic View. *Microbiol. Mol. Biol. Rev.* 73, 71–133 (2009).
64. Dilling, S., Imkamp, F., Schmidt, S. & Müller, V. Regulation of caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*. *Appl. Environ. Microbiol.* 73, 3630–3636 (2007).
65. Zhang, Z. et al. Priming effects of soil organic matter decomposition with addition of different carbon substrates. *J. Soils Sediments* 19, 1171–1178 (2019).
66. Brune, A., Schnell, S. & Schink, B. Sequential Transhydroxylations Converting Hydroxyhydroquinone to Phloroglucinol in the Strictly Anaerobic, Fermentative Bacterium *Pelobacter massiliensis*. *Appl. Environ. Microbiol.* 58, 1861–1868 (1992).
67. Vasta, V. et al. Plant polyphenols and rumen microbiota responsible for fatty acid biohydrogenation, fiber digestion, and methane emission: Experimental evidence and methodological approaches. *J. Dairy Sci.* 102, 3781–3804 (2019).

68. Kato, S. et al. Methanogenic degradation of lignin-derived monoaromatic compounds by microbial enrichments from rice paddy field soil. *Sci. Rep.* 5, 14295 (2015).
69. Kuusk, S. et al. Kinetics of H<sub>2</sub>O<sub>2</sub>-driven degradation of chitin by a bacterial lytic polysaccharide monooxygenase. *J. Biol. Chem.* 293, 523–531 (2018).
70. Roy Chowdhury, T. et al. Metaphenomic Responses of a Native Prairie Soil Microbiome to Moisture Perturbations. *mSystems* 4, (2019).
71. Wilhelm, R. C., Singh, R., Eltis, L. D. & Mohn, W. W. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J.* 13, 413–429 (2019).
72. Hatzepichler, R. et al. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal - bacterial consortia. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4069–E4078 (2016).
73. Pett-Ridge, J. & Weber, P. K. NanoSIP: NanoSIMS applications for microbial biology. *Methods Mol. Biol.* 881, 375–408 (2012).
74. Bailey, V. L. et al. Micrometer-scale physical structure and microbial composition of soil macroaggregates. *Soil Biol. Biochem.* 65, 60–68 (2013).
75. Min, K., Freeman, C., Kang, H. & Choi, S. U. The Regulation by Phenolic Compounds of Soil Organic Matter Dynamics under a Changing Environment. *Biomed Res. Int.* 2015, (2015).
76. Quideau, S., Deffieux, D., Douat-Casassus, C. & Pouységu, L. Plant polyphenols: Chemical properties, biological activities, and synthesis. *Angew. Chemie - Int. Ed.* 50, 586–621 (2011).

77. Hagerman, A. E. Fifty Years of Polyphenol-Protein Complexes. in *Recent Advances in Polyphenol Research* 3, 71–97 (Wiley-Blackwell, 2012).
78. Porter, L. J., Hrstich, L. N. & Chan, B. G. The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin. *Phytochemistry* 25, 223–230 (1985).
79. Tannin Handbook - Ann E. Hagerman. Available at:  
<http://www.users.miamioh.edu/hagermae/>. (Accessed: 2nd July 2020)
80. Wrighton, K. C. et al. Bacterial community structure corresponds to performance during cathodic nitrate reduction. *ISME J.* 4, 1443–1455 (2010).
81. van Trump Ian J., I. J. et al. Humic acid-oxidizing, nitrate-reducing bacteria in agricultural soils. *MBio* 2, 44–55 (2011).
82. Borton, M. A. et al. Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. *Proc. Natl. Acad. Sci. U. S. A.* 115, E6585–E6594 (2018).
83. Lovley, D. R. & Phillips, E. J. P. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Appl. Environ. Microbiol.* 54, 1472–1480 (1988).
84. Gregory, K. B., Bond, D. R. & Lovley, D. R. Graphite electrodes as electron donors for anaerobic respiration. *Environ. Microbiol.* 6, 596–604 (2004).
85. Caporaso, J. G. et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6, 1621–1624 (2012).
86. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583 (2016).

87. Joshi, N. & Fass, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle>. 2011 (2011).
88. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428 (2012).
89. Kang, D. D. et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019, e7359 (2019).
90. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28, 1033–1034 (2012).
91. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055 (2015).
92. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* 11, 2864–2868 (2017).
93. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* 48, 8883–8900 (2020).
94. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858 (2015).

95. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. (2014). Available at: [sourceforge.net/projects/bbmap/](https://sourceforge.net/projects/bbmap/). (Accessed: 2nd July 2020)
96. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079 (2009).
97. CoverM: Read coverage calculator for metagenomics. Available at: <https://github.com/wwood/CoverM>. (Accessed: 2nd July 2020)
98. Kim, S. & Pevzner, P. A. MS-GF+ makes progress towards a universal database search tool for proteomics. *Nat. Commun.* 5, 1–10 (2014).
99. Herold, M. et al. Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. *Nat. Commun.* 11, (2020).
100. Diamond, S. et al. Mediterranean grassland soil C–N compound turnover is dependent on rainfall and depth, and is mediated by genomically divergent microorganisms. *Nat. Microbiol.* 4, 1356–1367 (2019).
101. Schneider, T. et al. Who is who in litter decomposition Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J.* 6, 1749–1762 (2012).
102. Hultman, J. et al. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212 (2015).
103. Von Bergen, M. et al. Insights from quantitative metaproteomics and protein stable isotope probing into microbial ecology. *ISME J.* 7, 1877–1885 (2013).
104. Zorz, J. K. et al. A shared core microbiome in soda lakes separated by large distances. *Nat. Commun.* 10, 1–10 (2019).

- 105.Hawley, A. K., Brewer, H. M., Norbeck, A. D., Paša-Tolić, L. & Hallam, S. J. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc. Natl. Acad. Sci. U. S. A.* 111, 11395–11400 (2014).
- 106.Neilson, K. A., Keighley, T., Pascovici, D., Cooke, B. & Haynes, P. A. Label-free quantitative shotgun proteomics using normalized spectral abundance factors. *Methods in Molecular Biology* 1002, 205–222 (2013).
- 107.Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G. & Qian, K Kendrick mass defect spectrum: A compact visual analysis for ultrahigh resolution broadband mass spectra. *Anal. Chem.* 73, 4676–81 (2001).
- 108.Sato, H., Nakamura, S., Teramoto, K. & Sato, T. Structural characterization of polymers by MALDI spiral-TOF mass spectrometry combined with kendrick mass defect analysis. *J. Am. Soc. Mass Spectrom.* 25, 1346–1355 (2014).
- 109.Tolić, N. et al. Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra. *Anal. Chem.* 89, 12659–12665 (2017).
- 110.Tfaily, M. M. et al. Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Anal. Chim. Acta* 972, 54–61 (2017).
- 111.Yao, Y. et al. Analysis of metabolomics datasets with high-performance computing and metabolite atlases. *Metabolites* 5, 431–442 (2015).
- 112.Bowen, B. P. & Northen, T. R. Dealing with the unknown: Metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* 21, 1471–1476 (2010).

113.Smyth, G. K. Limma: linear models for microarray data. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (2005).

114.Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.* 47, (2019).

## Chapter 3 – CAMPER: Curated Annotations for Microbial Polyphenol Enzymes and Reactions<sup>2</sup>

### 3.1 Overview

Curated Annotations for Microbial (Poly)phenol Enzymes and Reactions (CAMPER) is a tool that annotates genes likely involved in transforming polyphenols and provides chemical context for these transformations in a summarized form (**Figure 3.1**). CAMPER aims to address a blind spot in microbial metabolism. It is currently challenging to infer polyphenol metabolism from genomic data because:

1. Genes encoding biochemically characterized enzymes have not been propagated into or labelled in annotation databases
2. Genes in databases can often be involved in multiple pathways, requiring expert knowledge to establish polyphenol context

These challenges limit widespread understanding of the transformation of polyphenolic compounds across environments.

To facilitate the inference of polyphenol metabolism from genomes, CAMPER includes 8 Hidden-Markov Model (HMM) profiles and 33 Basic Local Alignment Search Tool (BLAST) searches for (poly)phenol-active genes. The citations and amino acid sequences used to generate these homology based annotations are provided (**Appendix C**). We also provide recommended score cut-offs for searches using two

---

<sup>2</sup> This chapter was reproduced verbatim from “McGivern, B, Flynn, R, CAMPER, (2022), GitHub repository, <https://github.com/WrightonLabCSU/CAMPER>”. The text benefited from writing and editing contributions from contributing authors. The ordering of the materials in this dissertation are consistent with the content available online but have been renumbered to reflect incorporation into this dissertation.

ranks: a more stringent, trusted rank (A) and a more relaxed, exploratory rank (B). Beyond these 41 newly developed annotations, nearly 300 other known polyphenol annotations from other databases (KEGG, dbCAN) are included in the CAMPER summary.

CAMPER summarizes the gene annotations into 101 modules representing different polyphenol transformations. These modules can be as small as a single gene, or can include up to a maximum of 12 genes in the largest module (**Figure 3.2**). Following Phenol-Explorer categories, these modules are classified by the established polyphenol family and sub-family. Then, the modules are classified by the oxygen requirements for the reactions encoded: (a) oxic for reactions that are either known to require oxygen, or have only been characterized under oxygenic conditions, (b) anoxic for reactions characterized under anoxic conditions, or (c) both for reactions reported to operate with or without oxygen. It is important to note that as our knowledge of biochemistry changes, these oxygen-classifications could change. The complete CAMPER road map shows the metabolic reactions encoded (**Figure 3.3**). There are currently three ways to run CAMPER, depending on the end users goals.

### **3.2 CAMPER Data**

The CAMPER data set consists of 5 files (**Appendix C**), each serving a key role in reproducibly annotating gene data:

1. CAMPER\_blast.fa: A fasta file of CAMPER genes used as a target in a BLAST style search provided by mmseqs search.
2. CAMPER.hmm: A HMM file used as the target in an HMM profile search provide by MMseqs profilesearch

3. CAMPER\_blast\_scores.tsv: Provides the minimum cut off scores for search results and quality ranks with BLAST style searches.
4. CAMPER\_hmm\_scores.tsv: Provides the minimum cut off scores for search results and quality ranks with HMM Profile searches.
5. CAMPER\_distillate.tsv: A custom distillate, for use with DRAM or with CAMPER\_DRAMKit, to summarize the annotation results.

### **3.3 Using CAMPER with DRAM**

This functionality will be available in next major release of DRAM, so CAMPER can be integrated into your regular genome annotation pipeline. The DRAM version will have the benefit of summarizing 300 annotations derived from KEGG and dbCAN databases, if you have these databases installed, in addition to the 41 CAMPER annotations. This will provide curated annotation and summarization of polyphenol transformation genes in addition to the regular DRAM databases.

There are two steps to running CAMPER in DRAM: (1) annotation and (2) summarization (distillation, in DRAM terminology). Supply the `--use_camper` flag during the annotation step, like so:

```
DRAM.py annotate --use_camper -i 'my_bins/*.fa' -o DRAM_wCAMPER
```

```
DRAM.py distill -i DRAM_wCAMPER/annotations.tsv -o DRAM_wCAMPER_distilled
```

The difference in outputs between DRAM with CAMPER and default DRAM is that you will find CAMPER-specific columns added to the annotations.tsv and you will find a CAMPER tab in your metabolism\_summary.xlsx output. For descriptions of the content in output files, see the CAMPER Outputs section below.

### **3.4 CAMPER standalone tool: CAMPER\_DRAMKit**

If your goal is to only look at CAMPER annotations, we recommend running CAMPER\_DRAMKit, a standalone tool that provides curated annotation and summarization of polyphenol transformation genes. Unless you run with DRAM as described below and have KEGG and dbCAN databases, here you will only get the 41 custom annotations within CAMPER. CAMPER\_DRAMKit is really a smaller version of DRAM that follows much the same workflow as DRAM and has similar capabilities. There are three ways to set up the CAMPER\_DRAMKit:

#### **3.4.1 Set up with Conda**

The simplest way to get started with the CAMPER\_DRAMKit is with Conda, using the environment.yaml provided in this repository. CAMPER\_DRAMKit comes with the latest version of CAMPER preloaded, so if all you want to do is annotate and distill called genes with CAMPER, you only need the following commands.

```
wget https://github.com/WrightonLabCSU/CAMPER/main/CAMPER_DRAMKit/environment.yaml  
conda create --name CAMPER -f ./environment.yaml
```

However, if you want to have all of CAMPER at your fingertips, you can use the following command to both download this repository, including all the CAMPER data and install the CAMPER\_DRAMKit tool.

```
git clone https://github.com/WrightonLabCSU/CAMPER.git  
cd CAMPER/CAMPER_DRAMKit  
conda create --name CAMPER -f ./environment.yaml
```

In both cases, you can activate the newly made environments with the command:

```
conda activate CAMPER
```

Provided all things have gone smoothly, you will be able to activate this environment at any time and use any of the commands outlined in the Usage section below. If there are any problems, please open an issue in the GitHub repository.

### 3.4.2 Setup with pip

If you are not able to use Conda, you can still install CAMPER\_DRAMKit with pip using the command below. Note that first you will need to manually install scikit-bio, and MMseqs2, as these tools can't be installed with the other pip dependencies.

```
pip install camper_dramkit
```

### 3.4.3 Installing with DRAM

If you intend to use CAMPER\_DRAMKit with DRAM, it may be expedient to install them in the same Conda environment. This is easy to do if you have already made a DRAM Conda environment with the instructions in the DRAM README, then you can add CAMPER with the following commands:

```
wget https://github.com/WrightonLabCSU/CAMPER/main/CAMPER_DRAMKit/environment.yaml  
conda env update --name DRAM -f ./environment.yaml
```

If you install CAMPER\_DRAMKit, you will get the latest version of the CAMPER database with it. If you want more control over the database, you can override the default data with the instructions in Other Tools and flags.

### 3.4.4 Using CAMPER\_DRAMKit

Once installed, CAMPER\_DRAMKit will provide three commands: `camper_annotate`, `camper_distill` and `combine_annotations_lowmem`. These commands alongside DRAM enable a variety of workflows. The simplest workflow is the two-step annotation and summarization of a single amino acid fasta file. An example of such a workflow is shown below:

```
camper_annotate -i my_genes.faa -o my_output
```

```
camper_distill -i my_output/annotations.tsv -o my_output/distillate.tsv
```

These commands will make two files in the output directory (above named my\_output, but this is customizable): annotations.tsv and distillate.tsv (or whatever you name it in your -o command). For descriptions of these files, see the CAMPER Outputs section below.

An alternative workflow is to use in combination with DRAM. As previously stated, DRAM1.4.0 will include CAMPER by default as an easy tool, but it is possible to use CAMPER with any version of DRAM after 1.3, with one additional command. First follow the instructions above to update your DRAM environment with CAMPER\_DRAMKit. Then with that environment activated, you should be able to run the following commands to make a new raw annotations file with all the DRAM data you expect, and the CAMPER data added in. If you are not able to update your DRAM environment the environments will need to be switched mid-workflow.

```
DRAM.py annotate -i 'my_bins/*.fa' -o dram_output
```

```
camper_annotate -i my_genes.faa -a dram_output_annotation -o camper_dram_output
```

This will create a new set of raw annotations with CAMPER data added, in this case the path of the new file will be camper\_dram\_output/annotations.tsv. Then, use the camper\_distill command to get a distillate with all the key genes from both DRAM and CAMPER.

```
camper_distill -i camper_dram_output/annotations.tsv -o
```

```
camper_dram_output/camper_distillate.tsv
```

For descriptions of the annotations.tsv and summary file, see the CAMPER Outputs section below.

### 3.4.5 Other Tools and flags

To further customize the analysis workflow, there are a few more options in the CAMPER\_DRAMKit package:

1. **Combine Annotations With low memory:** You may want to re-annotate many existing DRAM annotation files, possibly from more than one version of DRAM. To this end we include the `combine_annotatons_lowmem` command which should combine many annotation files quickly and with a small memory footprint, even if they come from different versions of DRAM. The command is used like so:

```
Combine_annotatons_lowmem -i /path/to/many/dramfolders/*/annotations.tsv -o  
combined_annotation.tsv
```

The input path needs to be a wild card pointing to a set of DRAM annotation files, this is passed to the python glob command. This format should be familiar to bash formatting and it can be tested with the ls command.

2. **Manually Specifying the Location of CAMPER Files:** The behavior of the `camper_annotate` and `camper_distill` commands is controlled by the latest version of the CAMPER dataset. If you want to use an older version of CAMPER, it is suggested you install the older version of the CAMPER\_DRAMKit tool, as they will be released together and be mutually compatible. However, if you must, you can also specify the files to use with `camper_annotate` and `camper_distill` using the appropriate arguments. An example is shown below.

```
camper_annotate -i my_genes.faa -o my_output \  
--camper_fa_db_loc CAMPER_blast.fa \  
--camper_fa_db_cutoffs_loc CAMPER.hmm \  

```

```
--camper_hmm_loc CAMPER_blast_scores.tsv \  
--camper_hmm_cutoffs_loc CAMPER_hmm_scores.tsv  
camper_distill -i my_output/annotations.tsv -o my_output/distillate.tsv \  
--camper_distillate CAMPER_distillate.tsv
```

Remember that running any script with the `--help` flag will provide more information. Also note that if one or more arguments is not specified, the default data will be used.

### **3.5 I just want to run your BLAST and HMM searches on my own!**

We get that sometimes this is all you want to do! This is the simplest way to use our annotations. See the above CAMPER Data section and download the CAMPER\_blast.fa and CAMPER.hmm files. These can be run using blast, hmmsearch, or mmseqs2 searches of the data, for example:

```
makeblastdb -in CAMPER_blast.fa -dbtype prot  
blastp -query my_genes.fa -db CAMPER_blast.fa -out BLAST_my_genes_CAMPER.txt -outfmt 6  
hmmsearch --tblout hmmsearch_my_genes_CAMPER.txt CAMPER.hmm my_genes.faa
```

We strongly recommend curating these outputs with the scores given in the CAMPER\_blast\_scores.tsv and CAMPER\_hmm\_scores.tsv files for each profile search.

### **3.6 CAMPER Outputs**

Approaches 1 and 2 output two files: the raw information for given searches (annotations.tsv) and the summarized information across searches (the distillate, either the metabolism\_summary.xlsx if run through DRAM or the distillate.tsv from CAMPER\_DRAMKit).

### 3.6.1 Raw annotations

This is either a standalone file, or columns added to a file, depending on search approach. This file annotates the genes in your dataset that pass CAMPER score thresholds, reports the annotation from the databases, and the CAMPER scores. It includes the following columns:

- `camper_hits`, A longer ID giving the CAMPER ID, gene abbreviation, and gene description.
- `camper_rank`, A match quality rank based on the value of the bit score (A or B). For BLAST-style searches, an A rank is a bitscore  $\geq 200$  and B  $\geq 120$ . For HMM-style searches, scores are specific to each profile (see `CAMPER_hmm_scores.tsv`).
- `camper_bitScore`, The bitscore from the best search result. If more than one search meets at least a B-rank for a given gene, the search with the higher score is reported.
- `camper_id`: Unique CAMPER ID used in the distillation step, of the form D000XX.
- `camper_definition`: A short description of the CAMPER match in the database.
- `camper_search_type`: Tells you if a HMM profile or blast search found this match.

### 3.6.2 Distillate

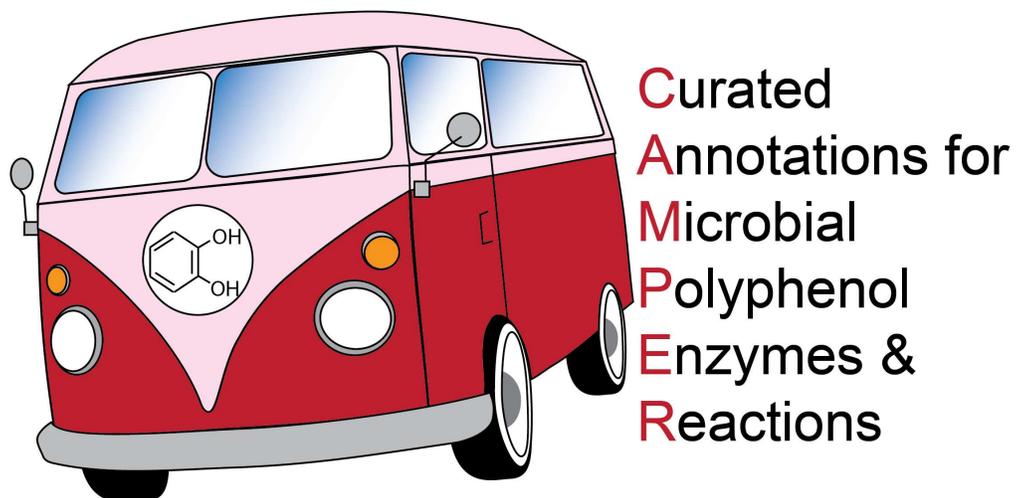
This is either a single file, or the CAMPER tab in the `metabolism_summary.xlsx` file. Each row in this file corresponds to a gene in a CAMPER module. This file gives you gene counts of genes in CAMPER modules. It includes the following columns:

- `gene_id`, the database IDs assigned to this gene. These can be from CAMPER (D000XX), KEGG (KXXXXX), dbCAN (AAX), or EC numbers. Note, some IDs are

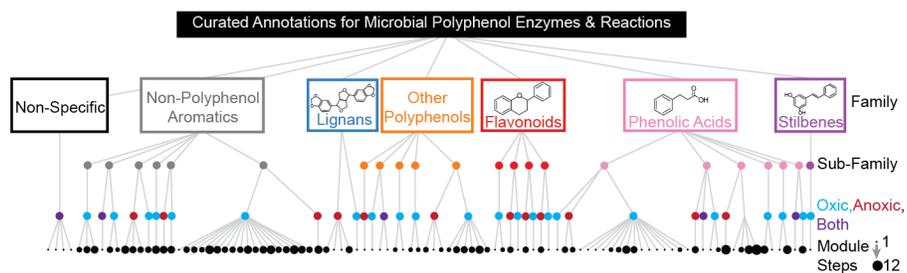
included more than once in the sheet if they are involved in more than one module!

- `gene_description`, A more informative description of the gene in the step including gene abbreviation and gene name.
- `module`, The CAMPER module that the given gene belongs to. There are 101 modules in CAMPER.
- `header`, The classification for the polyphenol substrate following Phenol-Explorer Ontology. In the form: Polyphenol;Family;Sub-Family;Compound.
- `subheader`, This contains information about routes, steps, and subunits. Sometimes, a given transformation can be accomplished in more than one sequence of steps: these are termed 'Routes'. Steps indicate the sequential transformations in the module. Subunits denote if the given gene encodes a subunit of a larger complex that carries out a step. Sometime steps are labelled as "optional" if they are not required.
- `specifc_reaction`, This gives examples of reactions when possible.
- `oxygen`, This is either "oxic", "anoxic", "or "both" for reactions that require oxygen, don't require oxygen, or can function with or without, respectively. Note: these are largely based on literature reporting and the systems they were characterized in, and should be used as guidelines.
- `EC`, The EC number (if known) for a reaction.
- `Notes`, Any important information to know, for example: should they be extracellular etc.

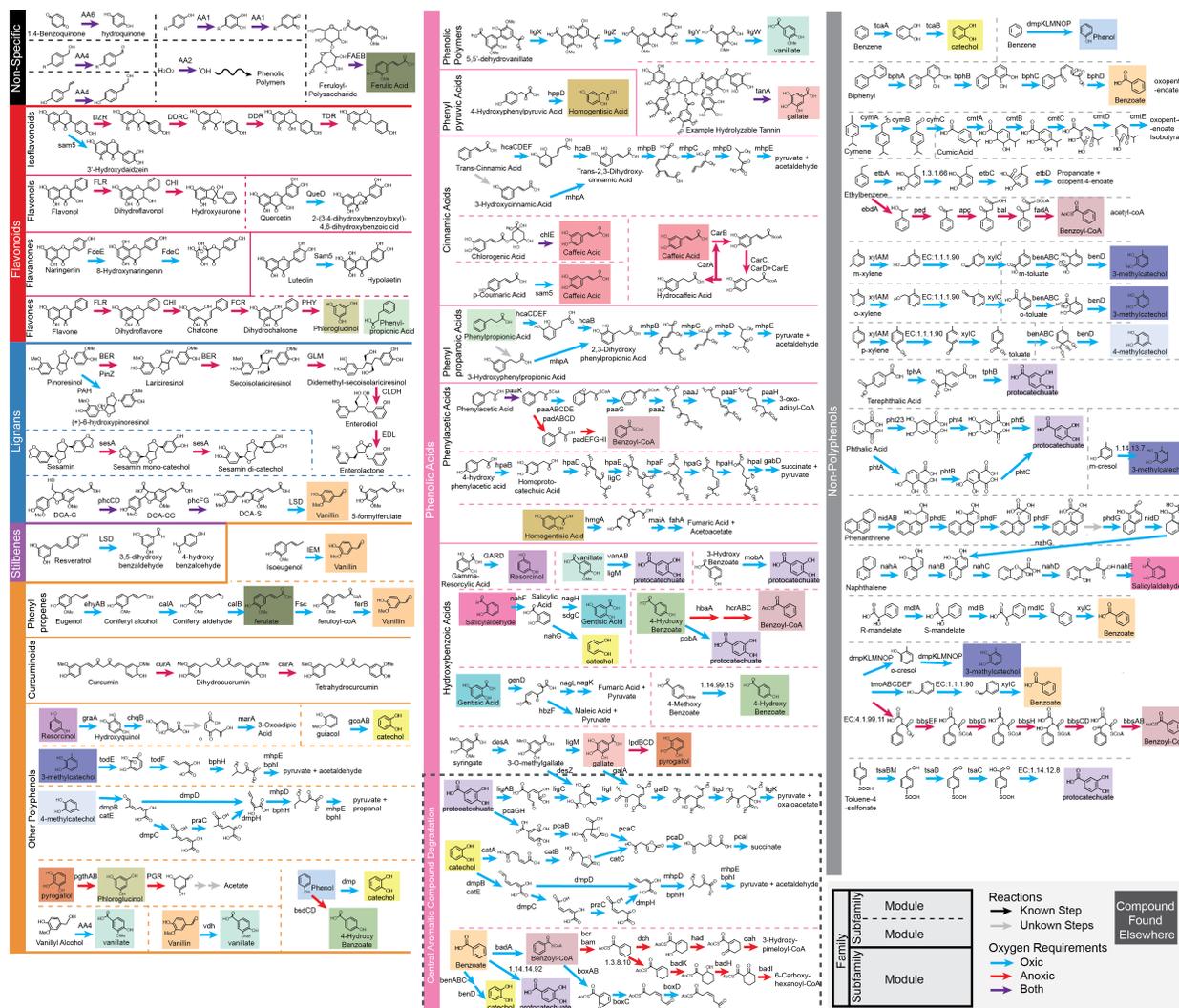
Happy CAMPER-ing!



**Figure 3.1.** Curated Annotations for microbial polyphenol enzymes and reactions (CAMPER) logo.



**Figure 3.2.** CAMPER consists of 101 polyphenol transformation modules, organized by substrates (Family and Subfamily) and oxygen requirements. Modules can be composed of 1-12 reaction steps.



## Chapter 4 – Auditing the enzyme latch unveils polyphenols as unbudgeted microbial carbon substrates in northern peatlands

### **4.1 Summary**

A prevailing theory in arctic peatlands is that polyphenols and phenolic compounds limit microbial decomposition, in effect sequestering soil carbon and limiting carbon dioxide (CO<sub>2</sub>) flux. The theory is founded on the assumption that polyphenol degradation is mediated by a single oxygen-requiring enzyme known as polyphenol oxidase, though molecular evidence supporting this assertion is lacking. In Chapter 3 I described a polyphenol metabolism annotator that I used in this study to assess polyphenol biotransformations in multi-omics microbiome data along a permafrost thaw gradient. I revealed (i) a positive relationship between soil polyphenols and carbon dioxide concentrations, (ii) multiple enzymes and microbial lineages with the ability to degrade polyphenols were present and active across thaw and depth gradients, and (iii) expressed polyphenol metabolism predicted bog porewater CO<sub>2</sub> concentrations implicating microbial greenhouse gas production outside traditionally considered substrates and redox conditions. Together these findings illustrate that in contrast to commonly-held assumptions, polyphenols are not an insurance for carbon sequestration in these climate-critical soils.

### **4.2 Introduction**

Permafrost stores an estimated 50% of soil carbon<sup>1</sup>. But with increased arctic temperatures, permafrost is thawing thus risking alterations to the global carbon economy<sup>2</sup>. The carbon found in newly thawed soils is decomposed by the microbial

communities in these soils, yielding carbon dioxide (CO<sub>2</sub>) and methane (CH<sub>4</sub>) that, and potentially accelerates climate warming<sup>3,4</sup>. Thus, understanding the facets of microbial carbon processing along these thaw-created habitats is critical to predicting the fate of permafrost carbon stores.

For decades, the enzyme latch<sup>5</sup> and derived theories<sup>6,7</sup> have aimed to explain controls on carbon mineralization in northern peatlands, targeting polyphenols as the primary inhibitors of microbial decomposition. Polyphenols are plant derived compounds, spanning more than 10,000 chemical formulas and several structural families<sup>8</sup>. In these theories, the sole microbial polyphenol degrading enzyme is polyphenol oxidase (PPO), an oxygen requiring enzyme. Under the anoxic conditions that occurs with waterlogging in these habitats, PPO is rendered inactive and polyphenols are presumed to accumulate, inhibiting soil carbon mineralization by blocking microbial metabolism. However, these theories have never been tested in the field using highly-resolved microbiome methods, instead relying on less specific chemical and enzyme assays. In light of our recent laboratory demonstration that a soil microbial community used non-PPO enzymes to degrade a model polyphenol under anoxic conditions<sup>9</sup>, we sought to comprehensively curate microbial polyphenol transformations active in thawing permafrost and revisit the fundamental assumptions of the enzyme latch in a field setting.

To test the assertions of enzyme latch, we selected a model arctic permafrost peatland, Stordalen Mire<sup>3,10</sup>. Permafrost thaw at Stordalen Mire has created differing habitats, from dry, intact permafrost (palsa) through a partially thawed bog, to a fully thawed and saturated fen. To track polyphenol metabolism across these environmental

gradients, we obtained genome-resolved metatranscriptomes, metabolomes (FT-ICRMS, LC-MS, NMR), and other geochemical data from three depths of *palsa*, bog, and fen cores. Leveraging this dataset, we provide a new view of polyphenols in microbially-catalyzed carbon cycling in these climate critical habitats.

### **4.3 Enzyme latch disproved with multiple methods**

To evaluate enzyme latch in Stordalen Mire we used more traditional methods that have long supported the theory, in addition to high-resolution microbial community focused methods. If the enzyme latch was controlling soil carbon cycling the following suppositions would be expected: (i) a positive relationship between PPO and oxygen availability, (ii) a negative relationship between PPO and polyphenol concentrations, (iii) negative relationships between polyphenols and microbial carbon processing measures, (iv) and negative relationship between polyphenols and CO<sub>2</sub> (**Figure 4.1A**). For the first supposition, our metatranscriptome data supported an oxygen-dependence of PPO, but this relationship was not observed using the traditional PPO enzyme assay (**Figure 4.1B-C**). This could be due to the convention of conducting the PPO assay under normal laboratory conditions instead of the environmentally relevant, potentially anoxic, conditions<sup>11</sup>. For the second supposition, we did not observe a significant negative relationship between PPO and polyphenol concentrations using traditional or multi-omic methods (**Figure 4.1 B-C**). For the third supposition, we did not observe any relationships between polyphenols and microbial or enzyme activity across data types, though we did see a significant negative correlation between polyphenol content and different forms of organic matter across habitats (**Figure 4.1 B-C**). Lastly, and most notably, in contrast to the primary tenet of enzyme latch, we showed significant and

positive relationships between polyphenol concentration measured several ways, and porewater carbon dioxide (CO<sub>2</sub>) concentrations (**Figure 4.1B-C**).

Collectively, our investigation of the enzyme latch hypothesis, that used both traditional and modern high-resolution methods, failed to support key steps in the biogeochemical cascade. The lack of relationship between PPO and polyphenol concentration suggests the oxygen-requiring PPO is not the sole enzyme controlling the fate of these compounds in this thaw gradient. This finding is not entirely surprising given that it is unlikely that a single enzyme modulates the behavior of this diverse pool of substrates. Paired to the positive relationship observed between polyphenol concentration and porewater CO<sub>2</sub> concentrations, it is probable that currently unaccounted microbial metabolisms contribute to polyphenol degradation *in situ*. Thus, we sought to investigate whether other avenues of microbial polyphenol metabolism were active in Stordalen Mire.

#### **4.4 Stordalen Mire polyphenols are diverse and habitat specific**

To more comprehensively profile microbe-polyphenol interactions, we characterized the polyphenol landscape in Stordalen Mire. According to the phenol-explorer classification system, polyphenols span five families: flavonoids, lignans, stilbenes, phenolic acids, and other polyphenols. Despite this chemical diversity, polyphenols have historically been regarded as a uniform group in many investigations. This is likely in part due to analytical limitations that make it difficult to examine individual compounds. Many studies have relied on methods like the Folin-Ciocalteu assay to assess “total phenolics” as a surrogate for compound-specific determinations<sup>12</sup>. At Stordalen Mire, the Folin-Ciocalteu assay reveals the bog has

significantly more soluble polyphenols than the fen ( $p=0.013$ , **Figure 4.2A**). However, this masks any information on the diversity and identity of the polyphenols in the samples. Therefore, we used Fourier Transform mass spectrometry (FT-ICRMS) to characterize the high-molecular weight, likely polymeric polyphenols, and Liquid Chromatography mass spectrometry approaches (LC-MS) to detect smaller polyphenols and phenolic compounds.

A FT-ICRMS data analysis convention is assigning identified masses a formula, and placing this formula in a biochemical class (e.g. tannin-like, lignin-like) using the ratios of O:C and H:C<sup>13</sup>. Using compounds with known polyphenol structures we refined the boundaries on these biochemical classes, defining lignin-like, hydrolysable tannin-like, and condensed tannin-like boundaries (see methods). Metabolites within these classes were all enriched in the palsa and bog (**Figure 4.2B**). We further identified 24 polyphenol metabolites in our LC-MS data exhibiting habitat-specific abundance patterns (**Figure 4.2B**). We detected different phenolic acids across the site, as well some interesting “other” polyphenols. We exclusively detected two stilbenes in the bog. Several monomeric and dimeric flavonoids were detected across the palsa and bog, while in the fen we detected three different flavonoids that were prenylated<sup>14</sup>. Collectively, this high-resolution view of Stordalen Mire metabolites was an important step to identify the prevalent classes of polyphenols microorganisms may encounter.

#### ***4.5 Diverse lineages express polyphenol metabolizing gene in Stordalen Mire microbiomes***

Polyphenol metabolism is challenging to infer from microbial genomes because the genes encoding their metabolism are either missing or unlabeled in misannotated in

databases. Therefore, to survey the potential for polyphenol metabolism across the microbial community at Stordalen Mire, we developed Curated Annotations for Microbial Polyphenol Enzymes and Reactions (CAMPER), a tool for annotating and summarizing genomic potential for polyphenol transformations. CAMPER can annotate homologs to 41 biochemically characterized polyphenol-active enzymes (**Chapter 3**) and nearly 300 previously recognized polyphenolic-related annotations. CAMPER defines 101 modules that describe polyphenol transformations (**Figure 4.3A**). These modules are first organized using the established ontology for polyphenols<sup>8</sup>, including the structural family and subfamily of polyphenols they act upon, then classified by the known oxygen requirements for the enzymes. CAMPER also broadly categorizes polyphenol metabolism, considering reactions where polyphenols directly (ex. complete degradation) or indirectly (ex. conversions, respiration, non-specific reactions) support microbial metabolism.

To investigate the potential for polyphenol metabolism in Stordalen Mire, we used CAMPER to annotate a database of 1,864 medium- and high-quality metagenome-assembled genomes (MAGs) recovered from the site. Of the 101 modules in CAMPER, we observed 84 in our MAGs. The potential for polyphenol metabolism spanned the standardized GTDB<sup>15</sup> bacterial and archaeal trees (**Figure 4.3B**), with 73% (n=1,372) of the MAGs encoding at least 5 CAMPER modules. The average and median number of modules encoded across the MAGs was 9 and 7, respectively (**Figure 4.3C**). In fact, of the 43 phyla-level lineages in this Stordalen MAG dataset, the Patescibacteria (n=58 MAGs) were the only lineage largely devoid of CAMPER modules.

MAGs from the Acidobacteriota, Proteobacteria, Actinobacteriota, Eremiobacteriota, and Desulfobacterota\_B were particularly enriched in polyphenolic metabolism, encoding over 23 CAMPER modules (**Figure 4.3D**). Among these are MAGs from well characterized polyphenol and aromatic active genera like *Novosphingobium*<sup>16</sup> and *Herbaspirillum*<sup>17</sup>. Interestingly, of the top 100 CAMPER encoding MAGs, 66 belong to new or undescribed lineages, illustrating the power of CAMPER to broaden knowledge of microbial polyphenol metabolism.

#### ***4.6 Microbial polyphenol metabolism is active across Stordalen Mire***

To complement our findings that polyphenol metabolism was widely encoded across the Stordalen Mire microbiome, we investigated the prevalence of CAMPER modules in field metatranscriptomes. Linking expression to the detected polyphenol metabolite pool, CAMPER module expression was significantly related to polyphenol metabolite abundance across habitats and depth gradients (Procrustes analysis,  $p=0.026$ , 999 permutations). Across the habitats and depths we detected microbial expression of 39 unique polyphenol transforming genes, which were assigned to 30 CAMPER modules. Module expression levels exhibited site and depth specific trends (**Figure 4.4A**). Notably, reflecting soil redox, the deep waterlogged samples (ex. 20-24 cm bog and fen) were enriched in exclusively anoxic CAMPER modules. This refutes all-or-nothing PPO assumptions of the enzyme latch, where polyphenolic degradation is only mediated by a single, oxygen requiring enzymes. Instead our data depicts microbial communities that are equipped and expressed genes for polyphenol metabolism regardless of redox. The vast number of polyphenolic genes, acting on a

wide array of polyphenol substrates, exemplifies the many ways polyphenols may contribute to microbial metabolism.

To evaluate this we highlight different polyphenol active metabolisms across the mire. A module highly enriched in the palsa metatranscriptome is chlorogenic acid cleavage (**Figure 4.4B**), comprised of a single gene encoding the enzyme chlorogenic acid esterase (chlE). Across palsa depths, we detected chlE expression from MAGs spanning 24 genera. The canonical reaction for this enzyme is cleavage of the ester bond in chlorogenic acid to release the phenolic compound caffeic acid and the nonphenolic quinic acid that can be further degraded via central aromatic metabolism. Chlorogenic acid is a reported metabolite of *Sphagnum*<sup>18</sup>, which is a dominant vegetation cover at Stordalen<sup>19</sup>. An alternative reported substrate is propyl-gallate, which is cleaved to propanol and gallic acid<sup>20</sup>. Integrating metabolite and metatranscriptome data, both quinic acid and propyl-gallate are detected metabolites (**Figure 4.2**). This module is an example of how polyphenols can be cleaved to release other metabolites for degradation.

Coniferyl alcohol degradation was enriched in the bog metatranscriptomes (**Figure 4.4C**), expressed from MAGs representing 96 genera. In this module, the enzymes coniferyl alcohol dehydrogenase (calA) and coniferyl aldehyde dehydrogenase (calB) convert coniferyl alcohol to ferulic acid, a compound we detected in bog porewaters. However, we also detected the metabolite 4-methoxycinnamaldehyde and given reports that these enzymes could also act on cinnamic acids<sup>21</sup>, the substrates cannot be deduced from expression data alone. It is possible that the broad substrate range for these enzymes using multiple confirmed substrates in these porewaters could

explain the expression across diverse lineages. This module highlights that polyphenol metabolism often involves converting one polyphenol to another.

While pyrogallol was detected across sites (**Figure 4.2**), expression of pyrogallol degradation was enriched in the fen (**Figure 4.4D**) expressed by 13 different genera. The enzyme pyrogallol-phloroglucinol transhydroxylase (pgthAB) converts pyrogallol to phloroglucinol, which is then degraded to acetate via the phloglucinol reductase and a series of unknown enzymes. The production of acetate from this metabolism is notable, as this fen is enriched in acetoclastic methanogenesis<sup>22</sup> relative to the other sites. Therefore, polyphenol metabolism may also generate substrates that drive other biogeochemically important metabolisms.

Most interestingly, the caffeic acid respiration module is enriched in the bog and the fen (**Figure 4.4E**). In this, caffeic acid serves as an electron acceptor, canonically paired to hydrogen as an electron donor. First, caffeic acid is activated to caffeyl-coA and subsequently reduced to hydrocaffeyl-coA, which transfers the coA to another caffeic acid in an energy saving loop<sup>23</sup>. The reduction step generates NAD<sup>+</sup> and reduced ferredoxin, which fuel the Rnf complex and a hydrogenase<sup>23</sup>. The car operon was characterized in *Acetobacterium woodii*, and we expand it to 167 active genera. This metabolism defies the view that polyphenols can only serve metabolism as a carbon source or electron donor, instead supporting anaerobic respiration.

#### **4.7 Active polyphenol metabolism by undescribed lineages**

We leveraged our genome-resolved metatranscriptome data to identify key polyphenol-active lineages in Stordalen Mire. Of the MAGs encoding CAMPER modules, 75% (n=1,007) had polyphenol modules that recruited metatranscriptome

reads, illustrating that polyphenol metabolism is cosmopolitan across the mire. We aggregated our MAGs at the genus level to yield 266 genera and found nearly half of these exhibited habitat-specific polyphenolic activity (**Figure 4.5A**). Reinforcing the status of the bog as a transition site in the thaw gradient, 95 genera were either active in the bog and palsa, or the bog and fen.

We used a combination of average CAMPER expression level and average number of modules expressed per site to determine top polyphenol metabolizing genera (**Figure 4.5B**). A genus designated Palsa-295, a member of an undescribed order of the Acidobacteriota, expressed on average 12 modules in the palsa. In the bog, a putative novel genera in the Proteobacterial Steroidobacteraceae, Bog-1198, expressed on average 15 modules. A novel genera of the Chloroflexota, Fen-1064, was the top lineage in the fen. Therefore, we show polyphenol metabolism as a function for these unexplored lineages. Of note, an Acidobacterial genus *Terracidiphilus* was active across all habitats, expressing on average 9 modules in the bog and fen. There is just one characterized species of *Terracidiphilus*<sup>24</sup>, and here we expand the potential metabolisms for this genus to include polyphenol metabolism. Collectively, we illustrate the combined power of CAMPER, a rich database of microbial MAGs, and field metatranscriptomes for uncovering novel lineages and functions.

#### **4.8 Polyphenol metabolism predicts porewater carbon dioxide**

We wanted to understand how this refined view of polyphenol metabolism could elucidate carbon mineralization. Using a sparse partial least squares regression (sPLSR), we predicted bog porewater CO<sub>2</sub> concentrations from CAMPER module expression (**Figure 4.6A**,  $r^2=0.55$ ,  $p=0.028$ ). This model predicted bog porewater CO<sub>2</sub>

concentrations nearly as well as a model constructed from geochemistry variables (ex. depth, temperature; **Figure 4.6B**,  $r^2=0.60$ ,  $p=0.030$ ). Together, polyphenol module expression and geochemistry explained a substantial amount of bog porewater CO<sub>2</sub> concentration variation (**Figure 4.6C**,  $r^2=0.81$ ,  $p=0.010$ ). Module expression did not explain bog porewater CH<sub>4</sub> concentrations possibly because the end products of this metabolism may not yield methanogenic substrates or may be toxic to methanogens. Module expression was also not explanatory for fen porewater gas concentrations, which could be because the polyphenol concentration was significantly higher in the bog samples (**Figure 4.2**) and thus more centrally linked to microbial carbon metabolism. Taken together there are many factors beyond a single compound type that likely regulate microbial decomposition in these habitats, but our findings do demonstrate polyphenol metabolism is broader than a single enzyme and that polyphenols do not appear to restrict carbon mineralization in these soils.

To better resolve the key components of the combined polyphenol gene expression and geochemistry model, we used a Variance in Projection (VIP) score and correlations to bog porewater CO<sub>2</sub> concentrations to define significance (**Figure 4.6D**). When considering both VIP and correlation significance, depth was the most significant variable in the model (VIP=3.81) for predicting CO<sub>2</sub> concentrations, while expression of several oxic CAMPER modules was also predictive and correlative. We note these genes were not PPO, but included modules for stilbene and lignan degradation. Interestingly, these modules were all negatively correlated to bog CO<sub>2</sub> concentrations, which may reflect a depth-dependence due to their oxic nature. This analysis also showed that there were predictive CAMPER modules that were both positively and

negatively related (not significant correlations) to bog porewater CO<sub>2</sub>, revealing that polyphenol metabolism is a dynamic and complicated part of bog carbon cycles. Together, these results illustrate that polyphenol metabolism is an important parameter for predicting future CO<sub>2</sub> emissions, albeit in a site-specific manner.

The centrality of polyphenol metabolism to bog carbon cycles shown here is notable, as the enzyme latch was originally developed in bogs. Beyond that, other studies have targeted bog polyphenols as being inhibitory and limiting to carbon mineralization<sup>25</sup>. To reconcile our results with those studies, we offer two points: (1) it can be true that polyphenols are inhibitory to some microorganisms while also being used by others, and (2) that complex webs of microbial metabolisms and environmental factors contribute to emitted CO<sub>2</sub> and CH<sub>4</sub>. Towards (1), we showed the bog microbial communities found along the depth gradient were actively using polyphenols. As such, more targeted studies are needed to understand which polyphenols are inhibitory, under which conditions, and what microorganisms they inhibit. Towards (2), we offer these findings and CAMPER as a public tool for further understanding how microbial polyphenol metabolism is wired into field carbon cycles.

#### **4.9 Conclusion**

Polyphenols have long been cited as controllers of peatland carbon cycles through the enzyme latch mechanism. As the studies proposing and supporting these roles have historically measured polyphenol impacts on net outputs of mineralization, we sought to focus on the interplay between polyphenols and the microbiome. Here, we refuted an enzyme latch mechanism in Stordalen Mire using both traditional and genome-enabled methods. We show microbial polyphenol metabolism is more than

polyphenol oxidase and that polyphenols can serve a variety of metabolic uses for microbial communities. In Stordalen Mire, we illustrated that polyphenol metabolism is broadly encoded in the microbial community and expressed across a permafrost thaw gradient regardless of redox conditions. Finally, we show that expressed polyphenol metabolism could help explain porewater CO<sub>2</sub> concentrations in bogs. Future assessments of peatland and permafrost microbial carbon cycles should consider the myriad ways that polyphenols support microbial metabolism to better understand and to predict ecosystem biogeochemistry.

#### **4.10 Methods**

##### *4.10.1 Field Site and Sampling*

Stordalen Mire is a peatland in Northern Sweden (68 22' N, 19 03' E). The site has three main habitats along a permafrost thaw gradient, with distinct vegetation and thaw status: (1) palsa, overlaying permafrost, well-drained, and dominated by woody plants and shrubs, (2) bog, intermediate-thaw with variable water table depth, dominated by *Sphagnum* mosses, and (3) fen, fully thawed and waterlogged, dominated by *Eriophorum* sedges.

In July 2016, triplicate cores were taken from palsa, bog, and fen sites (all within a 120m radius) in Stordalen Mire using an 11-cm diameter push-corer. Cores were sectioned in the field at three depths: surface (1-4cm), middle (10-14cm), and deep (20-24cm). Core sections were stored at -20°C until analysis.

##### *4.10.2 Field Geochemical Analyses*

Before coring, porewater from each core site was collected using a perforated stainless-steel piezometer inserted into the peat to the desired depth, and extracting

with an airtight syringe. Due to the dry nature of the palsa, no pore water was collected, and often was unable to be collected in bog and fen surfaces. Porewater was used to determine pH, and used to measure porewater CO<sub>2</sub> and CH<sub>4</sub> using a Flame-Ionization-Detector Gas Chromatography equipped with a methanizer<sup>25</sup>.

#### *4.10.3 Metagenome Assembled Genome Taxonomy, Phylogeny, and Annotation*

A database of 13,290 medium- and high-quality metagenome-assembled genomes (MAGs) has constructed from 882 Stordalen Mire field and incubation metagenomes, from 2011-2017. These 13,290 MAGs were dereplicated at 97% identity using dRep<sup>26</sup> into 1,864 representative MAGs. MAGs were annotated using DRAM<sup>27</sup> (v1.3.2).

MAG taxonomy was assessed using the GTDB-tk<sup>15</sup> (v2, r207). Phylogenetic trees of the MAGs were inferred using the GTDB de\_novo\_wf workflow for both bacterial and archaeal MAGs, using p\_\_Patescibacteria and p\_\_Micrarchaeota as outgroups, respectively. Phylogenetic trees were visualized using the ggtreeEXTRA<sup>28</sup> R packages.

#### *4.10.4 RNA Extraction*

DNA and RNA were co-extracted using the Mobio PowerMax Soil DNA/RNA isolation kit (cat# 12966-10) with slight modifications. Briefly, sample vials were thawed on ice. After thawing, 5-10 g of peat materials (preserved in Lifeguard soil preservation solution, Qiagen) was added into the bead tubes in the kit, and the nucleic acids were extracted following the manufacturer's protocol without the addition of beta-mercaptoethanol at the beginning. Reagents were proportionally increased to maintain the concentration and strength of solutions. An additional ethanol wash of nucleic acids-bound column was performed to further wash out impurities. The resulting nucleic acids were eluted with 5ml RNase-free DI water and further

concentrated using ethanol precipitation overnight and re-eluted in 100 ul of TE. The resulting nucleic acids are then further processed for DNA and RNA separation and purification. Briefly, the extracted nucleic acids were aliquoted into two 2ml tubes at a ratio of 1:2. RNase treatment and DNase treatment (Roche) were performed following the manufacturers' instructions on the recovery of DNA and RNA. After the treatments, DNA and RNA were further purified by phenol:Chloroform purification to remove the enzymes and impurities. DNA and RNA were then ethanol precipitated and the pellets were eluted in TE buffer. Final purified DNA and RNA were quantified with Qubit 3.0. Quality of extracted RNA was also examined with TapeStation analysis at Genome shared resources (GSR) at the Ohio State University. Extracts were stored at -80C till downstream sequencing analysis.

#### *4.10.5 Metatranscriptome Analysis*

Metatranscriptome libraries were prepared for the 27 samples. Using 10ng RNA as input, rRNA was depleted using the QIAseq FastSelect -5S/16S/23S (Qiagen) kit following the kit protocol with the following modifications: addition of probes for plant and yeast also, and using one-third the probe volumes. Then, the TruSeq Stranded Library Preparation kit (Illumina) was used to prepare the sequencing library. Libraries were sequenced on an Illumina NovaSeq 6000 system at the Genomics Core at the University of Colorado Anschutz Medical Campus.

Raw metatranscriptome reads were quality trimmed and adapters removed using bbduk<sup>29</sup> with the following flags: k=23 mink=11 hdist=1 qtrim=rl trimq=20 minlength=75. Reads were filtered with rqcfilter2[ref] using the following flags: jni=t ma=t trimfragadapter=t qtrim=r trimq=0 maxns=1 maq=10 minlen=51 mlf=0.33 phix=t removeribo=t removehuman=t removedog=t removecat=t removemouse=t khist=t removemicrobes=t mtst=t sketch kapa=t clumpify=t tmpdir=null barcodefilter=f

trimpolyg=5. Filtered and trimmed reads were mapped against the database of 97% dereplicated MAGs using bowtie2<sup>30</sup> using the following settings: -D 10 -R 2 -N 1 -L 22 -i S,0,2.50. The output SAM file was converted to BAM using samtools<sup>31</sup>, and filtered using the reformat.sh script in the bbttools<sup>29</sup> package using: idfilter=0.95 pairedonly=t primaryonly=t. Mapped reads were counted using htseq-count<sup>32</sup> using the following flags: -a 0 -t CDS -i ID --stranded=reverse. Read counts were filtered to remove counts <5, and were converted to geTMM<sup>33</sup> in R.

#### 4.10.6 Enzyme Assays

Activity of five hydrolytic enzymes  $\beta$ -d-glucosidase,  $\beta$ -d-xylosidase, N-acetyl- $\beta$ -d-glucosaminidase, arylsulphatase, and phosphatase was assessed with fluorometric enzyme assays following methods adapted from Saiya-Cork et al. 2002<sup>34</sup> and DeForest, 2009<sup>35</sup>. Briefly, a soil slurry was prepared for each sample by blending 1 g of soil with 125 mL of sodium acetate buffer (50 mM; pH 6.2). The soil slurry was then transferred to a 96 well flat-bottom black microplate which included a buffer only control column, as well as controls containing only soil and standard. Following, 4-methylumbelliferyl (MUB) standard solution and fluorescently linked enzyme substrates were added to the respective wells and the plates were incubated at 25°C for 45 min for  $\beta$ -d-glucosidase, N-acetyl- $\beta$ -d-glucosaminidase, arylsulphatase, and phosphatase, and 30 min for  $\beta$ -d-xylosidase. Incubation times and substrate concentrations were chosen based on a V-max test performed to capture peak enzyme activity. Fluorescence was read in a BioTek Synergy HT microplate reader at a wavelength of 460 nm emission and 360 nm excitation. Final enzyme activity was reported as  $\mu\text{mol activity g}^{-1} \text{ dry soil h}^{-1}$ .

The oxidative enzyme activity of phenol oxidase was assessed with a colorimetric enzyme assay<sup>35</sup>. A soil slurry was prepared for each sample by blending 1 g of soil with 125 mL of sodium acetate buffer (50 mM; pH 6.2), and slurries were transferred to a 96 deep-well plate. A blank column containing only buffer was included in the plate, as well as controls containing only buffer and substrate. L-3,4,-dihydroxy phenylalanine (L-DOPA) was chosen as the substrate for measuring phenol oxidase activity. After adding 25 mM L-DOPA substrate, plates were incubated at 25°C for 24 hours. Following incubation, the supernatant was transferred to a 96 well flat-bottom clear microplate, and absorbance was read in a BioTek Synergy HT microplate reader at 460 nm. Final activity was reported as  $\mu\text{mol activity g}^{-1} \text{ soil h}^{-1}$ .

#### *4.10.7 Folin-Ciocalteu Assay*

To determine total phenolics, we used the Folin-Ciocalteu assay on the water extracts of the soil samples. Extracts were prepared from all the samples, centrifuged to remove the soil, and the extracts were stored at -80 until analysis. On the day of analysis, each extract was thawed and centrifuged briefly (10,000 x g) to remove insoluble material, and 25  $\mu\text{L}$  aliquots were transferred to 96 well plates. The methyl gallate standard (0.2 mg/mL) was dispensed into the wells to obtain a series of samples containing 0-4.25  $\mu\text{g}$  methyl gallate. On each plate, both samples and standards were run in triplicate. Each extract was analyzed on two different days.. After adding 50  $\mu\text{L}$  of Folin reagent (Sigma), 20  $\mu\text{L}$  of 20% (m/v)  $\text{Na}_2\text{CO}_3$ , and enough water to bring the volume of each well to 95  $\mu\text{L}$ , the plate was briefly mixed and then incubated in the dark for 20. The samples were then read at 750 nm (Biotek Synergy LX, a standard curve

was generated, and the total phenolic content of each sample established by interpolation.

#### *4.10.8 Organic Matter metabolite Extraction*

Water soluble metabolites were extracted from peat by adding 7 mL of autoclaved milliQ water to 1g of peat in a sterile 15 mL Eppendorf tube. Tubes were vortexed twice for 30 seconds, and then the peat-water mixture was sonicated for 2 hours at 22°C. Samples were then centrifuged to separate the supernatant, which served as the water extract.

#### *4.10.9 LC-MS*

Water extracted metabolites were thawed at room temperature and centrifuged again to remove any potential particles that formed after thawing. Next, each sample was split into two 2ml glass tube vials (1 ml each), one for hydrophilic interaction liquid chromatography (HILIC) and the other for reverse-phase (RP) liquid chromatography. Samples in both vials were then dried down completely on a Vacufuge plus (Eppendorf, USA). Samples were resuspended in a solution of 50% Acetonitrile and 50% water for HILIC and a solution of 80% water and 20% HPLC grade methanol for RP.

A Thermo Scientific Vanquish Duo ultra-high performance liquid chromatography system (UHPLC) was used for the liquid chromatography step. Extracts were separated using a Waters ACQUITY HSS T3 C18 column for RP separation and a Waters ACQUITY BEH amide column for HILIC separation.

Samples were injected in a 1  $\mu$ L volume on column and eluted as follows: for RP the gradient went from 99% mobile phase A (0.1% formic acid in H<sub>2</sub>O) to 95% mobile phase B (0.1% formic acid in methanol) over 16 minutes. For HILIC the gradient went from 99% mobile phase A (0.1% formic acid, 10 mM ammonium acetate, 90% acetonitrile, 10% H<sub>2</sub>O) to 95%

mobile phase B (0.1% formic acid, 10 mM ammonium acetate, 50% acetonitrile, 50% H<sub>2</sub>O). Both columns were run at 45 °C with a flowrate of 300 µL/min.

A Thermo Scientific Orbitrap Exploris 480 was used for spectral data collection with a spray voltage of 3500 V for positive mode (for RP) and 2500 V for negative mode (for HILIC) using the H-ESI source. The ion transfer tube and vaporizer temperature were both 350 °C. Compounds were fragmented using data-dependent MS/MS with HCD collision energies of 20, 40, and 80.

The Compound Discoverer 3.2 software (Thermo Fisher Scientific) was used to analyze the data using the untargeted metabolomics workflow. Briefly, the spectra were first aligned followed by a peak picking step. Putative elemental compositions of unknown compounds were predicted using the exact mass, isotopic pattern, fine isotopic pattern, and MS/MS data using the built in HighChem Fragmentation Library of reference fragmentation mechanisms. Metabolite annotation was performed using spectral libraries and compound databases. First, fragmentation scans (MS<sub>2</sub>) searches in mzCloud were performed, which is a curated database of MS<sub>n</sub> spectra containing more than 9 million spectra and 20000 compounds.

Second, predicted compositions were obtained based on mass error, matched isotopes, missing number of matched fragments, spectral similarity score (calculated by matching theoretical and measured isotope pattern), matched intensity percentage of the theoretical pattern, the relevant portion of MS, and the MS/MS scan. The mass tolerance used for estimating predicted composition was 5 ppm. Finally, annotation was complemented by searching MS<sub>1</sub> scans on different online databases with ChemSpider (using either the exact mass or the predicted formula). Based on the annotation results, metabolites were divided into three categories: 1) full match on the three methods used (mzCloud, predicted composition, and ChemSpider), 2) full match by two methods (Predicted composition and ChemSpider) and 3) annotated only by one method (ChemSpider).

#### 4.10.10 FT-ICRMS Analysis

Water extracts were first purified using solid phase extraction (SPE) to remove contaminants (i.e., salts) according to Dittmar et al., 2008<sup>36</sup>. Briefly, water extracts were acidified to pH 2 using 1M HCL. Then, extracts were filtered through a 3 mL Bond Elut PPE cartridge (Aligent) that was previously activated using methanol. Cartridges were washed 3 mL of a 0.01 M HCl solution for five times, then dried using filtered air. Finally, extracts were eluted using 1.5 mL of methanol and stored at -80 °C until used.

Purified extracts were analyzed by direct injection using a 12 Tesla Bruker FTICR mass spectrometer located at the Pacific Northwest National Laboratory (PNNL). Positive and negative charged molecular ions were generated using a Bruker electrospray ionization (ESI) source. The instrument stability was optimized using A Suwannee River Fulvic Acid standard (SRFA), obtained from the International Humic Substance Society (IHCC). Potential carry-over between samples was monitored by injecting HPLC grade methanol. The instrument was flushed between samples using a combination of milliQ water and methanol. In order to account for variations in carbon concentrations in different samples the ion accumulation (IAT) was varied between 0.03 and 0.05 s. A total of 144 individual scans per sample were collected, averaged, and calibrated using an organic matter homologous series separated by 14 Da (CH<sub>2</sub>). Mass accuracy was < 1 ppm for single charged ions measured across a m/z range of 100–1,200 m/z, the mass resolution was ~240K at 341 m/z and the transient was 0.8 s. Raw spectra collected per sample was transformed into a list of m/z values using the FT-MS peak picker module within the BrukerDaltonik version 4.2 software using a signal to

noise ratio of 7 and absolute intensity threshold of 100 (default). Formularity<sup>37</sup> software was used to assign putative chemical formulae following Tfaily et al., 2018<sup>38</sup>.

#### 4.10.11 Refinement of Tannin and Lignin Van Krevlen classes

Van Krevlen analyses assign molecular formula to classes using H:C and O:C ratios. Two classical regions are the lignin-like and tannin-like region, however these mask a huge amount of complexity (ex. Condensed tannins vs. hydrolysable tannins). Using an approach similar to that in Laszakovits et al. 2022<sup>39</sup>, we sought to refine these classifications using known formula from over 60 characterized natural substrates. We determined the following H:C and O:C ratios for hydrolysable tannins (HT), condensed tannins (CT), and lignin:

Lignin:  $0.3 < O:C < 0.48$ ,  $1.08 < H:C < 1.28$

Hydrolyzable Tannin:  $0.6 < O:C < 0.7$ ,  $0.58 < H:C < 0.89$

Condensed Tannin:  $0.4 < O:C < 0.5$ ,  $0.74 < H:C < 0.88$

#### 4.10.12 Nuclear Magnetic Resonance (NMR) Analysis

To follow bio-available metabolites, we used NMR on the water extracted supernatant samples. Supernatant samples (180  $\mu$ L) were combined with 2,2-dimethyl-2-silapentane-5-sulfonate- $d_6$  (DSS- $d_6$ ) in  $D_2O$  (20  $\mu$ L, 5 mM) and thoroughly mixed prior to transfer to 3 mm NMR tubes. NMR spectra were acquired on a Bruker Neo spectrometer operating at 18.8T ( $^1H$   $\nu_0$  of 800.30 MHz) equipped with a 5mm Bruker TCI/CP HCN (inverse) cryoprobe with Z-gradient at a regulated temperature of 298.0 K. The  $90^\circ$   $^1H$  pulse was calibrated prior to the measurement of each sample. The one-dimensional  $^1H$  spectra were acquired using a nuclear Overhauser effect spectroscopy (noesypr1d) pulse sequence with a spectral width of 20.1 ppm and 2048 transients. The

NOESY mixing time was 100ms and the acquisition time was 4s followed by a relaxation delay of 1.5 s during which presaturation of the water signal was applied. The 1D  $^1\text{H}$  spectra were manually processed, assigned metabolite identifications and quantified using Chenomx NMR Suite 9.0. Time domain free induction decays (65536 total points) were zero filled to 131072 total points prior to Fourier transform, followed by exponential multiplication (0.3 Hz line-broadening), and semi-automatic multipoint smooth segments baseline correction. Chemical shifts were referenced to the  $^1\text{H}$  methyl signal in DSS- $\text{d}_6$  at 0 ppm. Metabolite identification was based on matching the chemical shift, J-coupling and intensity of experimental signals to compound signals in the Chenomx, HMDB and custom in-house databases. Quantification was based on fitted metabolite signals relative to the internal standard (DSS- $\text{d}_6$ ). Signal to noise ratios (S/N) were checked using MestReNova 14.1 with the limit of quantification equal to a S/N of 10 and the limit of detection equal to a S/N of 3.

#### *4.10.13 CAMPER Construction*

CAMPER is comprised of both Hidden-Markov Model (HMM) and Basic Local Alignment Search Tool (BLAST) searches. HMMs were created in cases where sufficient characterized sequences existed (ex. >3, with some phylogenetic diversity). For each gene type, we identified decoy sequences that represented related but functionally distinct sequences to the gene of interest. To create the profiles, we BLAST'ed the characterized sequences and the decoy sequences against the UniProt90<sup>40</sup> database and pulled the top 200 hits for each gene. Using the pulled sequences, characterized sequences, and decoy sequences, we created a sequence alignment using MAFFT<sup>41</sup> with the “-auto” flag, and the alignment was trimmed using

trimal<sup>42</sup> -gappyout. This trimmed alignment was fed to IQTree<sup>43</sup> with following flags: -alrt 1000 -bb 1000 -m MFP -nt AUTO -ntmax 10. The “.tree” file was visualized in iTOL<sup>44</sup>, and rooted on the clade containing all decoy sequences. Then, we identified clades that contained all characterized sequences, and pulled the sequences in the clade. Importantly, we removed the characterized sequences from this set. Using the remaining sequences, we constructed HMMs in graftM<sup>45</sup> using the graftm create command, and searched them against the sequences in the tree using the graftm graft command. HMM bitscores were visualized on the tree. We made two score cutoffs: the “A” score is the lowest bitscore assigned to a characterized sequence. The “B” score was curated from a combination of tree placement and score distribution.

MAGs were annotated for CAMPER genes using CAMPER within DRAM v1.3.2. MAGs were said to encode a module if they possessed >50% of the genes in the module. MAGs were said to express a module if they encoded >50% of the genes in the module, and at least 1 gene recruited metatranscriptome reads.

#### *4.10.14 Statistics*

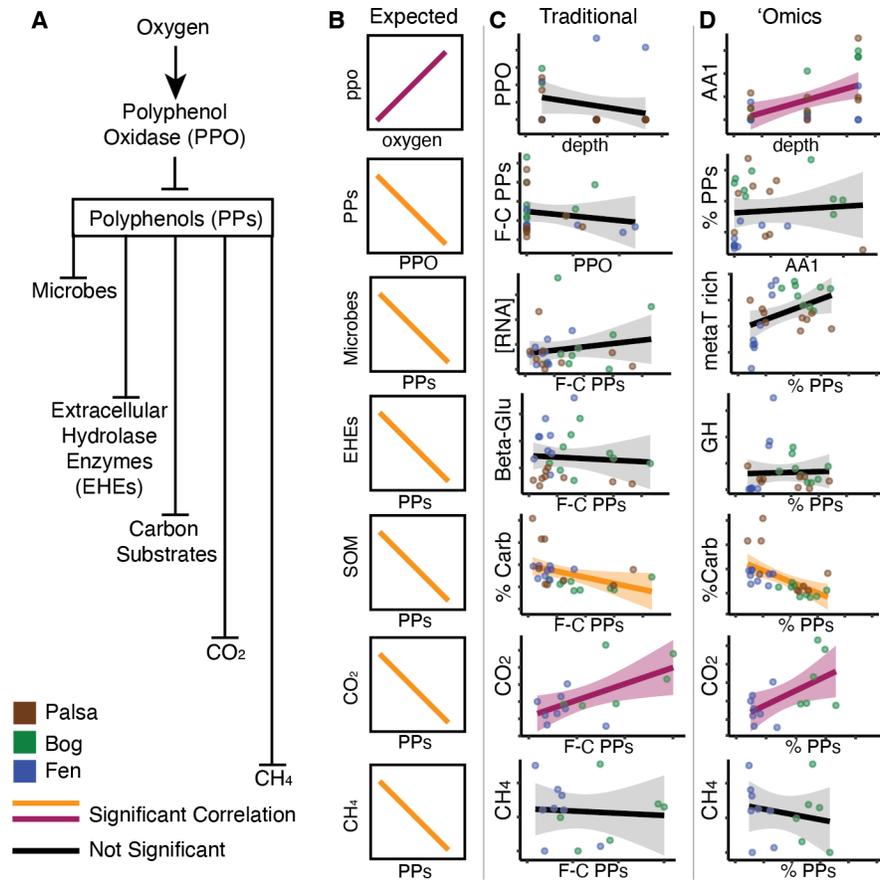
All data analysis and visualization was done in R (v4.2.1). For correlations of enzyme latch data types, we used Pearson correlations in R, and p-values were adjusted within groups of comparisons using Benjamini-Hochberg adjustment.

Procrustes analysis was done in R with vegan protest to compare Bray-Curtis Dissimilarity matrices for polyphenol metabolites and CAMPER module expression.

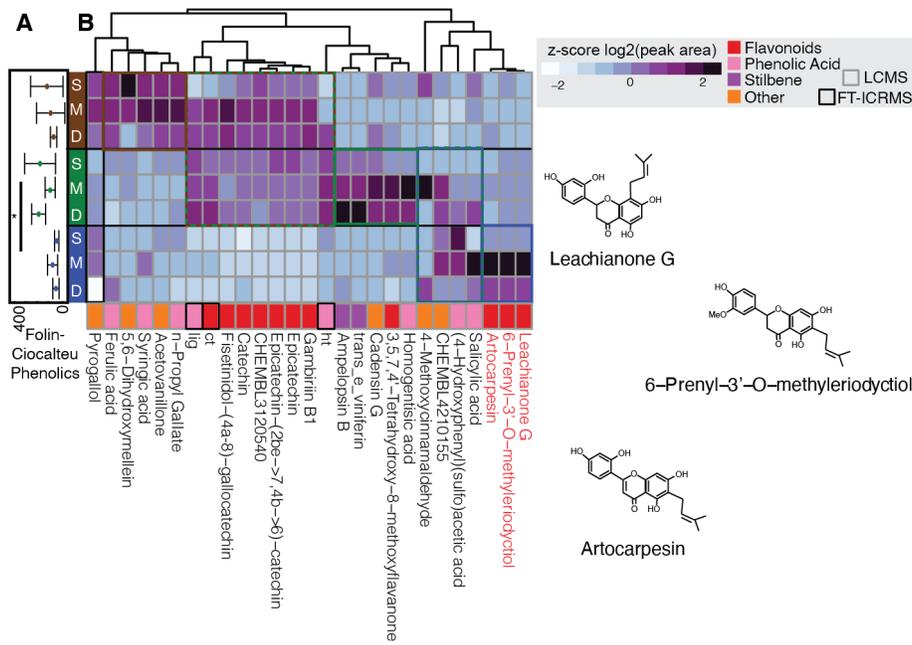
To predict porewater gas concentrations from geochemistry and CAMPER module expression, we used sparse PLS (sPLS)<sup>46</sup>, as implemented in the R package

mixOmics<sup>47</sup>. Geochemistry variables were depth, soil temperature, Folin-Ciocalteu phenolics, and pH. All VIP scores<sup>48</sup> were reported.

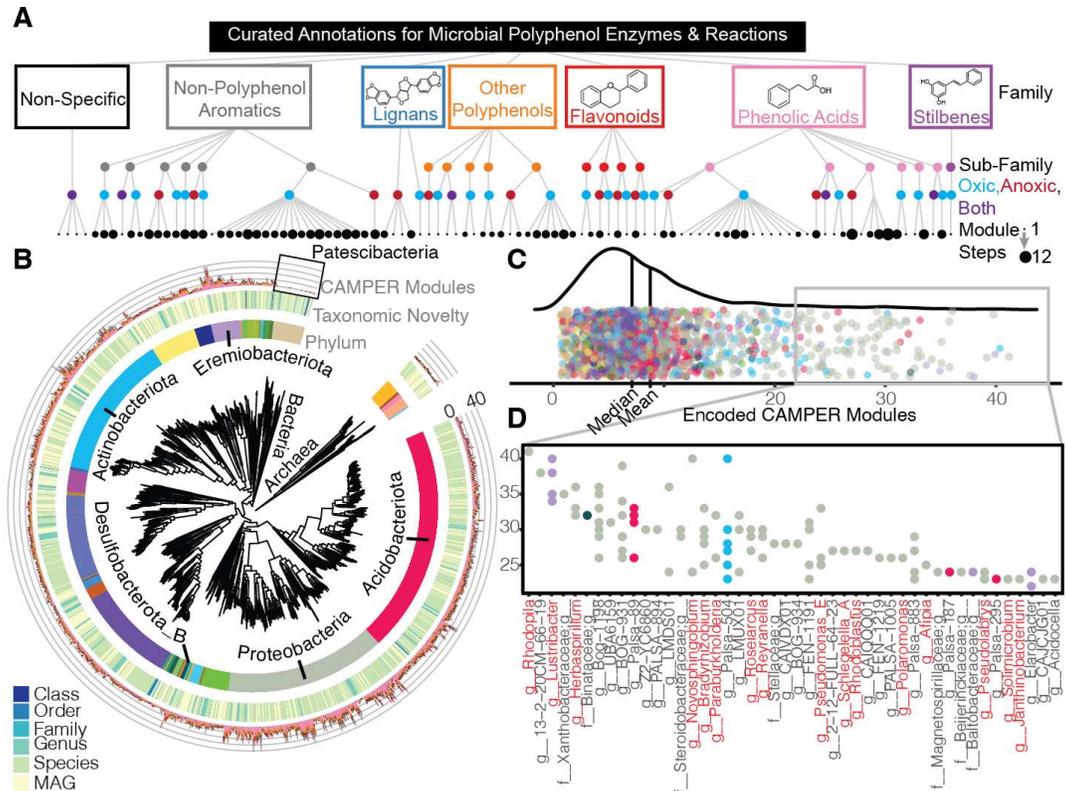
## Chapter 4 Figures



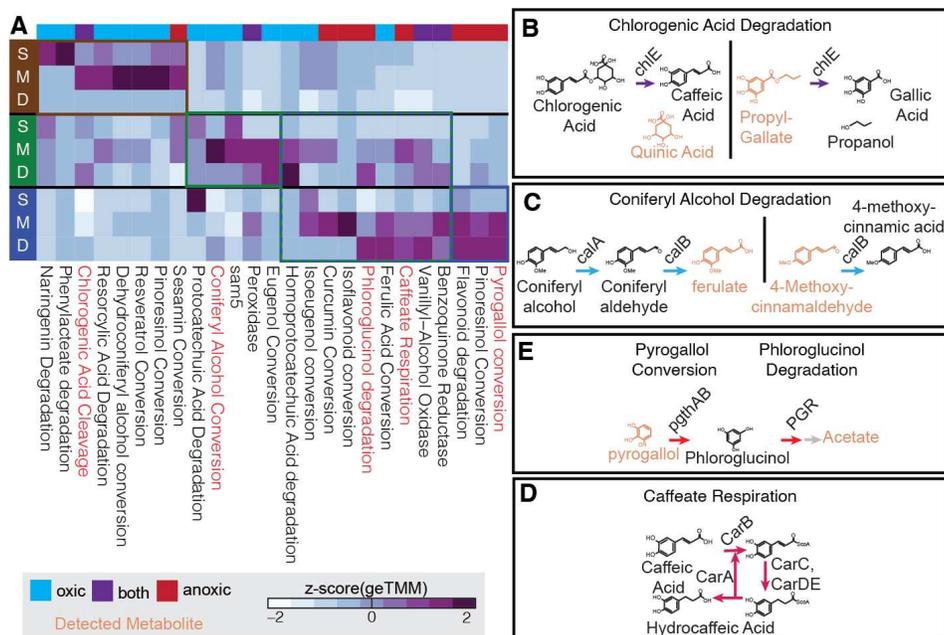
**Figure 4.1.** Multi-method investigation of the Enzyme Latch. (A) The Enzyme Latch cascade. (B) The expected relationships proposed by the Enzyme Latch. (C) Observed relationships using ‘Traditional’ assay methods. (D) Observed relationships using ‘Omics and high-resolution methods. Significant Pearson correlations (Benjamini-Hochberg adjusted  $p < 0.05$ ) are colored red and orange for positive and negative correlations, respectively. Depth is used as a proxy for oxygen, shown moving from deep to surface (left to right). Polyphenol content is given by Folin-Ciocalteu phenolics (F-C PPs, mg Methyl-Gallate equivalents/dry g soil) and polyphenol-like compounds identified in FTICRMS (%PPs). Polyphenol Oxidase activity is given by Phenol Oxidase assay (PPO, nmol activity/g/hr) and metatranscriptome expression of polyphenol oxidase (AA1, geTMM). Microbial activity is inferred through extracted RNA ([RNA], ng/ $\mu$ L) and number of taxa that recruited reads in metatranscriptome (metaT rich). Extracellular hydrolase enzymes are given by Beta-glucosidase activity (Beta-Glu, nmol activity /g/hr) and glycoside hydrolase expression in metatranscriptome (GH, geTMM). Soil organic matter (SOM) is given by carbohydrate-like compounds identified in FTICRMS (%Carb).



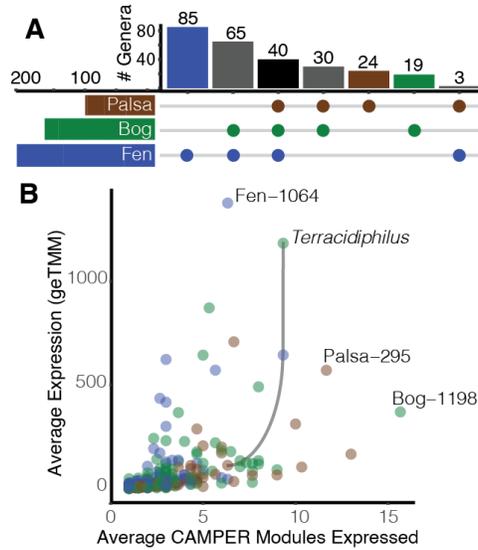
**Figure 4.2.** Polyphenol characterization at Stordalen Mire. (A) The average of Folin-Ciocalteu phenolics, in mg Methyl-Gallate equivalents per gram dry soil, per compartment (n=3). Error bars represent one standard deviation. The bog assay values are significantly higher than the fen, marked by an asterisk (\*, t-test,  $p$ -value=0.013). (B) Heatmap showing mass spectrometry detected polyphenols across the habitats, palsa (brown), bog (green), fen (blue). Pyrogallol (black box) was detected across all habitats. Bottom row corresponds to polyphenol family based on established ontology. Prenylated flavonoids (red names) are shown at right.



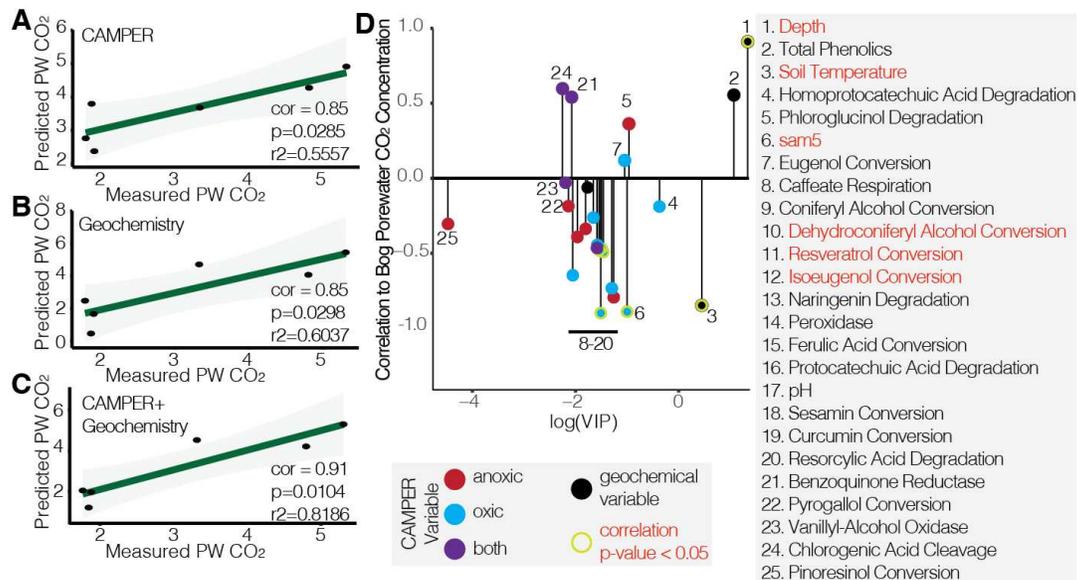
**Figure 4.3.** Polyphenol metabolic potential encoded in Stordalen Mire. (A) Curated Annotations for Microbial Polyphenol Enzymes and Reactions (CAMPER) consists of 101 polyphenol transformation modules, organized by substrates (Family and Subfamily) and oxygen requirements. Modules can be composed of 1-12 reaction steps, indicated by black circle size. (B) The 1,864 dereplicated Stordalen mire metagenome assembled genomes (MAGs) were placed onto the GTDB backbone trees for bacteria and archaea. The inner ring corresponds to phylum and the middle ring corresponds to the first unnamed taxonomic level for the MAG (ex. g\_\_Rhodospila;s\_\_ would be a new species). The outer ring displays the number of CAMPER modules encoded by each MAG, colored according to the polyphenol families in (A). (C) The distribution of the number of CAMPER modules encoded per MAG. Each circle corresponds to a MAG, and is colored by phylum. (D) The top 100 CAMPER module encoding MAGs are plotted, organized by genus. The y-axis displays the number of modules encoded. Names colored in red correspond to named genera, while black names represent potentially novel or undescribed lineages.



**Figure 4.4.** CAMPER module expression across Stordalen Mire. (A) Heatmap showing the standardized expression of CAMPER modules across the habitats, palsa (brown), bog (green), fen (blue). The top row displays the oxygen requirements of the modules: blue (oxic), red(anoxic), purple (both). Boxes highlight habitat expression trends. Modules colored in red are shown in (B-D). Structures shown in orange are detected in metabolite data.



**Figure 4.5.** Active CAMPER expressing genera. (A) Upset plot showing the habitat distribution of CAMPER-active genera across the habitats and those shared between habitats. (B) For each genus, the summed expression of CAMPER modules was calculated per habitat depth, then averaged across all depths within a habitat to give an average of polyphenol gene expression by habitat. Similarly, the number of modules expressed was summed across MAGs per compartment, then averaged across compartments in a site. Circles correspond to a given genus in a given site, colored by the site. The “top” polyphenol-active genera are labelled. A line connects the *Terracidiphilus* datapoints across habitats.



**Figure 4.6.** CAMPER expression predicts bog porewater carbon dioxide (CO<sub>2</sub>) concentrations. (A) Linear regression of measured and sparse partial least squares (sPLS)-predicted bog porewater (PW) CO<sub>2</sub> concentrations considering summed expression of CAMPER modules. Error bands represent 95% confidence intervals, each point is a sample. (B) Linear regression of measured and sparse partial least squares (sPLS)-predicted bog porewater (PW) CO<sub>2</sub> concentrations considering geochemical data. Error bands represent 95% confidence intervals, each point is a sample. (C) Linear regression of measured and sparse partial least squares (sPLS)-predicted bog porewater (PW) CO<sub>2</sub> concentrations constructed from summed expression of CAMPER modules and geochemistry. Error bands represent 95% confidence intervals, each point is a sample. (D) All explanatory variables from the combined geochemistry-CAMPER model are shown. On the y-axis, the Pearson correlation coefficient for that variable to bog porewater CO<sub>2</sub> is plotted, with significant (p-value < 0.05) correlations highlighted in green and named in red.

## Chapter 4 References

1. Hugelius, G. *et al.* Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* **11**, 6573–6593 (2014).
2. Turetsky, M. R. *et al.* Carbon release through abrupt permafrost thaw. *Nat. Geosci.* **13**, 138–143 (2020).
3. Woodcroft, B. J. *et al.* Genome-centric view of carbon processing in thawing permafrost. *Nature* **560**, 49–54 (2018).
4. Koven, C. D. *et al.* Permafrost carbon-climate feedbacks accelerate global warming. *Proc. Natl. Acad. Sci.* **108**, 14769–14774 (2011).
5. Freeman, C., Ostle, N. & Kang, H. An enzymic “latch” on a global carbon store: A shortage of oxygen locks up carbon in peatlands by restraining a single enzymes. *Nature* **409**, 149 (2001).
6. Fenner, N. & Freeman, C. Drought-induced carbon loss in peatlands. *Nat. Geosci.* **4**, 895–900 (2011).
7. Fenner, N. & Freeman, C. Woody litter protects peat carbon stocks during drought. *Nat. Clim. Chang.* **10**, 363–369 (2020).
8. Rothwell, J. A. *et al.* Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database* **2013**, bat070–bat070 (2013).
9. McGivern, B. B. *et al.* Decrypting bacterial polyphenol metabolism in an anoxic wetland soil. *Nat. Commun.* **12**, 2466 (2021).

10. Johansson, T. *et al.* Decadal vegetation changes in a northern peatland, greenhouse gas fluxes and net radiative forcing. *Glob. Chang. Biol.* **12**, 2352–2369 (2006).
11. Urbanová, Z. & Hájek, T. Revisiting the concept of ‘enzymic latch’ on carbon in peatlands. *Sci. Total Environ.* **779**, 146384 (2021).
12. Sánchez-Rangel, J. C., Benavides, J., Heredia, J. B., Cisneros-Zevallos, L. & Jacobo-Velázquez, D. A. The Folin-Ciocalteu assay revisited: Improvement of its specificity for total phenolic content determination. *Anal. Methods* **5**, 5990–5999 (2013).
13. Tfaily, M. M. *et al.* Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Anal. Chim. Acta* (2017). doi:10.1016/j.aca.2017.03.031
14. Yang, X. *et al.* Prenylated flavonoids, promising nutraceuticals with impressive biological activities. *Trends Food Sci. Technol.* **44**, 93–104 (2015).
15. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**, 1925–1927 (2020).
16. Perez, J. M. *et al.* Redundancy in Aromatic O -Demethylation and Ring-Opening Reactions in *Novosphingobium aromaticivorans* and Their Impact in the Metabolism of Plant-Derived Phenolics. *Appl. Environ. Microbiol.* **87**, (2021).
17. Marin, A. M. *et al.* Naringenin degradation by the endophytic diazotroph *Herbaspirillum seropedicae* SmR1. *Microbiology* **159**, 167–175 (2013).
18. Fudyma, J. D. *et al.* Untargeted metabolomic profiling of *Sphagnum fallax* reveals

- novel antimicrobial metabolites. *Plant Direct* **3**, e00179 (2019).
19. Hough, M. *et al.* Coupling plant litter quantity to a novel metric for litter quality explains C storage changes in a thawing permafrost peatland. *Glob. Chang. Biol.* **28**, 950–968 (2022).
  20. Yao, J., Gui, L. & Long, Q. A chlorogenic acid esterase from a metagenomic library with unique substrate specificity and its application in caffeic and ferulic acid production from agricultural byproducts. *Biocatal. Biotransformation* 1–9 (2021). doi:10.1080/10242422.2021.1969370
  21. Achterholt, S., Priefert, H. & Steinbüchel, A. Purification and Characterization of the Coniferyl Aldehyde Dehydrogenase from *Pseudomonas* sp. Strain HR199 and Molecular Characterization of the Gene. *J. Bacteriol.* **180**, 4387–4391 (1998).
  22. McCalley, C. K. *et al.* Methane dynamics regulated by microbial community response to permafrost thaw. *Nature* **514**, 478–481 (2014).
  23. Hess, V., Gonzalez, J. M., Parthasarathy, A., Buckel, W. & Mueller, V. Caffeate Respiration in the Acetogenic Bacterium *Acetobacterium woodii*: a Coenzyme A Loop Saves Energy for Caffeate Activation. *Appl. Environ. Microbiol.* **79**, 1942–1947 (2013).
  24. García-Fraile, P., Benada, O., Cajthaml, T., Baldrian, P. & Lladó, S. *Terracidiphilus gabretensis* gen. nov., sp. nov., an Abundant and Active Forest Soil Acidobacterium Important in Organic Matter Transformation. *Appl. Environ. Microbiol.* **82**, 560–569 (2016).
  25. Cory, A. B. *et al.* Quantifying the inhibitory impact of soluble phenolics on anaerobic carbon mineralization in a thawing permafrost peatland. *PLoS One* **17**,

- e0252743 (2022).
26. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
  27. Shaffer, M. *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).
  28. Xu, S. *et al.* ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data. *Mol. Biol. Evol.* **38**, 4039–4042 (2021).
  29. Bushnell, B. BBtools. <https://jgi.doe.gov/data-and-tools/software-tools/bbtools/>
  30. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
  31. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
  32. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
  33. Smid, M. *et al.* Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics* **19**, 236 (2018).
  34. Saiya-Cork, K. ., Sinsabaugh, R. . & Zak, D. . The effects of long term nitrogen deposition on extracellular enzyme activity in an *Acer saccharum* forest soil. *Soil Biol. Biochem.* **34**, 1309–1315 (2002).
  35. DeForest, J. L. The influence of time, storage temperature, and substrate age on potential soil enzyme activity in acidic forest soils using MUB-linked substrates

- and l-DOPA. *Soil Biol. Biochem.* **41**, 1180–1186 (2009).
36. Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* **6**, 230–235 (2008).
  37. Tolić, N. *et al.* Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra. *Anal. Chem.* **89**, 12659–12665 (2017).
  38. Tfaily, M. M. *et al.* Vertical Stratification of Peat Pore Water Dissolved Organic Matter Composition in a Peat Bog in Northern Minnesota. *J. Geophys. Res. Biogeosciences* **123**, 479–494 (2018).
  39. Laszakovits, J. R. & MacKay, A. A. Data-Based Chemical Class Regions for Van Krevelen Diagrams. *J. Am. Soc. Mass Spectrom.* **33**, 198–202 (2022).
  40. Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R. & Wu, C. H. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* **23**, 1282–1288 (2007).
  41. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
  42. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
  43. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood

- Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
44. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
  45. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res.* **46**, e59–e59 (2018).
  46. Shen, H. & Huang, J. Z. Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivar. Anal.* **99**, 1015–1034 (2008).
  47. Rohart, F., Gautier, B., Singh, A. & Lê Cao, K.-A. mixOmics: An R package for ‘omics feature selection and multiple data integration. *PLOS Comput. Biol.* **13**, e1005752 (2017).
  48. Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **78**, 103–112 (2005).

## Chapter 5 – Conclusion

### **5.1 Summary**

My dissertation research describes microbial polyphenol metabolism under anoxic conditions and explores how this metabolism relates to soil carbon cycles. This was first accomplished in a lab setting, showing soil microcosm microbial communities could degrade a model polyphenol under anoxic conditions (**Chapter 2**). The known microbial genes for degrading polyphenol and phenolic compounds were then assessed and used to develop a publicly-available tool for annotating microbial polyphenol metabolism in genomic data (**Chapter 3**). Finally, polyphenol metabolism was surveyed in a natural permafrost peatland system, revealing it is broadly encoded across this microbiome and actively expressed in field metatranscriptome data. More so, active polyphenol metabolism could predict porewater carbon dioxide concentrations in a bog system, highlighting the potential for using this metabolism to predict carbon emissions (**Chapter 4**).

### **5.2 Future Research Directions**

Beyond anoxic soils, I believe polyphenol metabolism impacts both crop production and the sustainability of livestock practices.

#### *5.2.1 Polyphenol metabolism in agricultural soils*

Chemical analyses of crop root exudates have highlighted the presence of polyphenols, and particularly flavonoids<sup>1</sup>. These compounds can serve as signaling molecules in the case of attracting nodule-forming diazotrophs in legumes, or be exuded in stress response where they are presumably recruiting a beneficial

rhizobiome<sup>2</sup>. However, flavonoids meet abiotic<sup>1</sup> and biotic<sup>3</sup> fates in agricultural soils, which ultimately serve to dampen their signaling efficacy. Therefore, it is important to understand the microbial mechanisms controlling polyphenol concentrations in agricultural soils to predict and optimize crop fitness. CAMPER can enable some of these investigations, but this area offers the opportunity to discover more polyphenol metabolizing organisms and their enzymes.

### *5.2.2 Polyphenol metabolism in ruminants*

Ruminants and their enteric fermentation contribute nearly 25% of global methane (CH<sub>4</sub>) emissions<sup>4</sup>. Previous data suggests addition of the polyphenol catechin to cattle feed prevents the emission of CH<sub>4</sub>. One hypothesized mechanism underlying this inhibition is reductive degradation of catechin by ruminant microbes, limiting hydrogen available for the ruminant methanogens and preventing their ability to produce methane<sup>5</sup>. However, the identity of catechin degrading microbes in the rumen, and the genes and enzymes enabling the degradation, are unknown. This knowledge could help in optimizing ruminant nutrition and reducing CH<sub>4</sub> outputs by pairing catechin-amended diets with a probiotic of catechin degrading microorganisms.

In summary, there is a lot of potential to better understand and manage our agricultural systems if we have a full picture of microbial polyphenol metabolism. While my work has endeavored to address this in anoxic soils, I am excited to see what more will be learned about the microbes who can use polyphenols and the enzymes they possess across systems.

## Chapter 5 References

1. Del Valle I, Webster TM, Cheng H-Y, Thies JE, Kessler A, Miller MK, Ball ZT, MacKenzie KR, Masiello CA, Silberg JJ, Lehmann J. 2020. Soil organic matter attenuates the efficacy of flavonoid-based plant-microbe communication. *Sci Adv.*
2. Cloutier M, Chatterjee D, Elango D, Cui J, Bruns MA, Chopra S. 2021. Sorghum Root Flavonoid Chemistry, Cultivar, and Frost Stress Effects on Rhizosphere Bacteria and Fungi. *Phytobiomes J* 5:39–50.
3. Shaw LJ, Hooker JE. 2008. The fate and toxicity of the flavonoids naringenin and formononetin in soil. *Soil Biol Biochem* 40:528–536.
4. Moss AR, Jouany J-P, Newbold J. 2000. Methane production by ruminants: its contribution to global warming. *Ann Zootech* 49:231–253.
5. Becker PM, van Wakselaar PG, Franssen MCR, de Vos RCH, Hall RD, Beekwilder J. 2014. Evidence for a hydrogen-sink mechanism of (+)catechin-mediated emission reduction of the ruminant greenhouse gas methane. *Metabolomics* 10:179–189.

## Appendices

### ***Appendix A: Chapter 2 Supplementary Text***

#### *Sterilized Soils*

The soil slurry used to inoculate samples was autoclaved three consecutive times on liquid-20 cycles. Unlike the biotic samples, there was no visible biomass pellet following centrifugation of the 5mL of this sample at day 20. We attempted to extract DNA from these samples at day 20, and failed to obtain Qubit-HS detectable DNA from 5mL of the of the sample. We attempted to obtain Illumina MiSeq 16S rRNA amplicon sequencing from this sample, and we were unable to obtain any amplicons.

#### *Comparison of CT degradation metabolites between soils and guts*

Many of the metabolites we detected in soils have also been reported in studies of gut microbiome metabolism of CT. However, we did not detect any phenyl valeric acids or lactones (C6-C5) that are widely reported as products of flavonoid metabolism in the gut<sup>1</sup>. These compounds are proposed to be formed by transformation of diphenylpropanes after ring opening at position 1 of the C-ring<sup>2</sup>, where we instead proposed CT depolymerization is followed by flavonoid transformations involving C-ring fission (position 1 and 4) and release of C6-C3 products such as 3-hydroxyphenylpropanoic acid (**Figure 2.4**). Regardless of the exact mechanisms, our data broadly echo the proposed steps for flavonoid degradation in the gut, with chain shortening by loss of CO<sub>2</sub> and/or acetate to yield small molecules that could enter central metabolic pathways (**Figure 2.9**).

#### *Database recruitment*

Our dereplicated MAG database recruited on average 43% and 12% of metagenomic reads for CT-amended and unamended samples, respectively. However, we knew the dominant MAGs Kosakonia, Holophaga, and the three UBA7701 MAGs were dominant in the CT-amended samples, and nearly absent in unamended samples (**Figure 2.12**). Therefore, we repeated this analysis removing these MAGs and found 9% and 12% of metagenomic reads mapped to the adjusted MAG database in CT-amended and unamended samples, respectively (**Figure 2.13**, black bars), thereby confirming we did not observe bias due to preferential assembly and binning with a treatment.

To assess overall metagenome assembly, we compared metagenomic read mapping to assembled scaffolds >2500 bp. We found on average 50% and 25% of reads were mapped to assembled scaffolds >2500 bp for CT-amended and unamended samples, respectively. Again, removing scaffolds from the dominant MAGs, we found 18% and 25% of reads were mapped to assembled scaffolds >2500 bp for CT-amended and unamended samples, respectively (**Figure 2.13**, green bars). Therefore, while we had very dominant members in our CT amended microcosms, when we quantified read recruitment beyond those genomes we recovered similar proportions of the microbial community at the MAG and assembly levels.

To ensure we had equal representation of genes from our microbial communities in our metaproteome gene database, we again mapped our metagenomic reads to this dereplicated gene database. We found on average 40% and 22% of our reads mapped to the database for CT-amended and unamended samples, respectively. When we removed genes derived from the five CT-dominant MAGs from this database, we saw

21% and 22% of metagenomic reads mapped from CT-amended and unamended samples, respectively (**Figure 2.13**, blue bars). This again reinforced that our database captured the same proportion of the microbial communities between treatments.

### *Nitrogen Cycling*

While we failed to find evidence that tannins inhibited carbon degradation, we did observe lower levels of protein-like compounds by FTICR-MS in live CT-amended reactors (**Figure 2.23**), which could be the result of tannin-bound organic nitrogen as has been previously suggested<sup>3</sup>. Furthermore, while we observed peptidases in both treatments, we detected expression of 15 peptidase-classes by 9 members under CT treatment, and just one class in the control using our unique (binned and unbinned) data. This increased peptidase diversity could reflect nitrogen limitation under CT exposure, demonstrating a shared response between our anoxic soils and previously reported aerobic soils<sup>6</sup>. However, we did not observe differences in the expression of denitrification enzymes (nitrate reductase, nitrite reductase, and nitrous oxide reductase) between treatments (**Figure 2.20**), suggesting tannins may not repress the entire nitrogen cycle in anoxic soils.

### *HPLC and NMR Methods for characterizing purified CT*

The separation used an Agilent 1100 quaternary pump system paired with an autosampler and a DAD detector collecting data at 220 nm. The column was a ThermoFisher Hypersil Gold C8, 160 x 4.6 mm column with 3 micron packing. The sample injection volume was typically 10  $\mu$ L. The gradient program was controlled with ChemStation Rev. A.09.03 software and employed 0.13% (v/v) trifluoroacetic acid (TFA) in nanopure water (A) and 0.10% (v/v) TFA in acetonitrile (B) at a flow rate of 0.5

mL/min for 48 min duration. The 28 min separation phase (0-3 min, isocratic at 15% B; 3-8 min, increase to 20% B; 8-10 min, increase to 30% B; 10-28 min, isocratic at 30% B) followed by a wash & re-equilibration phase (28-32 min, increase to 70% B; 37-40 min, decrease to 15% B and hold isocratic at 15% for 8 min before the next injection).

<sup>1</sup>H-<sup>13</sup>C HSQC NMR spectra were recorded at 27°C on a BrukerBiospin DMX-500 (<sup>1</sup>H 500.13 MHz, <sup>13</sup>C 125.76 MHz) instrument equipped with TopSpin 3.5 software and a cryogenically cooled 5 mm TXI <sup>1</sup>H/<sup>13</sup>C/<sup>15</sup>N gradient probe in inverse geometry. Spectra were recorded in DMSO-d<sub>6</sub> and were referenced to the residual signals of DMSO-d<sub>6</sub> (2.49 ppm for <sup>1</sup>H and 39.5 ppm for <sup>13</sup>C spectra). For <sup>1</sup>H-<sup>13</sup>C HSQC experiments, spectra were obtained using between 200 and 620 scans (depending on sample size and instrument availability) obtained using the standard Bruker pulse program (hsqcetgpsisp.2) with the following parameters: Acquisition: TD 1024 (F2), 256 (F1); SW 16.0 ppm (F2), 165 ppm (F1); O1 2350.61 Hz; O2 9431.83 Hz; D1 = 1.50 s; CNST2 = 145. Acquisition time: F2 channel, 64 ms, F1 channel 6.17 ms. Processing: SI =1024 (F2, F1), WDW = QSINE, LB = 1.00 Hz (F2), 0.30 Hz (F1); PH\_mod = pk; Baseline correction ABSG = 5 (F2, F1), BCFW = 1.00 ppm, BC\_mod = quad (F2), no (F1); Linear prediction = no (F2), LPfr (F1). Sample sizes used for these spectra ranged from 5-10 mg providing NMR sample solutions with concentrations of 10-20 mg/mL.

#### *Additional FTICR-MS methods*

A 9.4T Tesla Bruker Solarix FTICR spectrometer located at the University of Arizona, was used to collect high resolution mass spectra of the supernatant samples from reactors (microcosms) by direct injection in negative ion mode. Samples were first desalted and concentrated using solid phase concentration (SPE) according to Dittmar

et al. 2008<sup>4</sup>. Briefly, SPE cartridges (PPL, 300 mg, Varian Mega Bond Elut, Varian Inc., Palo Alto, CA, USA) were rinsed with one cartridge volume (3 mL) of methanol (p.a.) immediately before use. Three milliliters of each of the supernatant samples was acidified with hydrochloric acid (p.a.) to pH 2 and pumped through the SPE cartridge, at a flow rate of <50 mL/min. Before elution of DOM with methanol, the cartridges were rinsed with at least two cartridge volumes of 0.01 M HCl for complete removal of salts. Sorbents were then dried under a stream of N<sub>2</sub> and DOM then eluted with 1 mL of methanol at a flow rate of <10 mL/min. A standard Bruker ESI source was used to generate negatively charged molecular ions and then eluted DOM samples were introduced directly to the ESI source. The instrument was externally calibrated weekly to a mass accuracy of <0.1 ppm using a tuning solution from Agilent, which contains the following compounds: C<sub>2</sub>F<sub>3</sub>O<sub>2</sub>, C<sub>6</sub>H<sub>9</sub>N<sub>3</sub>O, C<sub>12</sub>H<sub>21</sub>N<sub>3</sub>O, C<sub>20</sub>H<sub>18</sub>F<sub>27</sub>N<sub>3</sub>O<sub>8</sub>P<sub>3</sub>, and C<sub>26</sub>H<sub>18</sub>F<sub>39</sub>N<sub>3</sub>O<sub>8</sub>P<sub>3</sub> with an m/z ranging between 112 and 1333. The instrument settings were optimized by tuning on a Suwannee River Fulvic Acid (SRFA) standard. Blanks (HPLC grade MeOH) were ran at the beginning and the end of the day to monitor potential carry over from one sample to another and the instrument was flushed between samples using a mixture of water and methanol. The ion accumulation time (IAT) was varied to account for differences in C concentration between samples. Three hundred individual scans were averaged for each sample and internally calibrated using organic matter and (epi)catechin CT oligomers homologous series separated by 14 Da (-CH<sub>2</sub> groups). The mass measurement accuracy was <1 ppm across a broad m/z range (i.e. 100 < m/z < 1000). The mass resolution was 350,000 at 339.112 Da.

The observed spectra contained a mixture of singly and doubly charged ions (inferred from KMD analysis), and some remaining salt clusters and background noise peaks that were deleted manually. We applied Kendrick Mass Defect analysis using  $\text{CH}_2$  as the base unit to data from all samples. In plotting this data (**Figure 2.5**), there were clear regions that were mainly composed of points from CT-amended samples (**Figure 2.6**, red ovals). We used this to extract peaks that likely corresponded to the different oligomers of the CT polymer and derived compounds.

To verify the presence of doubly and singly charged (epi)catechin CT oligomers within our samples, we used a combination of Compass Isotope Pattern calculation and isotopic fine structure (IFS) simulation and evaluation using Bruker Data Analysis software (**Figure 2.6**). In mass spectrometry, isotope profiles are displayed dependent on the effective resolution of the instrument to resolve the isotope variants. In this sense, FTICR-MS differs from lower resolution instruments since its resolving power is usually high enough to baseline separate the isotope profile of intact and multiple charged ions. Thus, IFS analysis is useful here for identification of (epi)catechin CT oligomers since (1) IFS is an exact fingerprint for every possible molecular configuration and (2) the high resolving power of the FTICR-MS.

Kendrick mass defect analysis<sup>5</sup> was then used to compare the fate of (epi)catechin CT oligomers over time in both biologically active and inactive (autoclaved soil) reactors amended with CT. Kendrick plots are generated by plotting nominal mass as a function of Kendrick mass defect, most commonly based on a methylene ( $\text{CH}_2$ ) subunit, although other subunits can be used (e.g., carboxyl), in a single 2D display. Thus, compounds whose elemental composition differs by a number of base units only

possess the same KMD value and line up horizontally in the associated KMD plot. Here we used a modified version of KMD commonly used for polymer ions, proposed by Sato et al<sup>6,7</sup> where the repeating unit of a polymer backbone (here (epi)catechin) was used as the base unit by arbitrarily setting its mass at the nearest integer and other masses are re-calculated based on that. To walk through the orientation of peaks corresponding to CT-oligomers within a Kendrick plot based on mass, charge, and isotopic composition, see **Figure 2.6**.

Putative chemical formulas of the remaining singly charged ions and in-situ soil metabolites were assigned using Formularity software<sup>7</sup>. First, to further reduce cumulative errors, all sample peak lists for the entire dataset were aligned to eliminate possible mass shifts that would impact formula assignment. Chemical formulas were assigned based on the following criteria:  $S/N > 7$ , and mass measurement error  $< 1$  ppm, taking into consideration the presence of C, H, O, N, S and P and excluding other elements. Peaks with large mass ratios ( $m/z$  values  $> 500$  Da) often have multiple possible candidate formulas. These peaks were assigned formulas through propagation of  $CH_2$ , O, and  $H_2$  homologous series. Additionally, to ensure consistent choice of molecular formula when multiple formula candidates are found the following rules were implemented: the formula with the lowest error with the lowest number of heteroatoms was consistently picked and the assignment of one phosphorus atom required the presence of at least four oxygen atoms. Biochemical compound classes were reported as relative abundance values based on counts of C, H, and O for the following H:C and O:C ranges only : lipid-like ( $0 < O:C \leq 0.3$  and  $1.5 \leq H:C \leq 2.5$ ), unsaturated hydrocarbons-like ( $0 \leq O:C \leq 0.125$  and  $0.8 \leq H:C < 2.5$ ), protein-like ( $0.3 < O:C \leq 0.55$

and  $1.5 \leq \text{H:C} \leq 2.3$ ), carbohydrates-like ( $0.55 < \text{O:C} \leq 0.7$  and  $1.5 \leq \text{H:C} \leq 2.2$ ), lignin-like ( $0.125 < \text{O:C} \leq 0.65$ ,  $0.8 \leq \text{H:C} < 1.5$ ), tannin-like ( $0.65 < \text{O:C} \leq 1.1$ ,  $0.8 \leq \text{H:C} \leq 1.5$ ), and condensed hydrocarbon-like ( $0 \leq \text{O:C} \leq 0.95$ ,  $0.2 \leq \text{H:C} < 0.8$ )<sup>8</sup>. To identify “polyphenol-like” compounds, we grouped lignin-like and tannin-like compounds (**Figure 2.3, Figure 2.23**).

#### *Additional LC-MS methods*

Metabolites were extracted into ethyl acetate from filtered supernatant samples that were acidified with HCl. Both the aqueous and organic phases were analyzed by liquid chromatography - tandem mass spectrometry (LC-MS/MS). Sample aliquots were stored frozen at  $-80^{\circ}\text{C}$  until extraction. Aliquots were thawed, and filtered through 0.2  $\mu\text{m}$  cellulose acetate to yield 1 mL filtrates that were vortexed, and then acidified by addition of 200  $\mu\text{L}$  of 1M HCl. The acidified samples were then extracted 3 times with addition of 200  $\mu\text{L}$  of water-saturated ethyl acetate, each time followed by vortexing, sonication for 10 minutes in a water bath and collection of the ethyl acetate layer (ethyl acetate layers from the 3 extraction steps were combined). At the end the water layer was also collected. The ethyl acetate extracts were dried at room temperature in a speed vacuum concentrator; the aqueous layer was frozen at  $-80^{\circ}\text{C}$  and then dried by lyophilization. Dried extracts of both types were resuspended in 50  $\mu\text{L}$  of 80% methanol containing 1  $\mu\text{g/mL}$  2-Amino-3-bromo-5-methylbenzoic acid (Sigma) as internal standard. Resuspended samples were vortexed, centrifuged (2070 RCF for 5 minutes at room temperature), then filtered via 0.22  $\mu\text{m}$  pvdf microcentrifuge filtration tubes. Filtrates were then analyzed by LC-MS/MS using an Agilent 1290 UHPLC system connected to a Thermo Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer

equipped with a Heated Electrospray Ionization (HESI-II) source probe. The filtrates from the ethyl acetate extraction were chromatographically separated on a ZORBAX RRHD Eclipse Plus C18, 95Å, 2.1 x 50 mm, 1.8 µm column (Agilent) and filtrates from the aqueous layer were chromatographically separated on an InfinityLab Poroshell 120 HILIC-Z, 2.1 × 150 mm, 2.7 µm (Agilent). Separation, ionization, fragmentation and data acquisition parameters are specified in **Appendix B**. Briefly, metabolites were separated by gradient elution followed by MS1 and data dependent (top 2 most abundant MS1 ions not previously fragmented in last 7 seconds) MS2 collection; targeted data analysis was performed by comparison of sample peaks to a library of analytical standards analyzed under the same conditions. Three parameters were compared: matching m/z, retention time and fragmentation spectra using Metabolite Atlas (<https://github.com/biorack/metatlas>)<sup>9,10</sup>. Identification and standard reference comparison details are provided in **Appendix B**. Data are available for download at the JGI Joint Genome Portal under ID 1281268. To determine significantly discriminating LC-MS exometabolites, we used limma<sup>11</sup> in R on log2-transformed data to compare metabolites in live and autoclaved treatments at each timepoint.

#### *Appendix A References*

1. Del Rio, D. et al. Dietary (poly)phenolics in human health: Structures, bioavailability, and evidence of protective effects against chronic diseases. *Antioxidants and Redox Signaling* 18, 1818–1892 (2013).
2. Sánchez-Patán, F. et al. Capability of lactobacillus plantarum IFPL935 to catabolize flavan-3-ol compounds and complex phenolic extracts. *J. Agric. Food Chem.* 60, 7142–7151 (2012).

3. Fierer, N., Schimel, J. P., Cates, R. G. & Zou, J. Influence of balsam poplar tannin fractions on carbon and nitrogen dynamics in Alaskan taiga floodplain soils. *Soil Biol. Biochem.* 33, 1827–1839 (2001).
4. Dittmar, T., Koch, B., Hertkorn, N. & Kattner, G. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. *Limnol. Oceanogr. Methods* 6, 230–235 (2008).
5. Hughey, C. A., Hendrickson, C. L., Rodgers, R. P., Marshall, A. G. & Qian, K. Kendrick mass defect spectrum: A compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem.* 73, 4676–81 (2001).
6. Sato, H., Nakamura, S., Teramoto, K. & Sato, T. Structural characterization of polymers by MALDI spiral-TOF mass spectrometry combined with Kendrick mass defect analysis. *J. Am. Soc. Mass Spectrom.* 25, 1346–1355 (2014).
7. Tolić, N. et al. Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra. *Anal. Chem.* 89, 12659–12665 (2017).
8. Tfaily, M. M. et al. Sequential extraction protocol for organic matter from soils and sediments using high resolution mass spectrometry. *Anal. Chim. Acta* 972, 54–61 (2017).
9. Yao, Y. et al. Analysis of metabolomics datasets with high-performance computing and metabolite atlases. *Metabolites* 5, 431–442 (2015).
10. Bowen, B. P. & Northen, T. R. Dealing with the unknown: Metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* 21, 1471–1476 (2010).

11. Smyth, G. K. limma: Linear Models for Microarray Data. in Bioinformatics and Computational Biology Solutions Using R and Bioconductor 397–420 (Springer-Verlag, 2005). doi:10.1007/0-387-29362-0\_23

## ***Appendix B: Chapter 2 Supplementary Files***

See supplemental file AppendixB.zip.

*Supplementary Data 1.* (xlsx) Metabolite identification data for LC-MS and NMR, including specifics of LC-MS methodology.

*Supplementary Data 2.* (xlsx) Information on metagenomes, assemblies, MAGs, and 16S rRNA gene reads, including accession numbers.

*Supplementary Data 3.* (xlsx) Information on metaproteomes including table of MAG peptide recruitment by sample, peptide recruitment/functional annotation for detected proteins, and enzyme information for flavonoid enzymes.

*Supplementary Data 4.* (xlsx) Feature table derived from 16S rRNA gene analyses.

*Supplementary Data 5.* (amino acid FASTA) Flavonoid enzymes queried against this dataset, with accession numbers in FASTA header.

*Supplementary Data 6.* (amino acid FASTA) Amino acid sequences of Flavonoid enzyme encoding genes CHI, FCR, PHY, and PGR, and *Kosakonia* KatG and AA6.

*Supplementary Data 7.* (xlsx) FTICR-MS data for all peaks across all samples, including formula assignment if applicable and Kendrick Mass defect analysis.

### ***Appendix C: Chapter 3 Supplementary Files***

See supplemental file AppendixC.zip.

*CAMPER\_blast.fa*: A fasta file of CAMPER genes used as a target in a BLAST style search provided by mmseqs search.

*CAMPER.hmm*: A HMM file used as the target in an HMM profile search provide by MMseqs profilesearch

*CAMPER\_blast\_scores.tsv*: Provides the minimum cut off scores for search results and quality ranks with BLAST style searches.

*CAMPER\_hmm\_scores.tsv*: Provides the minimum cut off scores for search results and quality ranks with HMM Profile searches.

*CAMPER\_distillate.tsv*: A custom distillate, for use with DRAM or with CAMPER\_DRAMKit, to summarize the annotation results.

*CAMPER\_refs.xlsx*: References for the genes behind the 41 custom CAMPER annotations.